Interactive Journal of Medical Research

Impact Factor (2022): 2.0
Volume 12 (2023) ISSN 1929-073X

Contents

Viewpoints

Reviews

Original Papers

Short Paper

Case Reports

Research Letter

Viewpoint

Using Decision Trees as an Expert System for Clinical Decision Support for COVID-19

Dillon Chrimes[1], MSc, PhD

School of Health Information Science, Human and Social Development, University of Victoria, Victoria, BC, Canada

Corresponding Author:
Dillon Chrimes, MSc, PhD
School of Health Information Science
Human and Social Development
University of Victoria
HSD Building, A202
Victoria, BC, V8W2Y2
Canada
Phone: 1 250 472 4474
Email: dchrimes@uvic.ca

Abstract

COVID-19 has impacted billions of people and health care systems globally. However, there is currently no publicly available chatbot for patients and care providers to determine the potential severity of a COVID-19 infection or the possible biological system responses and comorbidities that can contribute to the development of severe cases of COVID-19. This preliminary investigation assesses this lack of a COVID-19 case-by-case chatbot into consideration when building a decision tree with binary classification that was stratified by age and body system, viral infection, comorbidities, and any manifestations. After reviewing the relevant literature, a decision tree was constructed using a suite of tools to build a stratified framework for a chatbot application and interaction with users. A total of 212 nodes were established that were stratified from lung to heart conditions along body systems, medical conditions, comorbidities, and relevant manifestations described in the literature. This resulted in a possible 63,360 scenarios, offering a method toward understanding the data needed to validate the decision tree and highlighting the complicated nature of severe cases of COVID-19. The decision tree confirms that stratification of the viral infection with the body system while incorporating comorbidities and manifestations strengthens the framework. Despite limitations of a viable clinical decision tree for COVID-19 cases, this prototype application provides insight into the type of data required for effective decision support.

(Interact J Med Res 2023;12:e42540) doi:10.2196/42540

KEYWORDS

assessment tool; chatbot; clinical decision support; COVID-19; decision tree; digital health tool; framework; health informatics; health intervention; prototype

Introduction

COVID-19 presents a major and urgent threat to global health. Since the original outbreak of the disease in early December 2019 in Wuhan, China, COVID-19 has spread to over 188 countries, with the number of cases recently estimated at ~50-100 million and the number of fatalities at ~1-2 million, putting the death rate at approximately 3% [1-4]. In Canada, 30-40% of Canadians have contracted COVID-19, and the proportion of the infected population who experience severe illness is estimated at ~10% [5]. Clinicians dealing with this disease would benefit from an expert tool to rapidly assess severe COVID-19 cases.

The sources and modes of transmission of COVID-19 remain unpredictable [1], despite well-established scientific knowledge of the biology of the virus. COVID-19 is highly contagious, with a transmission capacity far greater than that of the previous severe acute respiratory syndrome (SARS) epidemic in 2003, and the virus reaches an extremely high abundance in infected people (up to 1 billion RNA copies per milliliter of sputum) [6]. Moreover, SARS-CoV-2 is stable on plastic and steel objects, where viable virus can remain detectable for more than 72 hours [6]. The wide, rapid spread of COVID-19, together with the population-level mortality rate have occasioned substantial research on its physiological and health effects, including its manifestations (ie, increasing effects on human biological functions) in conjunction with existing morbidities [1]. This formative study aimed at making accurate causal inferences with a view toward predictive modeling. Yet, the physiological interaction between COVID-19 and the human body is not yet well-established. Therefore, research models need to be reframed or even reinvented based on an assessment of the problems and strengths of targeting COVID-19, which can be achieved by using several intersecting knowledge bases concerned with human pathophysiology.

Mild cases of COVID-19 are typically characterized by early viral clearance, with 90% of mildly affected patients repeatedly testing negative on reverse transcriptase-polymerase chain reaction tests by day 10 postonset [7]. In contrast, all severe cases were still evaluated as positive for COVID-19 at or beyond day 10 postonset, with the median duration of viral shedding in survivors being 20 days, during which time the affected individuals are highly contagious [8]. In one study, the longest observed duration of viral shedding in survivors was 37 days [9]; other studies reported even longer durations [10]. Modeling of COVID-19 suggests that severe cases can result in chronic respiratory and cardiovascular conditions lasting months and even longer, with recoveries remaining only partial without treatment [7,8]. Therefore, an important aspect of diagnosis is the ability to determine if a case will be mild or severe, especially since—for patients with slow recovery—mild cases can become severe. Conversely, if a case is quickly diagnosed as mild, then the risk of intubation and mechanical ventilation causing further chronic respiratory problems could be minimized [11].

Current scientific research in the epidemiology of COVID-19, as well as public health surveillance, relies on modeling COVID-19 infections to predict outbreaks [1]. One systematic review of over 100 models of COVID-19 found that none of these models was successful either in modeling the outbreak or in predicting recovery versus mortality [8]. Recent studies have used medical images and radiology techniques to develop prognostic and diagnostic models for detecting COVID-19 using applied statistics and machine learning, and a "COVID-19 vulnerability index" was developed related to hospital admissions [7,12]. Radiology can quickly detect COVID-19 as a respiratory disease at first onset and offer a quick indication of COVID-19 severity; however, it does not have a high success rate in detecting severe cases within the important 10-day latency-to-replication phase of this virus [13,14]. Another potential approach is to use blood samples to verify cases of severe COVID-19, without needing X-rays or other tissue samples or biopsy [15]. Because uncertainties remain and patients' responses to infection vary greatly—and because there is no cure yet available for this disease—we need new methods to detect severe cases accurately and quickly.

Epidemiologists and other health professionals are currently working together to model risks using applied Bayes theorem with machine-learning techniques. This approach promises to produce models that can accurately assess the probability of severe cases within a given infected population, as well as complications of cardiovascular diseases [8,16]. Although updates on available treatments are frequent, gaps in the clinical guidelines for treating COVID-19 remain [8], with few validated clinical decision-support tools [17] or clinical guidelines in constructing chatbots for COVID-19 (Table 1).

As the pandemic continues, clinicians need reliable information on human pathophysiological responses to COVID-19. Models commonly applied in prognosis and treatment include the simple nomogram (ie, a diagram between variables such as age, gender, renal function, medication dosage, and body weight), decision trees, score systems, and online tools that provide statistics on a range of metrics: in-hospital deaths, prolonged mechanical ventilation needs, and a composite of poor outcomes [18]. While there is little agreement as to which modeling techniques are useful in a hospital setting [8,19], studies seem to have found value in decision trees. One study of recovery in severe COVID-19 cases, based on testing of blood samples, showed that decision trees using Gini coefficients are more effective than models using support vector machines, logistic regression, or random forest classification [15]. Other studies indicate that decision trees can define complex outcomes of COVID-19, especially for severe in-hospital cases [20]. Thus, decision trees can be valuable to incorporate knowledge structures that support the effective treatment of severe COVID-19.

Recently verified predictive models used specific parameters as predictors in the diagnosis and prognosis of COVID-19: age, body temperature, lymphocyte count, and data obtained through lung imaging [19]. In addition, flu-like symptoms and neutrophil count are frequently used as predictors in diagnosis, while comorbidities, gender, presence of creatinine-reactive protein (CRP), and overall amount of creatinine are frequently used in prognosis [19]. Findings regarding the biology and differentiation of lung inflammation suggest that excessive quantities of cytokines result in inner (microvascular) and outer (macrovascular) heart problems and elevated risk of mortality [16]. Therefore, the main diagnostic-to-prognostic problem is the level of integration of COVID-19 with human biology and the variable degree of immune response. That is, some immune responses are beneficial in the body's fight against COVID-19, while others are detrimental [21]. The diagnostic problem of ascertaining the probabilities of mortality and recovery can be simplified using binary classification, allowing for a variety of recovery-versus-death scenarios. The problem here is that there may be a variety of manifestations, each with its own specific physiological sequence, which can severely impact the body's lung or heart [22]. Therefore, the design of a clinical decision-support tool to diagnose severe cases of COVID-19 would have to consider both these manifestations and any remaining gaps in that understanding.

The knowledge of COVID-19 manifestations is complex, since a range of chronic conditions are associated with lung, cardiovascular, and gastrointestinal diseases. One study showed that preceding coronavirus outbreaks such as SARS and Middle East respiratory syndrome (MERS) were associated with a significant burden of cardiovascular comorbidities [23]. Furthermore, diagnostic workups during SARS infections revealed detailed changes in electrocardiographic results, subclinical left ventricular diastolic impairment, and troponin elevation, all of which varied widely among patients [24]. Therefore, in the context of severe COVID-19 infection, any tools designed to support clinical decision-making need to take into account the interconnectedness of the body's respiratory and cardiovascular systems. This accounting for interconnectedness can be accomplished by decision trees, which can synthesize knowledge structures in the architectural construction of their branches and leaves.

Table 1. Environmental scan of chatbots used for COVID-19 information.

Chatbot	Ages	Description	Setting	Inference method	Intervention effectiveness	Guidelines used	Prescription
Apple and Siri	All	Apple and Siri to help people who ask if they have the coronavirus	Anywhere	Uses data from Johns Hopkins University; rules-based	Only symptom-based	No clinical guidelines	No
Intermountain Healthcare	All	COVID-19 symptom checker	Research, primary care, and acute care in Utah at Intermountain Healthcare	Uses data from acute-care settings; rules-based	Effective to separate mild and severe cases based on symptoms	Simple clinical guidelines	Unknown
Google Dialogflow and Google Assistant	All	Chatbot designed with extensive prompts to entities and intents	Anywhere	Prototype, fact checker, and Q&A[a] for COVID-19	Answers a variety of questions, including data on COVID-19 and symptoms	No clinical guidelines	No
COVIDradar	All	Chatbot and app to track health status	Anywhere within the United Kingdom and National Health Service	Prototype, fact checker, and Q&A for COVID-19	Effective to separate mild and severe cases based on symptoms and daily updates	Simple clinical guidelines	Unknown
Facebook Messenger with WHO[b]	All	COVID-19 fact and symptom checker	Anywhere	Prototype, fact checker, and Q&A for COVID-19	Answers a variety of questions, including data on COVID-19 and symptoms	WHO guidelines	No
BC CDC[c]	All	CDSS[d]-generated care suggestions based on agreed guidelines. These include what to do if testing negative for COVID-19	Research, hospital-based academic groups, possibly including clinicians	Prototype, fact checker, and Q&A for COVID-19	Answers a variety of questions, including data on COVID-19 and symptoms	Simple clinical guidelines	No

[a]Q&A: question and answer.

[b]WHO: World Health Organization.

[c]BC CDC: British Columbia Centre for Disease Control.

[d]CDSS: clinical decision support system.

Design

Decision trees can incorporate medical knowledge [25], including human physiological responses to influenza and other diseases [20]. Human immune responses to COVID-19 can reveal a normal increase in T cells due to inflammation; however, this leads to a subsequent cytokine storm that increases the risk of mortality rather than reducing it. T-cell inflammation as a biological immune response to COVID-19 makes construction of the diagnostic decision branches in the knowledge tree difficult, requiring the introduction of additional calculations such as Gini coefficient thresholds [23]. The knowledge tree also needs to include a transition from diagnosis of chronic lung-to-heart conditions in the tissues, vessels, muscles, and valves. While the stratification of interactions between COVID-19 and chronic lung conditions and heart disease is not yet well understood [26], we have nonetheless been able to draw on existing clinical guidelines to construct the framework of a decision-support tool that models the transition from mild to severe COVID-19 cases.

After reviewing expert systems and other modeling technologies that use decision trees, we decided to use the infrastructure of a chatbot tool [27] that would either design decision trees manually or use Predictive Model Markup Language (PMML) schematic formats to create them. PMML can be utilized in tools such as KNIME and RapidMiner to automate the formation of a decision tree; in such procedures, data sets and trained data can be employed iteratively. In clinical applications, PMML has been applied to binary classification such as wound care management [27]. The graphical artificial intelligence software VisiRule has knowledge engineering, interface, and control/editor tools (Figure 1) with embedded inductive techniques with inheritance settings, set as singular "Depth First" with maximum of 9 from the root node (Figure 2) and forward chaining settings (Figure 3).

Induction (as well as deduction) can be used in the diagnosis of COVID-19. For example, the landmark decision tree program called c4.5 algorithm is a machine-learning workhorse that in its sequence of decision points can establish decision endpoints for classification [28]. Inductive and deductive reasoning can also be employed to modify the construction of the architecture of decision trees.

We attempted to create a diagnostic tool that could develop a decision tree based on existing COVID-19 data sets using a number of applications: Google's Dialogflow, ZenChat, and KNIME. Dialogflow—a decision-support tool in the form of a chatbot—contains intents and entities and a knowledge base; its construction enables easy integration into web-based applications [29]. However, Dialogflow does not allow for binary classification; all it can do is enable a series of prompts in a specific sequence of COVID-19 queries. This makes modeling severe cases of COVID-19 extremely difficult, as we had no ability to incorporate Bayes theorem. Dialogflow thus could not be used to develop our planned chatbot. Another tool we considered, ZenChat, proved similar to Dialogflow, as it had no apparent capacity to generate decision trees using binary classification [30]. In the decision-tree framework in KNIME, which autogenerates a PMML schema, we used data sets on COVID-19 cases drawn from the website Kaggle.com. However, these COVID-19 data sets were not easily formed to binary classifications; notably, the data could not be easily sorted in terms of recovery, short or long recovery with onset of chronic conditions, or even risk of mortality. This difficulty in forming binary classifications was confirmed by running Bayes modeling of the data set in the Microsoft Azure Machine Learning Studio, which did not produce accurate results. Accordingly, we constructed the decision tree to account for health outcomes, stratified by comorbidities in the transition from the lung to the heart, while establishing binary classifications to manually calculate Bayes probabilities for each leaf endpoint (Table 2).

With the VisiRule construct (via upload of a .vsr file), prototypes were developed that allowed the user to trace the response in the decision tree. The maximum run was approximately 10-12 leaves, with an endpoint indicating a high risk of mortality, low risk of mortality with risk of morbidity, or prolonged recovery from COVID-19 [31]. These prototypes generated a report after each iteration of a user selecting "yes" or "no" in a set of questions linked to a patient's health and biological responses. Results were also obtainable in other forms such as an HTML application that generated a list of results and a chatbot. Upon testing, it became clear that the decision tree used biological and physiological knowledge in deductive reasoning, with some inductive reasoning in the knowledge acquisition. However, rebalancing the decision tree (ie, user response to flow from the top node to leaf nodes that indicate the status of a person's health with severe COVID-19 symptoms) will require additional data sets to allow for reordering the sequence of prompts, reducing or increasing the number of leaves leading to a decision point, and resampling data [25,32]. Furthermore, inductive, or inferential, reasoning is the process of moving from concrete examples to general models; that is, of learning to classify objects by analyzing a set of instances (eg, cases of illness that have already been resolved) whose classes are known [32].

In a previous study, data mining models were developed for the prediction of COVID-19 patients' recovery using an epidemiological data set of COVID-19 patients in South Korea [23]. In that study, a decision tree, a support vector machine, naïve Bayes classifiers, logistic regression, random forest, and a K-nearest neighbor algorithm were applied directly to the data set and a model was developed using Python. The model predicted the age ranges of patients who are at elevated risk of dying from COVID-19, of those who are likely to recover, and of those likely to recover rapidly [23]. The results show that a model developed with a decision-tree algorithm can be most efficient in predicting the probability of recovery for COVID-19 patients.

To improve the tool's inductive power, we included a set of lung conditions that can exacerbate cases of COVID-19 in the upper stratification of the decision tree: asthma, pneumonia, and chronic obstructive pulmonary disease [11]. In the next stratification, we incorporated elevated long-term risk of cardiovascular disease and hyperlipoidemia responses, because these have been linked with acute complications of COVID-19 [33].

Figure 1. Expert system of a COVID-19 decision support web-based (chatbot) tool. There is an important interaction between the knowledge base (controlled) and an interface to display the chatbot to the user with a sequence of questions linked to the stratification of the COVID-19 disease course on an individual basis.

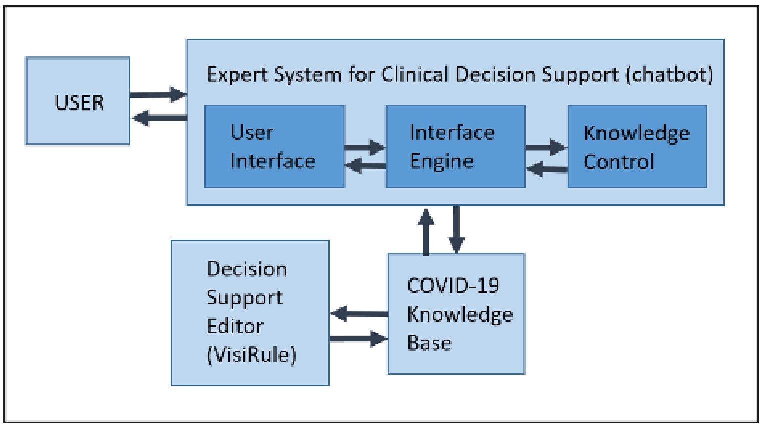


Figure 2. Screenshot from VisiRule 7.021. VisiRule inheritance settings, set as singular "Depth First" with a maximum of 9 from the root node. These settings are important as the complexity of the decision tree framework can be enhanced with the VisiRule applications and its PROLOG backend program.

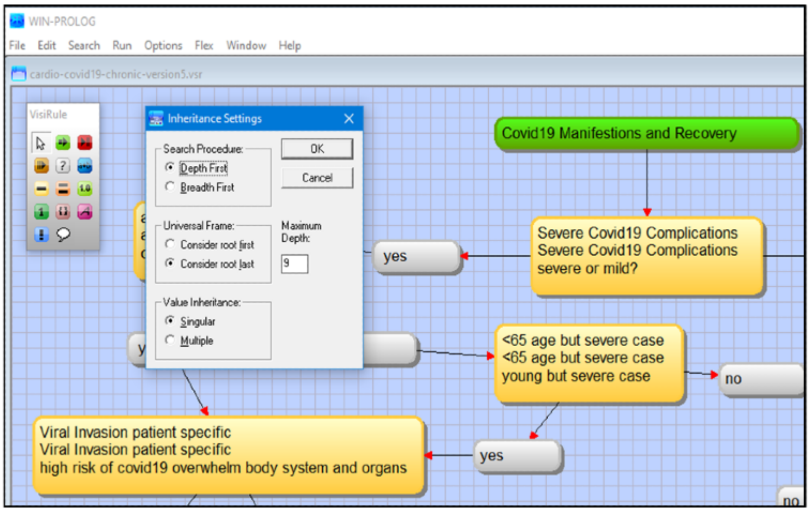


Figure 3. Screenshot from VisiRule 7.021 showing the forward chaining settings. Forward chaining is the default in this application for decision tree induction. As the pop-up display indicates, there are many more settings and customizations to add to the complexity of the forward chaining or traversing through the decision tree framework based on user input across the stratification.

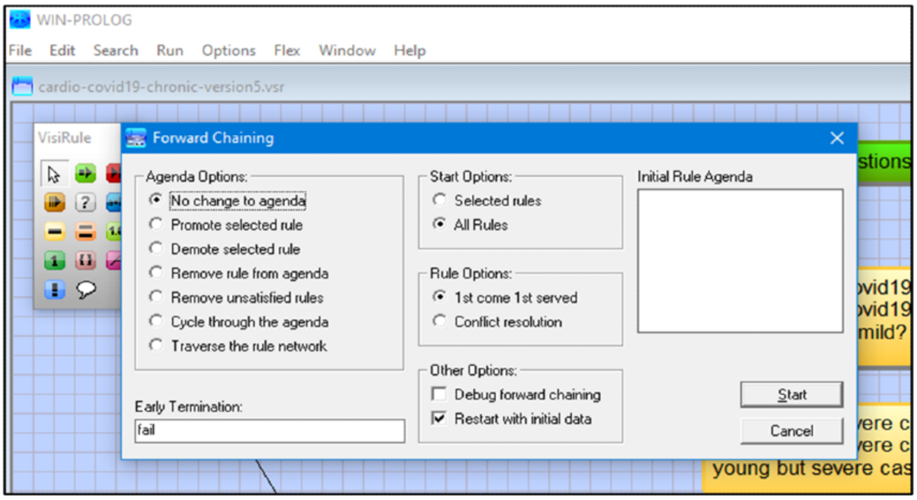


Table 2. Stratified approach to the decision tree hierarchy for severe COVID-19 cases.

Stratification level	Sequence	Description
1	Viral invasion	Suggestive of the level of SARS-CoV-2 invading the cytoplasm of cells and binding to ACE2[a]
2	Age and body systems	Suggestive of higher risk of being aged ≥65 years with impacts on the respiratory and circulatory systems
3	Comorbidities	Suggestive of underlying chronic diseases
4	Manifestations	Suggestive of the level of viral infection and the human body's responses

[a]ACE2: angiotensin-converting enzyme 2

Objectives

The strength of clinical decision-support tools relies on the difficulty of detecting COVID-19 early in its interaction with human biological systems. Because of the high percentage of mild and asymptomatic cases, several pathophysiological responses to the disease need to be fully documented for risk of severity to be adopted [8,19]. Assessing severity, for instance, requires a prominent level of diagnostic accuracy, together with monitoring and periodic reassessment [17,26], all of which are more easily accomplished with the use of a decision-tree tool. In this study, the main objective of the decision-support tool was to diagnose severe cases of COVID-19 based on knowledge using VisiRule control/editor tools (Figure 1). Decision support in both diagnosis and prognosis takes into account a variety of signs and symptoms of COVID-19 on the one hand and the ability of human biological responses to combat the virus on the other hand. This main objective can be broken down into the following subobjectives:

1. Minimize the severity of COVID-19 through early detection via binary classification.
2. Assess the severity of cases with the extent of respiratory and cardiovascular involvement.
3. Reduce inaccuracies in the diagnosis of COVID-19, including both false negatives and false positives calculated by Bayes theorem.
4. Assess the risk of prolonged recovery, morbidity, and mortality.

In the process of achieving these subobjectives, a clear set of clinical guidelines for dealing with COVID-19 will be achieved.

In the diagnosis and treatment of COVID-19, an effective clinical decision-support tool would ensure that best practices are followed (refer to Figure 4, Table 2, and Table 3). Among the benefits provided by a clinical decision-support tool are: (1) incorporating health outcomes of severe COVID-19 cases; (2) considering respiratory and cardiovascular symptoms related to severe COVID-19; (3) correlating COVID-19 infection with other indicators such as obesity, diabetes, blood type, age, and heart and lung complications and illnesses; (4) covering a wide gamut of human phytopathology issues relevant to severe COVID-19; (5) separating mild COVID-19 cases from severe COVID-19 cases (Figure 5); and (6) predicting the risk of mortality.

The diagnostic and risk stratification used in a clinical decision-support tool can also be updated as knowledge of health outcomes and treatments are validated.

Figure 4. List of parameters for a decision-support (expert system) tool developed using Protégé 5.5.0. This image was extracted from the application showing the mapped stratifications of comorbidities, manifestations, heart and lung, and mortality and recovery.

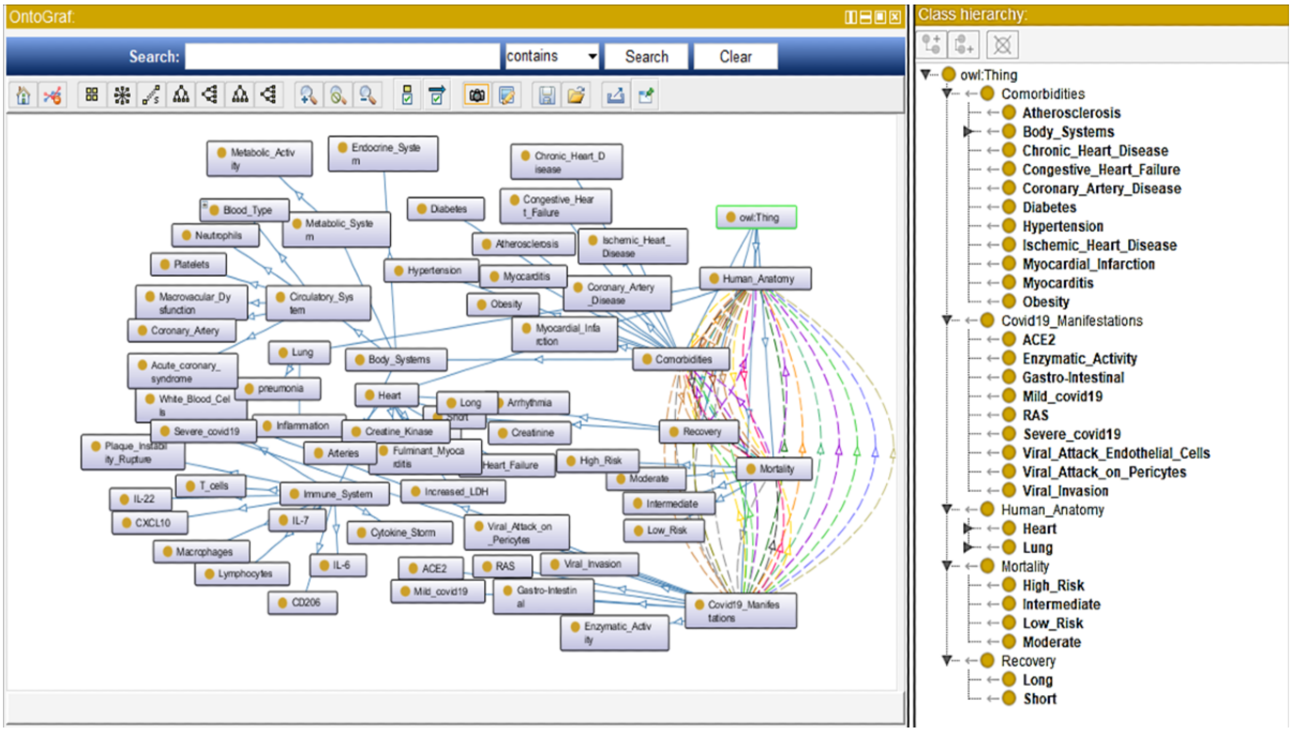
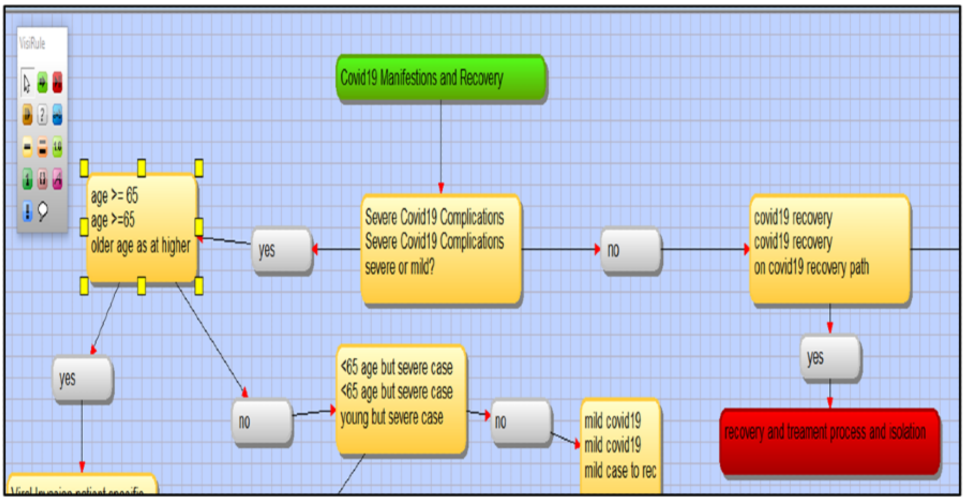


Table 3. Risk factors for severe cases of COVID-19.

Risk status	Relative risk (%)
High risk	
Personal history of multiple comorbidities, including hypertension	35
Previous lung cancer or chronic pneumonia	17
Immune deficiencies	5 to 15
Personal history of cardiovascular diseases	9 to 10
History of heart issues such as arrhythmia	8
Immunosuppression	6 to 8
Moderate risk	
Asthma and chronic pulmonary obstructive disease	4.9 to 7.3
Slight inflammation	3.0 to 5.4
Issues with coronary artery and microvessels	5.4
Intermediate risk	
Asthma without chronic conditions but could have pneumonia or bronchitis	3.8
Asthma, hypertension, and slight inflammation	3.0
Age <65 years with microvessel and coronary artery issues	2.2
Low risk	
Age <65 years	0.5
Normal level of SARS-CoV-2 replication for 10 days and testing negative for COVID-19	0

Figure 5. Screenshot from VisiRule 7.021 showing the separation of mild and severe COVID-19 cases. This display shows the start (in green), the question prompts (in yellow), and endpoint (in red).



Decision Tree Architecture

Overall Design

Architectural construction of the decision tree had 212 yes/no questions, which integrate a range of immune, cardiovascular, and other biological responses to COVID-19 infection. We ran a script to generate a truth table with 990 rows and 64 columns, giving us 63,360 possible combinations of contracted COVID-19 cases. It has 48 end points in total, including 18 long recovery, 13 deaths, 10 chronic disease onsets with recovery, and 7 recoveries. The main knowledge details of the 959 lines of code are described below.

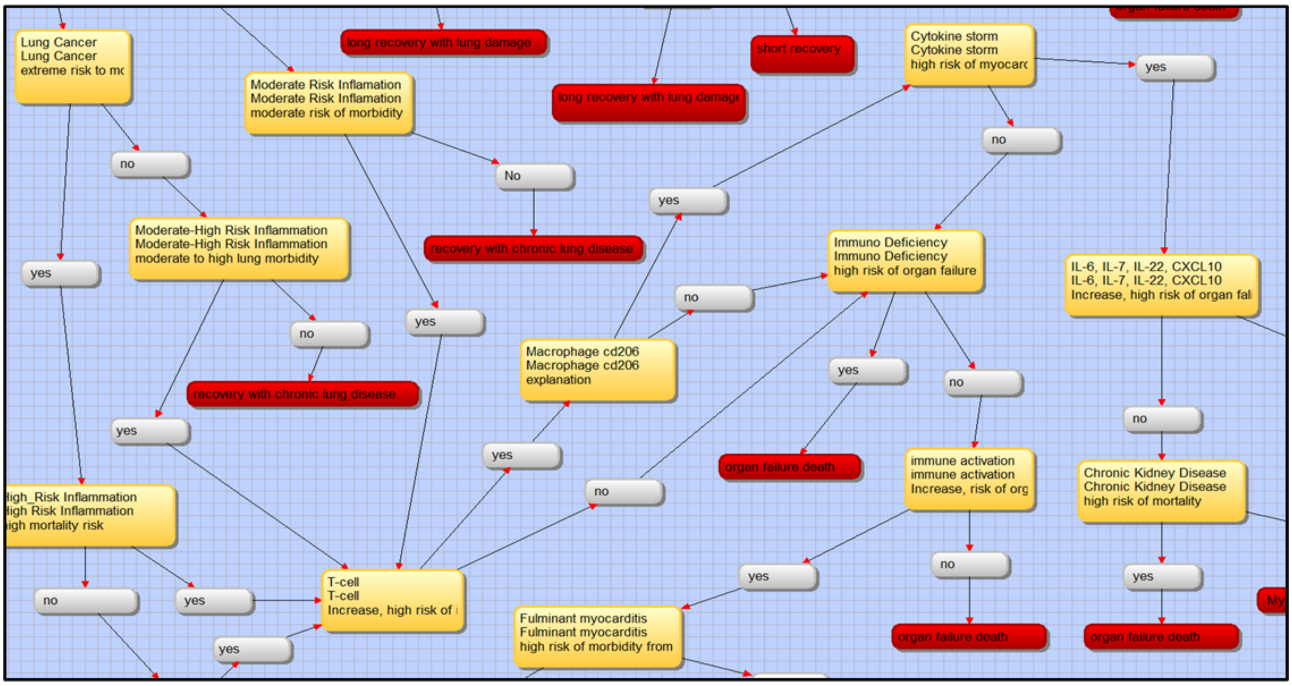
COVID-19 and the Lungs

The host receptor through which SARS-CoV-2 enters cells to trigger infection is angiotensin-converting enzyme 2 (ACE2), which is expressed in the lungs, heart, and blood vessels. This process facilitates entry of the virus into the alveolar epithelial cells within the cytoplasm of the host's skin. The viral RNA then starts to replicate, followed by viral shedding, which likely plays a pathogenic role, resulting in severe cases of lung injury and respiratory failure [33].

COVID-19 and the Heart

COVID-19 is primarily a respiratory disease, but many patients also develop cardiovascular disease, including hypertension that can exacerbate the effects of COVID-19 [34-37], and acute cardiac injury [17]. ACE2 is highly expressed in the human heart, blood vessels, and gastrointestinal tract [17]; when COVID-19 infection dysregulates the ACE2 system, we see cardiovascular disease as the result. In a study of 416 COVID-19 patients, 57 of whom died of the illness, Shi et al [38] found that cardiac injury was common. Therefore, cardiac status and influences of COVID-19 should be part of the decision tree to determine the severity of the potential effects of COVID-19 on that person or patient. COVID-19 infections are also likely associated with infection-induced myocarditis and ischemia [36]. Elevated troponin T levels due to cardiac injury have been associated with a significant rise in mortality [33-37]. A cytokine storm, resulting from a combination of T-cell activation and dysregulated release of interleukin (IL)-6, IL-17, and other cytokines, may also contribute to cardiovascular disease in COVID-19 cases [17]. It is possible that activated T cells and macrophages may infiltrate the infected myocardium, resulting in the development of fulminant myocarditis and severe cardiac damage (Figure 6) [35,36]. This condition eventuality will also be fully integrated in the decision-support tool, with several linkages in the decision tree between cytokine storms, arrhythmias, microvascular dysfunction, and acute coronary syndrome (Figure 6) [35]. Cardiovascular disease may be a primary phenomenon in COVID-19, but it may also be secondary to acute lung injury, which increases the cardiac workload, a condition that is especially problematic in patients with congestive heart failure [21].

Figure 6. Screenshot from VisiRule 7.021 showing lung inflammation to T-cell activation and cytokine storm. This display shows the transition of question prompts (in yellow) from lung to heart conditions, including moderate to high risk, inflammation to immune deficiency, cytokine storm, and chronic kidney disease with binary responses (yes or no) toward the endpoint (in red).



COVID-19 and Comorbidities

Comorbidities identified in COVID-19 studies include chronic cardiac disease, chronic respiratory disease, chronic renal disease (estimated glomerular filtration rate ≤30), mild to severe liver disease, dementia, chronic neurological conditions, connective tissue disease, diabetes mellitus, and various malignancies [2,26]. Clinician-defined obesity is also classified as a comorbidity owing to its probable association with adverse outcomes in patients with COVID-19 in New York City [26].

COVID-19 and the Immune System

Immune system activation may result in plaque instability, increasing the risk of acute coronary events such as stroke [21]. Additionally, research indicates that COVID-19 positivity is associated with lymphopenia (ie, abnormally low levels of white cells in the blood), damage to the liver and muscle tissue, and significantly increased levels of CRP [39,40].

Limitations and Future Work

All of this knowledge from the research literature was incorporated into the flow of questions in the decision tree to form the COVID-19 clinical decision-support tool. However, there are a number of smaller limitations that limit the usefulness of the present design. Gender was not included as a variable in the decision tree; however, the effects of COVID-19 are covered in biological responses with underlying health issues of hypertension and coronary heart disease. COVID-19 patients are more likely to be male than female, and to have more comorbidities such as hypertension and coronary heart disease [9,24,41]. Men have a higher risk of cardiovascular complications than women. In addition, no geographical data were included and no sensitivity analysis was applied.

Another limitation of our work is that the decision-support tool was not based on data that we collected; rather, we made decisions for its construction based on research published during the pandemic. Information about the impact of COVID-19 on the human body is rapidly changing as the pandemic unfolds. For example, new findings indicate that the blood tests used in diagnosing congestive heart failure may also help to indicate severe cases of COVID-19 [39].

The decision tree is not considered to be very deep and would take less than 1 minute for a user to derive a result, status, or event based on certain health outcomes. I conducted stratification of 6-7 leaves to include the relevant parameters for the diagnosis and prognosis of COVID-19 [25]. However, the decision tree used in this decision-support tool did not set a threshold on the number of leaves per node, and instead added knowledge structures of COVID-19 with human anatomy and biology.

The use of clinical decision-support tools that incorporate decision trees is subject to an inherent limitation: the lack of circular references. Best practice for the diagnosis of severe cases of COVID-19 with comorbidities requires circular references and feedback loops in a system-dynamic approach, especially regarding pathophysiology. However, circular references and feedback loops are not possible in decision trees, which makes the application less dynamic to a real-time diagnosis.

The question remains: how would this decision tree framework and use of chatbot to assess COVID-19 integrate within the clinical setting? The answer relies on the inductive power of the chatbot to improve the sequence of the stratification of the level of viral infection, which can be assessed in a clinical setting [1], and the biological responses such as risk assessment of the human cardiovascular system (lung to heart), as Guzik et al [16] stated that stratification from COVID-19 severity increases from the lung to the heart. Furthermore, Knight et al [19] stated that no model can predict the outbreak and spread of COVID-19 in the population; however, there are indications on the duration of COVID-19 infection in individuals of populations that can range from mild to severe. Therefore, a decision tree framework that stratifies the impact of COVID-19 (on an individual basis) can contribute to a clinical setting in real time dealing with patients in the hospital compared to running reports and models (eg, “COVID-19 vulnerability index” related to hospital admissions [7,12]) that can be broadcasted to the public and influence policies such as travel, masks, and distance between people in a hospital rather than clinical patient care [13-19].

Another limitation is that the application of inference probabilities of Bayes theorem to the leaves of the decision tree were not interactive and static. There are uncertainties in the health outcomes of patients with severe COVID-19; therefore, the application of Bayes probabilities is important. Since we could not automate the construction of a decision tree from COVID-19 data, we had no method to automate a valid application of Bayes probabilities for the risk of mortality versus recovery. The structure of decision trees allows for knowledge acquisition and application of inference probabilities, although this framework cannot be clinically validated at this time.

Moreover, if the decision tree is primarily ontology-based on its binary classification, then the probabilities of COVID-19 could become more accurate and plausible based on medical conditions such as comorbidities. Khan et al [42] demonstrated a possible way to integrate a trained data set (using WEKA, MATLAB) and then integrate the data set with the ontology of relationships (via the Protégé application) to establish an ontology-based decision tree model. This method could be applied to our COVID-19 decision tree and could integrate ontology in its stratification, which would reduce the number of scenarios from 63,360 to a much lower number; the scenarios would have to incorporate ontology rules such as age and body system with comorbidities in a more succinct manner. In turn, a data set could be simulated and ratified toward the type of data that need to be collected to form a similar decision tree for accurate binary classification of stratified severe COVID-19 cases.

Finally, the use of decision trees makes it difficult to control for certain biases and overlapping. Decision trees use induction-to-deduction algorithms that range from traditional heuristic-based techniques to more recent hybrid data-to-tree approaches. These algorithms are essential in constructing a sequence of questions that flow from one to the next. For this reason, the basic features of the decision-support tool mitigate against the quantification of any inherent biases. For example, the stratification could be biased by the decision to design a sequence from chronic lung conditions to the heart instead of to gastrointestinal illnesses. This places an emphasis on knowledge of the cardiovascular system while ignoring underlying conditions affecting other body systems.

Conclusion

In conclusion, a decision tree with stratification of COVID-19 effects on biological systems is important knowledge to prototype and simulate. Additionally, stratification of the human physiology within the decision tree proved to indicate that the questions of the health and status of the person with COVID-19 would result in an appropriate summary or list of conditions that are involved in clinical decision support in a specific sequence of events.

Acknowledgments

The author thanks Mr. Hamid Zamani for his contribution of many interesting discussions and review of the framework and theory of a chatbot for severe COVID-19 cases. Additionally, the author thanks and acknowledges Mr. Clive Spenser and Mr. Alan Westwood for their technical expertise and insight in designing chatbots with special emphasis on using VisiRule.

Conflicts of Interest

None declared.

References

1. Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, et al. Virological assessment of hospitalized patients with COVID-2019. Nature 2020 May;581(7809):465-469. [doi: 10.1038/s41586-020-2196-x] [Medline: 32235945]
2. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. Lancet Respir Med 2020 May;8(5):475-481 [FREE Full text] [doi: 10.1016/S2213-2600(20)30079-5] [Medline: 32105632]
3. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534 [FREE Full text] [doi: 10.1016/S1473-3099(20)30120-1] [Medline: 32087114]
4. COVID-19 Dashboard. Johns Hopkins University and Medicine Coronavirus Research Center. 2020. URL: https://coronavirus.jhu.edu/map.html [accessed 2020-11-12]
5. Coronavirus disease (COVID-19). Government of Canada. URL: https://www.canada.ca/en/public-health/services/diseases/coronavirus-disease-covid-19.html [accessed 2020-11-12]
6. van Doremalen N, Bushmaker T, Morris DH, Holbrook MG, Gamble A, Williamson BN, et al. Aerosol and surface stability of SARS-CoV-2 as compared with SARS-CoV-1. N Engl J Med 2020 Apr 16;382(16):1564-1567 [FREE Full text] [doi: 10.1056/NEJMc2004973] [Medline: 32182409]
7. Ai T, Yang Z, Hou H, Zhan C, Chen C, Lv W, et al. Correlation of chest CT and RT-PCR testing for coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases. Radiology 2020 Aug;296(2):E32-E40 [FREE Full text] [doi: 10.1148/radiol.2020200642] [Medline: 32101510]
8. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ 2020 Apr 07;369:m1328 [FREE Full text] [doi: 10.1136/bmj.m1328] [Medline: 32265220]
9. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020 Mar 28;395(10229):1054-1062 [FREE Full text] [doi: 10.1016/S0140-6736(20)30566-3] [Medline: 32171076]
10. Roozenbeek J, Schneider CR, Dryhurst S, Kerr J, Freeman ALJ, Recchia G, et al. Susceptibility to misinformation about COVID-19 around the world. R Soc Open Sci 2020 Oct 14;7(10):201199 [FREE Full text] [doi: 10.1098/rsos.201199] [Medline: 33204475]
11. Driggin E, Madhavan MV, Bikdeli B, Chuich T, Laracy J, Biondi-Zoccai G, et al. Cardiovascular considerations for patients, health care workers, and health systems during the COVID-19 pandemic. J Am Coll Cardiol 2020 May 12;75(18):2352-2371 [FREE Full text] [doi: 10.1016/j.jacc.2020.03.031] [Medline: 32201335]
12. Moutounet-Cartan PGB. Deep convolutional neural networks to diagnose COVID-19 and other pneumonia diseases from posteroanterior chest X-rays. arXiv. 2020 May 02. URL: https://arxiv.org/abs/2005.00845 [accessed 2020-11-01]
13. Gao Y, Cai GY, Fang W, Li HY, Wang SY, Chen L, et al. Machine learning based early warning system enables accurate mortality risk prediction for COVID-19. Nat Commun 2020 Oct 06;11(1):5033. [doi: 10.1038/s41467-020-18684-2] [Medline: 33024092]
14. Diaz-Quijano FA, da Silva JMN, Ganem F, Oliveira S, Vesga-Varela AL, Croda J. A model to predict SARS-CoV-2 infection based on the first three-month surveillance data in Brazil. Trop Med Int Health 2020 Nov 07;25(11):1385-1394 [FREE Full text] [doi: 10.1111/tmi.13476] [Medline: 32790891]
15. Brinati D, Campagner A, Ferrari D, Locatelli M, Banfi G, Cabitza F. Detection of COVID-19 infection from routine blood exams with machine learning: a feasibility study. J Med Syst 2020 Jul 01;44(8):135 [FREE Full text] [doi: 10.1007/s10916-020-01597-4] [Medline: 32607737]
16. Guzik TJ, Mohiddin SA, Dimarco A, Patel V, Savvatis K, Marelli-Berg FM, et al. COVID-19 and the cardiovascular system: implications for risk assessment, diagnosis, and treatment options. Cardiovasc Res 2020 Aug 01;116(10):1666-1687 [FREE Full text] [doi: 10.1093/cvr/cvaa106] [Medline: 32352535]
17. McRae MP, Dapkins IP, Sharif I, Anderman J, Fenyo D, Sinokrot O, et al. Managing COVID-19 with a clinical decision support tool in a community health network: algorithm development and validation. J Med Internet Res 2020 Aug 24;22(8):e22033 [FREE Full text] [doi: 10.2196/22033] [Medline: 32750010]
18. Xiong TY, Redwood S, Prendergast B, Chen M. Coronaviruses and the cardiovascular system: acute and long-term implications. Eur Heart J 2020 May 14;41(19):1798-1800 [FREE Full text] [doi: 10.1093/eurheartj/ehaa231] [Medline: 32186331]
19. Knight SR, Ho A, Pius R, Buchan I, Carson G, Drake TM, ISARIC4C investigators. Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score. BMJ 2020 Sep 09;370:m3339 [FREE Full text] [doi: 10.1136/bmj.m3339] [Medline: 32907855]
20. Oyelade ON, Ezugwu AE. A case-based reasoning framework for early detection and diagnosis of novel coronavirus. Inform Med Unlocked 2020;20:100395 [FREE Full text] [doi: 10.1016/j.imu.2020.100395] [Medline: 32835080]
21. Madjid M, Safavi-Naeini P, Solomon SD, Vardeny O. Potential effects of coronaviruses on the cardiovascular system: a review. JAMA Cardiol 2020 Jul 01;5(7):831-840. [doi: 10.1001/jamacardio.2020.1286] [Medline: 32219363]
22. Wu Y. Compensation of ACE2 function for possible clinical management of 2019-nCoV-induced acute lung injury. Virol Sin 2020 Jun;35(3):256-258 [FREE Full text] [doi: 10.1007/s12250-020-00205-6] [Medline: 32034638]

23. Muhammad LJ, Islam MM, Usman SS, Ayon SI. Predictive data mining models for novel coronavirus (COVID-19) infected patients' recovery. SN Comput Sci 2020;1(4):206 [FREE Full text] [doi: 10.1007/s42979-020-00216-w] [Medline: 33063049]
24. Flores D, Walter J, Wussler D, Kozhuharov N, Nowak A, Dinort J, et al. Direct comparison of high-sensitivity cardiac troponin T and I for prediction of mortality in patients with pneumonia. J Clin Chem Lab Med 2019;2(2):1000131 [FREE Full text]
25. Azar AT, El-Metwally SM. Decision tree classifiers for automated medical diagnosis. Neural Comput Appl 2012 Oct 7;23(7-8):2387-2403. [doi: 10.1007/s00521-012-1196-7]
26. Richardson S, Hirsch J, Narasimhan M, Crawford J, McGinn T, Davidson K, the Northwell COVID-19 Research Consortium, et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City area. JAMA 2020 May 26;323(20):2052-2059 [FREE Full text] [doi: 10.1001/jama.2020.6775] [Medline: 32320003]
27. Chrimes D. Medical Expert System Demos (Covid-19). VisiRule Expert System Software. URL: https://www.visirule.co.uk/medical-demos [accessed 2020-10-30]
28. Wiguna W, Riana D. Diagnosis of coronavirus disease 2019 (COVID-19) surveillance using C4.5 algorithm. Pilar 2020 Mar 16;16(1):71-80. [doi: 10.33480/pilar.v16i1.1293]
29. Dialogflow. Google Cloud. URL: https://cloud.google.com/dialogflow [accessed 2020-11-01]
30. Chat + Tasks = Magic. Zenchat. URL: https://zenkit.com/en/chat/ [accessed 2020-11-02]
31. Espinoza J, Crown K, Kulkarni O. A guide to chatbots for COVID-19 screening at pediatric health care facilities. JMIR Public Health Surveill 2020 Apr 30;6(2):e18808 [FREE Full text] [doi: 10.2196/18808] [Medline: 32325425]
32. Albu A. From logical inference to decision trees in medical diagnosis. 2017 Presented at: The 6th IEEE International Conference on E-Health Bioengineering - EHB . Grigore T. Popa University of Medicine and Pharmacy, Sinaia, Romania, June; June 2017; Sinaia, Romania p. 22-24. [doi: 10.1109/ehb.2017.7995362]
33. Santos RAS, Sampaio WO, Alzamora AC, Motta-Santos D, Alenina N, Bader M, et al. The ACE2/angiotensin-(1-7)/MAS axis of the renin-angiotensin system: Focus on angiotensin-(1-7). Physiol Rev 2018 Jan 01;98(1):505-553 [FREE Full text] [doi: 10.1152/physrev.00023.2016] [Medline: 29351514]
34. Guo T, Fan Y, Chen M, Wu X, Zhang L, He T, et al. Cardiovascular implications of fatal outcomes of patients with coronavirus disease 2019 (COVID-19). JAMA Cardiol 2020 Jul 01;5(7):811-818 [FREE Full text] [doi: 10.1001/jamacardio.2020.1017] [Medline: 32219356]
35. Reynolds HR, Adhikari S, Pulgarin C, Troxel AB, Iturrate E, Johnson SB, et al. Renin-angiotensin-aldosterone system inhibitors and risk of Covid-19. N Engl J Med 2020 Jun 18;382(25):2441-2448 [FREE Full text] [doi: 10.1056/NEJMoa2008975] [Medline: 32356628]
36. Roffi M, Patrono C, Collet J, Mueller C, Valgimigli M, Andreotti F, ESC Scientific Document Group. 2015 ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: Task Force for the Management of Acute Coronary Syndromes in Patients Presenting without Persistent ST-Segment Elevation of the European Society of Cardiology (ESC). Eur Heart J 2016 Jan 14;37(3):267-315 [FREE Full text] [doi: 10.1093/eurheartj/ehv320] [Medline: 26320110]
37. Hu H, Ma F, Wei X, Fang Y. Coronavirus fulminant myocarditis treated with glucocorticoid and human immunoglobulin. Eur Heart J 2021 Jan 07;42(2):206 [FREE Full text] [doi: 10.1093/eurheartj/ehaa190] [Medline: 32176300]
38. Shi S, Qin M, Shen B, Cai Y, Liu T, Yang F, et al. Association of cardiac injury with mortality in hospitalized patients with COVID-19 in Wuhan, China. JAMA Cardiol 2020 Jul 01;5(7):802-810 [FREE Full text] [doi: 10.1001/jamacardio.2020.0950] [Medline: 32211816]
39. Siedlinski M, Jozefczuk E, Xu X, Teumer A, Evangelou E, Schnabel RB, et al. White blood cells and blood pressure: a Mendelian randomization study. Circulation 2020 Apr 21;141(16):1307-1317 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.119.045102] [Medline: 32148083]
40. Tan L, Wang Q, Zhang D, Ding J, Huang Q, Tang Y, et al. Lymphopenia predicts disease severity of COVID-19: a descriptive and predictive study. Signal Transduct Target Ther 2020 Mar 27;5(1):33. [doi: 10.1038/s41392-020-0148-4] [Medline: 32296069]
41. Wu C, Chen X, Cai Y, Xia J, Zhou X, Xu S, et al. Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. JAMA Intern Med 2020 Jul 01;180(7):934-943 [FREE Full text] [doi: 10.1001/jamainternmed.2020.0994] [Medline: 32167524]
42. Khan Z, Saeidlou S, Saadat M. Ontology-based decision tree model for prediction in a manufacturing network. Product Manufact Res 2019 May 30;7(1):335-349 [FREE Full text] [doi: 10.1080/21693277.2019.1621228]

Abbreviations

ACE2: angiotensin converting enzyme
CRP: creatine-reactive protein
IL: interleukin
MERS: Middle East Respiratory Syndrome

PMML: Predictive Model Markup Language
SARS: Severe Acute Respiratory Syndrome

Edited by A Mavragani; submitted 07.09.22; peer-reviewed by V Ochs, R Sampieri-Cabrera, H Bangash; comments to author 03.12.22; revised version received 22.12.22; accepted 10.01.23; published 30.01.23.

Please cite as:
Chrimes D
Using Decision Trees as an Expert System for Clinical Decision Support for COVID-19
Interact J Med Res 2023;12:e42540
URL: https://www.i-jmr.org/2023/1/e42540
doi:10.2196/42540
PMID:36645840

Viewpoint

Web-Based Co-design in Health Care: Considerations for Renewed Participation

Maryam Mallakin[1*], MDes; Christina Dery[1*], MDes; Samuel Vaillancourt[2], MD; Sahil Gupta[3], MD; Katherine Sellen[1*], MSc, MRes, PhD

[1]Health Design Studio, OCAD University, Toronto, ON, Canada

[2]Department of Medicine, University of Toronto, Toronto, ON, Canada

[3]Department of Emergency Medicine, St Michael's Hospital, Toronto, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Katherine Sellen, MSc, MRes, PhD
Health Design Studio
OCAD University
42 Parkway Ave
Toronto, ON, M6R 1T5
Canada
Phone: 1 647 448 4902
Email: katherinesellen@gmail.com

Abstract

The COVID-19 pandemic has shifted the work environment to a new reality of remote work and virtual collaboration. This shift has occurred in various work settings with an impact on spaces, approaches, applied techniques, and tools. This has resulted in the broad use of virtual tools in the health care sector to avoid physical encounters and in-person interactions that will likely outlast the COVID-19 pandemic. Developing effective virtual approaches requires the knowledge and skills of using digital technologies collaboratively combined with a deep understanding of the context or contexts in which these approaches may be used. The implementation of virtual health design methods, including web-based co-design, has increased to meet the realities of COVID-19 restrictions and is likely to outlast them. Adapting the use of co-design methodologies to a virtual configuration requires rethinking methods of collaboration and communication, adapting to virtual environments, and creating new methods of engagement and facilitation. With this viewpoint, we reviewed the current work on co-design (in person and web based) to propose techniques for the design, planning, and implementation of web-based co-design. We propose 7 considerations that may enable web-based co-design projects in the health care sector. The key considerations that affect the success of a web-based co-design approach should be considered in the process of planning, developing, and conducting web-based co-design sessions. These include facilitation, collaboration, accessibility and equity, communication, sensemaking, tangible tools and games, and web-based research ethics. We illustrate this work with a case study of co-design for an emergency department discharge tool developed during the pandemic.

(Interact J Med Res 2023;12:e36765) doi:10.2196/36765

KEYWORDS

web-based design research; co-design; web-based co-design; virtual platform; virtual learning platforms; internet research ethics; collaboration; health communication; sensemaking; health design; tangible tools and games

Introduction

Background

In recent years, co-design methods have been widely applied in health care systems [1]. The application of these methods is rapidly expanding, specifically in the new era of remote work and virtual collaboration owing to the COVID-19 pandemic.

A co-design approach provides an opportunity to share, mobilize, and activate knowledge by engaging patients and health stakeholders in a collaborative research and design process [2]. Co-design is a design-led approach to change, with a set of creative and participatory principles, practices, and tools. Co-design has the potential to be used in many areas of health care such as improving the quality of care and patient experience [3] by drawing on a collaborative and equitable lens that brings health stakeholders and patients together to explore complex problems. Various activities and tools (eg, tangible tools, design games, and play-like activities) are used in co-design processes to support idea generation and foster communication among participants. These in-person activities and tools highlight the exploratory, imaginative, dialogical, and empathic aspects of co-design [4]. Ordinarily, these tools are used directly by participants in person and have not been actively developed for virtual spaces or modes before the COVID-19 pandemic.

The Covid-19 pandemic has resulted in a dramatic shift in the use of digital technologies, the internet, and internet-based services for communication, interaction, and collaboration in all aspects of work and life [5]. Many people now have increased exposure to web-based engagement and collaboration as well as a greater willingness and the required skills to engage in web-based activities [6]. As a result, virtual connections have become more acceptable, and digital engagement opportunities have increased and diversified [6]. The pandemic required many groups to begin working remotely, including design researchers and practitioners who needed to shift to web-based design activities [6]. Although the COVID-19 pandemic has increased the development and application of web-based co-design methods and tools, they are not limited to the pandemic and can be used as an extension to co-design practices even after the pandemic.

Collaborative activities and tools for both in-person and web-based co-design are central to facilitating meaningful and productive engagement and collaborative discussions among stakeholders. The use of tangible tools, game-like activities, and scenarios can elicit novel responses on a subject matter because the playfulness of these activities tends to foster creative behavior [7]. Imaginative and pretend play are effective strategies for idea generation and for moving toward mutual or shared understanding. Brown and Vaughan [8] emphasize the importance of "play" or "tinkering" for problem-solving and working with our hands (tangible experiences) to "see solutions" that otherwise would not be seen [7], an aspect largely missing in web-based experiences. In addition, communication is a powerful tool in any game-like activity and plays an essential role in participatory activities such as co-design techniques [8,9]. Although in-person and web-based communication share common characteristics, web-based communication has additional challenges that should be addressed, such as the lack of nonverbal cues that can compromise connection and empathy [10]. Lack of tangible interaction and compromises to connection are just 2 of the several issues that must be addressed in web-based co-design [11]. Although there are challenges, web-based co-design may also afford a broader reach for stakeholder engagement, with the constraints of a physical location and day or time meeting removed as barriers for some.

Objectives

The objective of this viewpoint is to describe a set of 7 considerations that were developed in response to the pandemic, in the design, development, and conduct of web-based co-design, and applied in the health sector.

Methods

In this viewpoint, we have captured considerations for web-based co-design in health care emerging from adaptations in co-design practices undertaken at the Health Design Studio, OCAD University, Canada during the pandemic. These considerations represent a combination of continuous review of the literature for possible adaptation strategies and the experience of adapting a project to test these considerations during the pandemic (illustrated by the case study below). In our review, we aimed to identify from existing and prepandemic studies the main challenges and opportunities for participatory design approaches and co-design activities in a web-based setting, specifically focused on web-based co-design within the health care system. We used an exploratory approach for this literature review, combining insights from research articles with available guidance and resources and other gray literature [12]. In parallel, we explored and made adaptations in an ongoing project (case study), adapting processes and techniques to achieve web-based co-design for a co-design project originally situated within an emergency department (ED). The case study is used in this viewpoint to illustrate adaptations of co-design during a pandemic in context [11,13].

The case study (Designing Discharge after Emergency Care [D.DEC] project [14]) was a parallel activity to experiment with techniques and draw additional insights about the main challenges, opportunities, benefits, and drawbacks of web-based co-design for health. The D.DEC project aimed to develop an improved and appropriate patient-centered approach for discharge information in the ED. As the COVID-19 pandemic emerged, the project had to move its co-design work to web over the duration of the project. As the team working on the D.DEC project, we were in a position to have firsthand experience with the adaptations made to the project and how those adaptations affected the project. We undertook the literature search as we developed adaptations, took notes as we went, and wrote this viewpoint together to capture the knowledge created through this experience. Ethics approval for the D.DEC project was granted by both the design research team's institution and the hospital in which the ED was situated.

Results

Overview

The web-based search identified information from the academic literature as well as relevant toolkits, handbooks, reports, guidelines, webinars, and presentations in the gray literature. Search terms included "co-design," "codesign," "participatory," "participatory design," "participatory design tools," "participatory approach in health," "health communication," "co-design in health," "web-based communication," "virtual learning," "sensemaking," "internet research ethics," "synchronous & asynchronous communication tools," "design games," "virtual play and creativity," and "health, health care," "health sector," "medical," and "web-based design research," "virtual collaboration platforms," "web-based or remote co-design," "guides on digital accessibility," "web-based tools for design thinking," and "internet research ethics."

Two researchers reviewed the abstracts of candidate articles for relevance to the aim of the literature review. Three researchers read each article for a final set of 52 peer-reviewed articles for inclusion in the review. The gray literature, in which design-related practice is more often documented, included a search for co-design toolkits (n=2), co-design handbooks (n=2), collaboration challenge (in-person or web-based) reports (n=4), facilitating virtual meeting guidelines (n=3), and webinar guidance (n=2). We retrieved resources from 15 organizations and institution websites, specifically design organizations and design in health groups.

We used a virtual meeting platform (Zoom, Zoom Video Communications) and a web-based visual collaboration platform (Miro) to conduct remote collaborative teamwork during the research process. The Miro web-based whiteboard was used to house articles and resources and share and study our findings, enabling us to explore and experience some of the challenges and opportunities of a web-based collaborative setting.

Case Study

The D.DEC project [15] was carried out in the ED in an urban center and focused on discharge information as a key opportunity to improve patient outcomes beyond the care provided in the ED. As stated above, the D.DEC project was one of the several projects undertaken by the Health Design Studio at Ontario College of Art and Design University when the pandemic hit. The project's co-design approach was intended to bring diverse stakeholders together to identify a creative solution for developing a precise, feasible, sustainable, and patient-centered tool. Initially structured as an in-person project with 2 large multistakeholder workshops centering on the co-design process, the actual process consisted of the following eight steps: (1) review of existing research, (2) intensive observations, (3) collaborative synthesis, (4) co-design sessions 1a and b, (5) web-based feedback, (6) co-design session 2a and b, (7) prototype implementation with feedback and testing, and (8) refinement of the solution. The project became an opportunity to experiment with adaptations in the co-design process and techniques through web-based means.

Shifting to a web-based co-design mode was an opportunity for the project team to learn about the challenges, gaps, and potential opportunities in making that shift. In total, 25 stakeholders participated in co-design sessions 1a and b and 2a and b, including patients, emergency physicians, emergency nurses, and family physicians, with the aim of discussing, identifying, and developing an improved method of discharge information in the ED setting. The first round of web-based co-design sessions (1a and b) focused on the discharge information delivery process and identifying design needs, and the second round of co-design sessions (2a and b) reviewed and refined possible design solutions. To support the delivery of these virtual sessions, Zoom was used as a web-based meeting platform, Google slides was used for sharing screens and documents, and Google Docs was used to take meeting notes together. The project team identified two main challenges in conducting the web-based co-design sessions:

1. Technical challenges and capabilities related to using web-based platforms
2. The level of participation, collaboration, and interaction from participants, which led to mostly conversations around existing knowledge instead of generative ideation and creating new knowledge.

The design of a web-based co-design process, specifically addressing health-related interventions, requires strong knowledge of the context and skills in problem formulation, shared processes, problem-solving, and collaborative solutioning. An appropriate web-based system is required to support a practical collaborative space with a variety of participation opportunities and tools [16].

Conducting and facilitating a web-based co-design process consists of 3 main phases [17]: the preworkshop phase (planning process), the workshop phase (conducting process), and the postworkshop phase (data analysis and evaluation process).

Well-planned virtual processes rely on applying the right combination of web-based or offline and synchronous or asynchronous tools to enable opportunities for both facilitators and participants to feel empowered during the co-design process. Providing combinations and alternatives supports participants in managing their time, space, and feedback and enables facilitators to analyze the outcomes and adjust their methods and agenda throughout the process [18].

Seven Considerations for Web-Based Co-design in Health

Overview

Through our scoping review and exploration of the D.DEC case study, we collated insights about various cocreation elements and the importance of selecting appropriate collaboration tools to improve participation, discussion, and ideation among participants. We identified 7 factors that affected participants' engagement and collaboration in a web-based co-design process (Textbox 1).

Textbox 1. Factors affecting participants' engagement and collaboration in a web-based co-design process.

1. Facilitation
2. Collaboration
3. Accessibility and equity
4. Communication
5. Sensemaking
6. Tangible tools and games
7. Web-based research ethics

The following section includes brief descriptions of these 7 factors, how each of these factors is reflected in the D.DEC project, and recommendations based on insights from the academic and gray literature.

Facilitation

Facilitation plays a critical role in providing meaningful cocreative opportunities (structure and space) in co-design to guide participants through sessions and ultimately plays a central role in facilitating the uncovering of new data and insights from participants. Facilitation requires expertise, resources and preparation for planning engagement, prioritizing tools, and exploring creative solutions [19].

It is more challenging to build an effective, creative, and encouraging collaboration in a virtual setting, making facilitator roles even more important than in-person settings.

In the planning and preparation phase of the D.DEC web-based co-design sessions, we thoughtfully developed a facilitator script to ensure that the facilitators created a collaborative and welcoming space. The script addressed ethics, accessibility, digital literacy, tone, language, and the specific activities and processes of the co-design sessions. The creation of the script served both the training and evaluation roles. We conducted a test web-based co-design session with participants naive to the project to refine materials and scripts, develop facilitators' comfort level with web-based delivery, communication channels, and facilitation of collaboration tools and materials. Two facilitators took turns to either facilitate dialogue and collaborative creation or manage technology, materials, and multiple communication channels.

Facilitators play a central role in web-based co-design to provide well-planned and focused processes, ensure equal opportunity for contribution from all participants, and keep participants motivated and focused [17]. Facilitation tasks in web-based co-design can be divided into 5 main categories: methodology, technical administration, content, user interaction, and results [17]. The facilitator's role includes deciding the topic, setting the context, planning rules and the agenda, inviting participants, providing access and motivation, managing and adapting technical considerations, stimulating discussions, identifying new topics, recognizing contributions, maintaining participation, and wrapping up. Having a trained and prepared facilitation team (consisting of the main facilitator, cofacilitators, and technical assistant for technical administration and support) and a well-planned process play an essential role in conducting a successful web-based co-design process [17].

Collaboration

Collaborating and designing with participants is the main component and central activity of a design-based research process. Collaboration within a co-design process includes communication, cooperation, and cocreation [10]. The original D.DEC project was structured in a 2-tiered manner with a core collaborative team including designers and physician leads with 2 large multistakeholder workshops. The workshops were intended to serve multiple purposes, including socializing the project with decision makers and eliciting design considerations from a broad range of roles in the ED. This approach was chosen in part owing to restrictions on in-person involvement in activities for ED staff, because a large meeting sanctioned by department decision makers was one mechanism for enabling unionized and highly time-constrained staff involvement. The pandemic hit just before the first workshop of 35 participants.

Given the variability and constraints of availability among the stakeholder groups (patients, emergency physicians, nurses, clerks, and family physicians) during the pandemic, the team decided to provide opportunities for feedback from frontline staff using web-based surveys between smaller and more frequent co-design sessions. This was deliberately intended to maintain a broader collaboration by providing access to otherwise unavailable stakeholders. Facilitation in the co-design session included deliberate focusing and privileging of diverse voices (focusing on roles other than emergency physicians who were already central to the project). In switching to a web-based engagement strategy, one of the challenges identified in the D.DEC web-based workshop was participant capabilities related to the use of the selected platform, which affected their engagement and collaboration in discussions.

There are various factors that affect participants' collaboration in a web-based co-design process in a health context such as an ED, including (1) selecting the right platform, programs, and tools that are appropriate for engaging participants in selected activities and tasks, specifically selecting tools that are equity positive by requiring as little technological proficiency as possible and the least sophisticated equipment as possible; (2) initiating initial interaction with participants to foster better relationships by establishing trust, connection, and commitment; (3) removing time constraints and planning shorter and more frequent sessions (eg, multiday engagement activities) to provide more flexible engagement for health care staff and patients, and providing facilitators the opportunity to analyze data and adjust agendas and activities to support continued contributions; (4) planning the right combination of synchronous or asynchronous activities and tasks to provide time and space for participants to manage their ideas; (5) including trained facilitators to avoid biased discussions and discussion breakdown, and providing better opportunity for participants to form shared understanding and commitment to the project's goals; (6) applying techniques and activities that are interactive, understandable, pleasing, and engaging; and (7) dividing participants into smaller groups by topic and logistics (eg, who is technology savvy? Who is comfortable with the camera? Which health care roles experience power dynamics that might restrict their voice) [10].

Accessibility and Equity

Although web-based research methods are rapidly expanding (accelerated by the COVID-19 pandemic) and becoming commonplace in health research, there are various challenges that affect their effectiveness. Access to digital technology and equitable resources, including time and space to participate, can affect the ability to maintain participation across stakeholders.

The main challenges identified through the D.DEC web-based co-design process include the distribution and access to digital resources (eg, hardware, software, and internet), level of skills required for participation (eg, literacy level and familiarity with computer technology and programs), and privacy and security considerations (eg, availability of staff spaces, family, culture, anonymity, and confidentiality) [20]. For some patients, digital video was not an option, and a physical version of the materials was mailed to their home, and participation by phone was made available. Flexibility and facilitation were key to maintaining access to the session and participation at the same level as other participants.

Accessibility requirements for patients with disabilities or impairments should be included in the guidelines for selecting platforms and activities that can fully support various needs such as hearing, vision, or speech impairments. It is important to select platforms that are compatible with assistive technologies (eg, screen readers) and accessible for people who are deaf or hard of hearing, blind or visually impaired, have sensory disabilities, and have intellectual or developmental disabilities [21]. In each collaborative case, the accessibility features of the platforms should be evaluated to establish sufficiency of access for the specific case and its participants. Among the various available platforms, Zoom and Microsoft Teams are 2 platforms that provide more inclusive (although, not entirely) virtual accessibility features for collaborative activities [22]. Providing a paper version of materials, asynchronous participation, or offline participation option should also be considered, especially when vulnerable groups and issues of health care access are central to the co-design project.

Trauma-Informed Practices

Remembering or recounting negative or harmful experiences in health care spaces may be traumatizing for participants. Addressing trauma is an important component that should be considered to maximize safety, accessibility, and equity in a web-based co-design process. Trauma-informed care is a strengths-based approach "that is grounded in an understanding of and responsiveness to the impact of trauma, that emphasizes physical, psychological, and emotional safety for both providers and survivors, and that creates opportunities for survivors to rebuild a sense of control and empowerment" [23]. The integration of trauma-informed care principles is critical for fostering more accessible and safe spaces when hosting virtual meetings (eg, web-based co-design). The 6 key principles of a trauma-informed approach are as follows: emotional and physical safety; cultural, historical, and gender considerations; trustworthiness and transparency; peer support and mutual self-help; collaboration and mutuality; and empowerment, voice, and choice [24].

Communication

Communication is a powerful tool that plays an essential role in participatory techniques. Communication should be accessible, inclusive, and generate shared understanding and empathy among designers, researchers, and stakeholder groups [9].

The initial co-design activities for the D.DEC project included a large-scale workshop of 20 to 35 people across various roles within the ED, with the addition of patients and family physicians. Larger-scale co-design activities are effective mechanisms for expressing complex issues and building empathy and understanding across disparate health care stakeholders. However, in the context of web-based co-design, we chose to reduce the scale, opting for 1 to 2 stakeholder representatives per stakeholder group per session. In this manner, the affordances of the Zoom platform, equal visual representation and single person auditory focus, mute function, and hand raising, could be used to reduce existing power dynamics that would otherwise have affected communication balance. We also chose to use previously crafted patient stories (video based) to communicate the focus of the co-design session, centering around patient voices from the beginning.

Effective communication requires an in-depth understanding of the context, priorities, needs, beliefs, environment, social norms, and preferences of the intended audiences. Communication in co-design creates consensus and ownership of the process and its outcomes [9]. Virtual and in-person communication may share common principles and motivating factors to enhance participants' engagement, but they require different implementation paths. The main communication challenges in virtual settings that may lead to confusion and misunderstanding include (1) lack of nonverbal cues such as body language, facial expression, and eye contact; (2) lack of strong connection, empathy, and trust among participants; and (3) lack of control over the process (for participants) [10].

To address these challenges, the study by McCarthy et al [12] suggests using multimedia platforms such as web-based meeting platforms (Zoom, Microsoft Team, etc) that support the use of nonverbal cues such as tone of voice, eye contact, or facial expression and dividing participants into smaller groups (5 people) to provide better opportunities for connection. Applying methods and tools such as storytelling, storyboards, and scenarios can help improve connections among participants to enhance emotional reciprocity, shared understanding, trust, and empathy [25].

Sensemaking

During in-person co-design, materials and interactions affect sensemaking. When co-design is conducted through web, the technical aspects of the experience play a role in the sensemaking process. Technical sensemaking refers to participants' interactions with technology that could be challenging, such as how to use technology or how to handle technological failures (eg, video freezing or audio cutting out). Interpersonal sensemaking refers to participants' interactions with other participants, which could be challenging owing to a lack of motivation or communication gaps (eg, lack of interpersonal feedback with those who are trying to communicate). There is a shortcoming in the literature about sensemaking for virtual environments that might otherwise point to adaptations that can be made to facilitate the move to web-based co-design. Story and narrative are strong sensemaking tools that play important roles in information sharing, collective interpretation of problems, and improving communication [26].

Supporting sensemaking for the D.DEC co-design sessions included carefully created visual support for each co-design subactivity and a highly structured co-design workshop agenda and script. This enabled open dialogue around specific aspects of the design process, shared sensemaking on gaps, and provided possible solutions for discrete aspects of the co-design of the discharge tool. We used filmed, re-enacted patient stories to try to bring about a sense of empathy. We also included multiple check-in points and verbal reiterations of insights and ideas, conclusions, and suggestions at each subactivity by the facilitator.

The existing literature largely focuses on addressing our understanding of sensemaking for in-person contexts [27]. Amber and Jorgen [27] point to unique aspects of the virtual environment that affect the sensemaking process, including (1) partial presence that may limit the capacity to detect opportunities for interaction and sensemaking; (2) concurrent states of being "in" and "out," which means the participants can be in 2 places at once, (ie, at their PC in their home or office and in the virtual space), which can add a new dimension and complexity to sensemaking in virtual environment owing to moving "in" and "out" of the virtual world; (3) disembodiment, which means subconscious, physical cues that are normally used to communicate with one another and make sense are lost in virtual settings; and (4) no known etiquette or norms in virtual environmental interactions that refer to the set of rules and norms that are followed in real-world interaction, which are ambiguous or nonexistent in virtual settings.

Tangible Tools and Games

A participatory design approach typically includes a variety of techniques and tools to engage participants in collaborative discussion and co-design, depending on the topic of the research, types of participants, and circumstances under which the research is conducted [17]. Tools are the material components that are used to connect design and research practices in the participatory co-design process [28] and can include probes, tangible tools, and games.

Probes are participatory design tools that often consist of material objects (eg, disposable cameras, postcards, stickers, maps, and art materials). Probes are often exploratory in nature and are intended to enhance dialogue and invite participants to be involved in different phases of the exploratory design process, including (1) probing knowledge and meaning, (2) provoking reflections, (3) projecting visions or ideas into the future, and (4) prototyping ideas and concepts that envisage future reality [29].

Tangible tools include visual tools such as graphic representations and artifacts and generative tools such as scenario boards, storyboards, videos, and collages [28]. Tangible tools are intended to enable collaborative, innovative, and active dialogue among participants in the design process. Tangible tools are defined as materials used in participatory design activities to facilitate knowledge exchange; shared understanding; and generating ideas among participants through making, telling, and enacting approaches.

Design games are generative, visual, and playful tools used to transfer knowledge and ideas and generate new ideas and insights into the co-design process, often including shared decision-making mechanisms. Design games have various applications in collaborative and participatory processes, including supporting creative thinking, engaging and empowering participants in an exploratory and human-centered design process, enhancing social collaboration, and understanding individual participants' experiences [30]. The exploratory design game framework by Brandt [31] takes advantage of the various skills and expertise of the participants to generate new ideas and design possibilities in the participatory process. The framework consists of various exploratory design games, including games to conceptualize designing, "exchange perspective," design games, negotiation and workflow-oriented design games, and scenario-oriented design games [31].

During the D.DEC project, when considering the design of participatory engagement as the team moved toward prototyping discharge communication options, it became necessary to develop activities that would engage a range of stakeholders in exploring new ideas about communication techniques and the content of discharge communication. Given the diverse technology access and acceptance among participants, the team chose to create tangible materials to represent different potential design options and to physically send them to all participants. These paper-based materials included prompting steps and options for participants to contribute during the session both verbally and by making notes or using stickers to provide feedback on each design option.

Many tangible or probing tools can be used in a web-based co-design approach to engage participants in web-based collaboration and cocreation (telling, making, and enacting) [32]. The following tools can be used in a web-based co-design process:

- Visual tools: sketches, diagrams, visual and graphic representations, and video
- Generative tools:
  - Telling: stories, storyboarding, self-observation (photo taking, short video, and drawing), diaries, voting, stickers, sorting, and categorizing to prioritize ideas
  - Making: 2D collages and 2D mapping
  - Prototyping apps (eg, Boards, Mockingbird, and Pop)
  - Enacting: scenario making, participatory envisioning, and improvisation
- Virtual design games for shared decision-making and prioritization.

Web-Based Research Ethics

Web-based research has uncovered new ethical challenges for researchers, requiring new considerations for various aspects of recruitment and participation. Despite the growing interest in web-based research, the ethical guidelines and policies needed to guide these practices are insufficient [33]. Web-based research refers to situations in which researchers set the research context as one with a significant interaction between the researcher and participants in a web-based setting.

Participants' privacy (ie, family and cultural considerations), anonymity and confidentiality, informed consent, and data security and integrity are some of the ethical challenges that should be considered in web-based research settings. Factors related to these challenges include the audience (with whom to consult), type of research activity, and epistemological perspectives (space or place and text based or person based), informed consent (public-private, degree of interaction, topic sensitivity, and subject vulnerability), researcher ethos (credibility and variability of roles), and ethical representation (publication in the age of remix, multimedia, and search engines) [20]. It is important to consider the different ethical issues involved in various types of web-based research. “Each type of research involves different levels of involvement and interaction from both the participant and researcher. The more involvement and interaction, the greater, one can assume, the ethical risk may be” [33].

Because the D.DEC project included a diverse participant group with a range of digital literacy and access, it was important to create an equitable and inclusive experience for participation. We created options for participation both in how and when to participate, as well as asynchronous and synchronous options, including flexibility during a session. Asynchronous activities, such as asynchronous feedback, increase flexibility and extend the potential breadth of participation [34]. We provided phone call availability for answering questions and put in place multiple facilitators to help ensure equal voice when a variety of engagement techniques or technologies were in use (phone or chat or virtual meeting or paper based). In addition, the identification of participants (names and faces) and recording of sessions were 2 other ethical challenges in the D.DEC project.

Discussion

Principal Findings

Web-based co-design methods and techniques have become increasingly common across industries, settings, and professions [35]. Consequently, conducting remote and virtual co-design has become an opportunity to advance participatory techniques. This viewpoint presents considerations for using a co-design developed from inquiry and adaptations during the pandemic. We examined the recent challenges in conducting co-design and identified potential opportunities to address them for projects conducted for a health context. The web-based co-design phases from the D.DEC case study revealed some of the drawbacks and challenges of the web-based setting, including technical challenges and capabilities related to using web-based platforms and the level of participation, collaboration, and interaction among participants. Through a literature review and scanning of web-based or offline co-design resources, we identified the main factors that can affect co-design in virtual settings, including facilitation, collaboration, platform accessibility, communication, sensemaking, using tangible tools, and web-based research ethics. These factors are intended to improve communication, increase shared understanding, support effective sensemaking, and support meaningful discussion among participants, which in turn may improve interaction, collaboration, and the generation of new ideas and creative solutions.

In developing a set of considerations for web-based co-design in health, we looked at existing work, including the participatory framework proposed by Sanders et al [28]. Sanders et al [28] proposed that a design framework can help design researchers determine which participatory techniques and tools are most relevant for a specific design process. The framework by Sanders et al [28] provides an overview of participatory design tools and techniques in virtual and in-person settings that may be complementary to the 7 considerations shared in this viewpoint. The framework by Sanders et al [28] is intended to orient practitioners to the purpose and context of participatory tools and techniques and to support the customization of those tools and techniques [18]. Identifying the project’s context, purpose, and goals is an essential step in planning a participatory research process in both the real world and virtual settings [10]. Important considerations for planning include (1) the context of the project including the purpose, goals, and objectives; (2) the target participants, in terms of numbers, abilities, motives, background, and experiences; (3) the characteristics and agenda of the process (eg, outputs and communication characteristics); (4) the characteristics of activities and tasks (eg, types of activities in terms of form, complexity, and timing); (5) the platforms and tools that fit with the goals and outcomes of the project; and (6) web-based research ethics [10]. These 6 aspects of the framework focus on planning co-design, and we would recommend consulting the framework alongside the 7 more conceptual considerations for co-design more so than some of the practical aspects.

In support of co-design outcomes, tools and techniques should aim to improve idea generation by facilitating communication and interaction among participants throughout the process (eg, visual tools such as graphic representations and generative tools such as cards and storyboards). Tangible tools, design games, and play-like activities are used in co-design to highlight the exploratory, imaginative, dialogical, and empathic aspects of co-design in improving idea generation and fostering communication between participants. “The means for reaching these objectives are drawn up in addition to the design (eg, tangible mock-ups and user representations) from the world of games (eg, role-playing, turn-taking, make-believe) to deliberately trigger participants’ imagination as a source of ideation for problem solving” [6]. There are several ways through which “play” can be initiated into a virtual co-design setting such as “play triggers” involving physical, verbal, or situational factors [7]. “Play” can be supported through tangible materials (such as game boards, playing cards, or prompt cards) and rules to provide a starting point or signal to the overall tone and expectations of a free and safe space to explore imaginative thoughts and ideas in a low-fidelity manner.

From a more practical perspective, it is critical to select the right platform or combination of software and platforms, activities, and materials that support the inclusivity of diverse participants in the research process. Well-planned web-based methods such as co-design workshops rely on integrating alternative access and communication methods to enhance inclusivity through increased accessibility. Alternatives include sharing information and materials via mail. Conducting and facilitating a web-based co-design process consists of 3 main phases [17]: the preworkshop phase (planning process), the workshop phase (conducting process), and the postworkshop phase (data analysis and evaluation process). Well-planned virtual processes rely on applying the right combination of web-based or offline and synchronous or asynchronous tools to enable opportunities for both facilitators and participants to feel empowered during the co-design process. Providing combinations and alternatives supports participants in managing their time, space, and feedback and enables facilitators to analyze the outcomes and adjust their methods and agenda throughout the process [18]. In addition, to address the challenge of agency over the process, applying a combination of web-based or offline and synchronous or asynchronous tools and techniques can help integrate opportunities for both facilitators and participants to include agency in the co-design process. This offers participants flexibility in managing their time, space, and feedback and enables facilitators to analyze the outcomes and adjust their methods and agenda as they see fit [18].

Conclusions

In recent years, participatory methods, including co-design, have been integrated into health care. The application of these methods in a web-based setting is rapidly expanding, specifically in the new era of remote work and collaboration owing to the COVID-19 pandemic. Adapting and using participatory design methods in a web-based setting requires the knowledge and skills to combine offline and virtual technologies, virtual collaboration, and creative methods and techniques.

We have been able to integrate existing work on the practical and conceptual aspects of co-design together with practical experience by adapting a co-design project for web-based engagement in the health sector. We present 1 example, but there are many projects that have experimented with adaptations out of necessity during the pandemic. Further research is required to fully capture the learnings from these experiences to improve co-design and to effectively transfer co-design methods to a web-based setting. However, transforming all the co-design methods and techniques into a web-based setting may neither be possible nor necessary.

Acknowledgments

The authors would like to thank the Designing Discharge After Emergency Care project team (Yesmeen Ghader, Matthew Beaubien, Alessandra Ceccacci, Galo Ginocchio, Christopher Rice, and Victoria Weng) who contributed to the case study that the authors have drawn on. The authors would also like to thank the expert reviewers for their insightful and constructive feedback.

Conflicts of Interest

None declared.

References

1. Ward ME, De Brún A, Beirne D, Conway C, Cunningham U, English A, et al. Using co-design to develop a collective leadership intervention for healthcare teams to improve safety culture. Int J Environ Res Public Health 2018 Jun 05;15(6):1182 [FREE Full text] [doi: 10.3390/ijerph15061182] [Medline: 29874883]
2. Langley J, Wolstenholme D, Cooke J. 'Collective making' as knowledge mobilisation: the contribution of participatory design in the co-creation of knowledge in healthcare. BMC Health Serv Res 2018 Jul 25;18(1):585 [FREE Full text] [doi: 10.1186/s12913-018-3397-y] [Medline: 30045726]
3. Donetto S, Pierri P, Tsianakas V, Robert G. Experience-based co-design and healthcare improvement: realizing participatory design in the public sector. Design J 2015 May 07;18(2):227-248. [doi: 10.2752/175630615X14212498964312]
4. De' R, Pandey N, Pal A. Impact of digital surge during Covid-19 pandemic: a viewpoint on research and practice. Int J Inf Manage 2020 Dec;55:102171 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2020.102171] [Medline: 32836633]
5. Kennedy A, Cosgrave C, Macdonald J, Gunn K, Dietrich T, Brumby S. Translating co-design from face-to-face to online: an Australian primary producer project conducted during COVID-19. Int J Environ Res Public Health 2021 Apr 14;18(8):4147 [FREE Full text] [doi: 10.3390/ijerph18084147] [Medline: 33919920]
6. Vaillancourt S, Gupta S, Ghader Y, Sellen K. Co-designing discharge after emergency care. DDEC Report. 2021 Jan 30. URL: http://stmichaelshospitalresearch.ca/wp-content/uploads/2021/06/DDEC-REPORT.pdf [accessed 2022-01-03]
7. Vaajakallio K, Mattelmäki T. Design games in codesign: as a tool, a mindset and a structure. CoDesign 2014 Mar 06;10(1):63-77. [doi: 10.1080/15710882.2014.881886]
8. Brown S, Vaughan C. Play How it Shapes the Brain, Opens the Imagination, and Invigorates the Soul. Melbourne: Scribe Publications Pty Limited; Jun 28, 2010.
9. Jull J, Giles A, Graham ID. Community-based participatory research and integrated knowledge translation: advancing the co-creation of knowledge. Implement Sci 2017 Dec 19;12(1):150 [FREE Full text] [doi: 10.1186/s13012-017-0696-3] [Medline: 29258551]
10. Schiavo R. Health Communication From Theory to Practice. Hoboken, New Jersey: Wiley; 2013.
11. Baxter P, Jack S. Qualitative case study methodology: study design and implementation for novice researchers. Qual Report 2015 Jan 14;13(4):544-559 [FREE Full text] [doi: 10.46743/2160-3715/2008.1573]
12. McCarthy K, Wainfan L, Davis P. Chapter three virtual versus face-to-face collaboration: a survey of the literature. In: Challenges in Virtual Collaboration. Santa Monica, California, United States: RAND Corporation; 2004.
13. Sucharew H, Macaluso M. Progress notes: methods for research evidence synthesis: the scoping review approach. J Hosp Med 2019 Jul 01;14(7):416-418. [doi: 10.12788/jhm.3248] [Medline: 31251164]

14. Crowe S, Cresswell K, Robertson A, Huby G, Avery A, Sheikh A. The case study approach. BMC Med Res Methodol 2011 Jun 27;11:100 [FREE Full text] [doi: 10.1186/1471-2288-11-100] [Medline: 21707982]
15. Vaillancourt S, Gupta S, Ghader Y, Sellen K. Co-designing discharge after emergency care: (D.DEC) - Summary Report. St. Michael's Unity Health toronto. 2020 Dec 21. URL: https://research.unityhealth.to/wp-content/uploads/2021/06/DDEC-REPORT.pdf [accessed 2023-02-07]
16. Karpova E, Correia A, Baran E. Learn to use and use to learn: technology in virtual collaboration experience. Internet Higher Educ 2009 Jan;12(1):45-52. [doi: 10.1016/j.iheduc.2008.10.006]
17. Web-based co-design: social media tools to enhance user-centred design and innovation processes. VTT Science 34. 2013 Jan 10. URL: https://www.vttresearch.com/sites/default/files/pdf/science/2013/S34.pdf [accessed 2022-01-03]
18. Lim F. An analysis of synchronous and asynchronous communication tools in e-learning. Advanced Sci Technol Letter 2017;143:230-234 [FREE Full text] [doi: 10.14257/astl.2017.143.46]
19. Salmi A, Mattelmäki T. From within and in-between – co-designing organizational change. CoDesign 2019 Mar 11;17(1):101-118. [doi: 10.1080/15710882.2019.1581817]
20. Markham A, Buchanan E. Ethical decision-making in internet research: recommendations from the AOIR ethics working committee. AOIR. 2012. URL: http://aoir.org/reports/ethics2.pdf [accessed 2022-01-03]
21. Leary A. How to make your virtual meetings and events accessible to the disability community. Rootedinrights. 2020 Apr 13. URL: https://rootedinrights.org/how-to-make-your-virtual-meetings-and-events-accessible-to-the-disability-community/ [accessed 2022-01-03]
22. Best video conferencing apps and software for accessibility. SCOPE. URL: https://bighack.org/best-videoconferencing-apps-and-software-for-accessibility/ [accessed 2021-02-08]
23. Hopper EK, Bassuk EL, Olivet J. Shelter from the storm: trauma-informed care in homelessness services settings~!2009-08-20~!2009-09-28~!2010-03-22~!. Open Health Serv Policy J 2010 Apr 07;3(2):80-100. [doi: 10.2174/1874924001003020080]
24. Satel S. Delivering services to individuals with severe mental illness: SAMHSA falls short. Psychiatr Serv 2014 Sep 01;65(9):1160-1161. [doi: 10.1176/appi.ps.201400259] [Medline: 25179186]
25. Lorini M, Sabiesco A, Memarovic N. Collective digital storytelling in community-based co-design projects: an emergent approach. J Community Informatics 2017 Mar 22;13(1) [FREE Full text] [doi: 10.15353/joci.v13i1.3296]
26. Soulier E, Caussanel J. Narrative tools to improve collaborative Sense-Making. AAAI. 2002. URL: https://aaai.org/Library/Workshops/2002/ws02-09-002.php [accessed 2022-01-03]
27. Amber M, Jorgen S. Sensemaking in 'real' versus virtual environments: a comparison and challenge. In: Proceedings of the Third International Symposium on Process Organization Studies. 2011 Presented at: Third International Symposium on Process Organization Studies; Jun 16-18, 2011; Corfu, Greece URL: https://espace.library.uq.edu.au/view/UQ:255683
28. Sanders E, Brandt E, Binder T. A framework for organizing the tools and techniques of participatory design. In: Proceedings of the 11th Biennial Participatory Design Conference. 2010 Presented at: PDC '10: The 11th Biennial Participatory Design Conference; Nov 29 - Dec 3, 2010; Sydney Australia. [doi: 10.1145/1900441.1900476]
29. Zamenopoulos T, Alexiou K. Co-design as collaborative research. In: Connected Communities Foundation Series. Bristol: University of Bristol/ AHRC Connected Communities Programme; 2018.
30. Mattelmäki T. Probing for co-exploring. CoDesign 2008 Mar;4(1):65-78. [doi: 10.1080/15710880701875027]
31. Brandt E. Designing exploratory design games: a framework for participation in Participatory Design? In: Proceedings of the ninth conference on Participatory design: Expanding boundaries in design - Volume 1. 2006 Presented at: PDC'06: Expanding Boundaries in Design; Aug 1 - 5, 2006; Trento Italy. [doi: 10.1145/1147261.1147271]
32. Constantin A, Alexandru C, Korte J, Wilson C, Fails JA, Sim G, et al. Distributing participation in design: addressing challenges of a global pandemic. Int J Child Comput Interact 2021 Jun;28:100255 [FREE Full text] [doi: 10.1016/j.ijcci.2021.100255] [Medline: 34976715]
33. Warrell JG, Jacobsen M. Internet research ethics and the policy gap for ethical practice in online research settings. Can J High Educ 2014 Apr 30;44(1):22-37. [doi: 10.47678/CJHE.v44i1.2594]
34. Davis A, Gwilt I, Wallace N, Langley J. Low-Contact Co-Design: considering more flexible spatiotemporal models for the co-design workshop. Strategic Design Res J 2021;14(1) [FREE Full text] [doi: 10.4013/sdrj.2021.141.11]
35. Steen M, Manschot M, De Koning N. Benefits of co-design in service design projects. Int J Design 2011;5(2) [FREE Full text]

Abbreviations

D-DEC: Designing Discharge After Emergency Care
ED: emergency department

Edited by A Mavragani; submitted 24.01.22; peer-reviewed by M(Hardey, YC Tseng, A Hidki; comments to author 20.02.22; revised version received 29.04.22; accepted 19.12.22; published 03.03.23.

Please cite as:
Mallakin M, Dery C, Vaillancourt S, Gupta S, Sellen K
Web-Based Co-design in Health Care: Considerations for Renewed Participation
Interact J Med Res 2023;12:e36765
URL: https://www.i-jmr.org/2023/1/e36765
doi:10.2196/36765
PMID:36595738

Viewpoint

Toward Public Health Resilience in the Eastern Mediterranean Region: Findings From the Seventh Eastern Mediterranean Public Health Network Regional Conference

The Eastern Mediterranean Public Health Network (EMPHNET)

Corresponding Author:
The Eastern Mediterranean Public Health Network (EMPHNET)
Abdallah Ben Abbas St
Building No 42
Amman, 11196
Jordan
Phone: 962 6 5519962
Email: executive.director@globalhealthdev.org

Abstract

The resilience of public health in the Eastern Mediterranean Region (EMR) varies from country to country, mostly based on the governmental and financial situation of the countries. With the theme of Towards Public Health Resilience in the EMR: Breaking Barriers, the seventh Eastern Mediterranean Public Health Network regional conference, held from November 14 to 18, 2021, was dedicated to exploring ways for achieving public health resilience. A total of 101 oral presentations and 13 poster presentations were presented on various public health topics. The conference included 6 keynote sessions, 10 roundtable sessions, and 5 preconference workshops. The preconference workshops were conducted on border health; the mobilization of Field Epidemiology Training Program (FETP) residents and graduates and rapid responders in EMR countries; continuous professional development for the public health workforce; brucellosis surveillance using the "One Health" approach; and strategies to integrate and use noncommunicable diseases data sources. The roundtable sessions included discussions on the following topics: the role of FETPs in responding to COVID-19, institutionalization of rapid response to public health emergencies, health systems resilience, integration of early warning and response with event-based and indicator-based surveillance, sustaining international health regulations, strengthening the "One Health" approach, the anticipated future of public health in the post COVID-19 era, supporting public health research capacity in a diverse region, and COVID-19 vaccines and routine immunization synergies and drawbacks. The keynote speaker sessions covered topics on essential public health functions and the universal health coverage challenge in the EMR, lessons from the US COVID-19 public health response, learning from COVID-19, reshaping public health after the pandemic era, COVID-19 resilient primary health care, and the cohesion of society during and after a pandemic. The conference sessions provided highly promising opportunities to explore ways to achieve such goals in the EMR and shed light on the latest scientific findings, important lessons learned, and discussions on the ways in which current barriers can be broken down through coordination and collaboration.

(Interact J Med Res 2023;12:e36356) doi:10.2196/36356

KEYWORDS

COVID-19; surveillance; public health; health system; conference; Eastern Mediterranean Region

Background

COVID-19 has caused an unprecedented global crisis, resulting in the loss of millions of lives, shock to public health systems, and economic and social disruption. The pandemic has disproportionately affected the most vulnerable populations [1]. It has challenged local, national, regional, and global capacities and exposed the limitations of many health systems, including those previously known for their high performance and resilience [2]. Health system resilience can be defined as the capacity of health actors, institutions, and populations to prepare for and effectively respond to crises; maintain core functions when a crisis hits; and, informed by lessons learned during the crisis, reorganize if conditions require it. Public health resilience is the ability to prepare for, manage (absorb, adapt, and transform), and learn from shocks. Shocks can predominantly affect the demand side of the health system (eg, an epidemic will increase health care needs) or its supply side (eg, an economic crisis will typically cause a reduction in available resources) or both and can be more or less severe.

Public health plays a pivotal role in building and strengthening resilience at the individual, community, and system levels. Public health resilience is needed more than ever before to recover from the extreme disruptions of today's life and to adapt and prosper in the face of challenges and barriers, while maintaining the core functions of the health systems. During crises, resilient health systems can effectively adapt in response to evolving situations and reduce vulnerability across and beyond the systems. The resilience of public health in the Eastern Mediterranean Region (EMR) varies from country to country, mostly based on the governmental and financial situation of the countries. The various national strategies used to control viral transmission are widely debated. However, the relative success of these strategies depends largely on the structure, governance, and financing of the existing health system across all levels.

The seventh Eastern Mediterranean Public Health Network (EMPHNET) regional conference, held from November 14 to 18, 2021, was dedicated to exploring ways to achieve public health resilience. To this effect, EMPHNET adopted the theme of *Towards Public Health Resilience in the EMR: Breaking Barriers*. Conference sessions provided highly promising opportunities to explore ways to achieve such goals in the EMR and shed light on the latest scientific findings, important lessons learned, and discussions on the ways in which current barriers can be broken down through coordination and collaboration. The objectives of the conference were to engage public health experts and national, regional, and international entities in a discussion of public health challenges hindering the achievement of public health resilience in the EMR; share public health lessons and expertise; and present the accomplishments of public health professionals from the region. The conference consisted of oral and poster presentations, preconference workshops, roundtable discussions, and keynote speeches.

Oral and Poster Abstract Sessions

A total of 101 oral presentations and 13 poster presentations were presented by FETP residents and graduates as well as public health specialists in the region. Their presentations offered opportunities to critically discuss new lessons learned from past years and explore new opportunities to become more resilient health systems. A total of 32 abstracts were in the area of COVID-19, 18 abstracts on surveillance, 7 abstracts on zoonotic diseases, 7 abstracts on vaccine-preventable diseases, 7 abstracts on noncommunicable diseases, 6 abstracts on vector-borne diseases, 6 abstracts on tuberculosis and HIV/AIDS, and 6 abstracts on maternal and child health. The remaining abstracts covered other areas of public health. All abstracts are made available in the abstract book (Multimedia Appendix 1).

Preconference Workshops

There were 5 preconference workshops, facilitated by experts within their respective fields, highlighting public health concepts and topics relevant to the region. These topics were border health approaches to mitigate cross-border communicable disease spread; the mobilization of FETPs and rapid responders in EMR countries; continuous professional development for public health workforce; brucellosis surveillance, diagnosis, and control using the "One Health" approach; and identifying barriers and strategies to integrate and use noncommunicable diseases data sources.

Border Health Approaches to Mitigate Cross-border Communicable Disease Spread

The COVID-19 pandemic has highlighted the fact that communicable diseases have no borders and has proved the vital role of border health strategies to limit the spread of such outbreaks globally. Border health measures aim to mitigate the international spread of communicable diseases through improved systems designed to detect, prevent, and respond to public health events [3]. In our increasingly connected world, border health strategies must remain innovative and resilient to effectively mitigate the cross-border spread of communicable diseases [4]. Therefore, it is important to strengthen cross-border public health surveillance, information sharing, and collaboration as part of a comprehensive border health system. This workshop aimed to understand population movement and connectivity to improve public health systems. The participants were able to identify border health factors and activities that a public health authority would consider when faced with a highly pathogenic outbreak in a neighboring country. Such factors include identifying resource needs, reviewing current public health emergency or pandemic plans and standard operating procedures (SOPs), enhancing surveillance, considering nonpharmaceutical interventions, and risk communication messaging. Participants were also able to explain how understanding population mobility and strengthening cross-border collaboration contribute to a comprehensive approach to border health and identify actions that a country should take to communicate a public health problem to its counterparts in a neighboring country. Challenges for cross-border collaborations were also identified, such as lack of up-to-date information from the borders, case admissions, lack of resources, language barriers, and political tension. At the end of the workshop, it was recommended that countries implement Public Health Corridors to harmonize their COVID-19 recovery road maps. Public Health Corridors is a strategy developed by the Collaborative Arrangement for the Prevention and Management of Public Health Events in Civil Aviation, which describes, from an aviation standpoint, how to handle essential flights in a manner that maintains safety and prevents the transmission of COVID-19, all while minimizing additional burdens. It is also important for countries to adopt multiple border health mechanisms as different layers of protection against the spread of internationally transmissible diseases. Moreover, there is a need to strengthen the detection, protection, and response capacities at the borders, including building the capacity of the officers responsible for public health activities at points of entry (POEs). Countries must ensure that they have relevant cross-border public health plans and SOPs developed and that relevant staff are familiar with and trained on these SOPs.

Mobilization of FETPs and Rapid Responders in EMR Countries

The roles of public health rapid response teams (RRTs), FETPs, and public health emergency operations centers (PHEOCs) were witnessed during the COVID-19 pandemic. In the EMR, the FETP residents, graduates, and mentors have contributed substantially to the public health responses to the COVID-19 pandemic in their respective countries by being involved in case investigations, POE and arrival screening, isolation protocols, transferring cases, risk communication, and training on infection control [5]. RRTs, FETPs, and PHEOCs are critical assets within a country's public health emergency response and manage disease outbreaks and other events [6]. Thus, coordinating the roles and responsibilities of these 3 elements is crucial for effective and efficient emergency response plans. This workshop aimed to introduce an operational framework that coordinates the roles of FETPs, RRTs, and PHEOCs. A framework that links FETPs, RRTs, and emergency operations centers (EOCs) was discussed and proposed during this workshop, as this will enhance the EMR emergency response capacity and strengthen the management and coordination among EMR countries. The operationalization of the proposed conceptual framework requires commitment from the EOCs, RRTs, and FETPs. The core element of an efficient coordinated response is preparedness. During the "nonemergency phase," the development of SOPs is essential to indicate the role of each actor in the response. Having standardized training and capacity building is important to ensure that all actors share the same competencies and skills. During the emergency phase, maintaining good communication and sharing information are essential to ensure a coordinated response. Including FETPs in the development of SOPs and response strategies as well as their involvement at various levels of the health system is beneficial.

Continuous Professional Development for Public Health Workforce

Continuing professional development (CPD) refers to all formal and informal activities that health workers engage in to maintain, update, develop, and improve their professional skills, knowledge, and attitudes [7]. CPD for public health workers is essential to ensure the best practices in a field that involves multiprofessional, multidisciplinary, and multiorganizational activities. FETPs are competency-based training and service programs in applied epidemiology and public health that help countries build public health system capacity. The COVID-19 pandemic revealed the need to equip FETP graduates with the additional competencies needed to manage the newly emerging threats. This regional workshop was conducted to propose a road map for CPD as part of FETP graduates' career pathway. The workshop participants identified FETP graduates' ongoing competency requirements and proposed a strategy for incorporating CPD activities to meet those requirements in the EMR in the coming years. This workshop devolved the participants' skills to identify the needs and potential opportunities for FETP graduates to continue their professional education and to outline a work plan for FETP graduate support in the EMR. At the end of the workshop, it was recommended that CPD be institutionalized within the Public Health system and that the FETP career path should include CPD activities. Each ministry of health should appraise the public health workforce and have clear objectives and CPD plans to meet the needs of individuals and teams.

Brucellosis Surveillance, Diagnosis, and Control Using the "One Health" Approach

Brucellosis is one of the most common zoonotic diseases worldwide [8] affecting public and veterinary health. It is reported to be particularly widespread in the Middle Eastern and North African countries. However, many countries do not yet have well-established surveillance systems or the laboratory capacities required to accurately confirm brucellosis in their human or animal populations [9]. This workshop aimed to present the latest recommendations regarding laboratory diagnosis of brucellosis, discuss how the concept of "One Health" can be applied in low-income settings to control the spread of brucellosis, and provide a proper methodology for conducting surveillance studies of zoonotic diseases. The workshop provided an excellent opportunity for public health professionals and officials from health and agricultural ministries of the EMR to reflect on the endemic of brucellosis and what needs to be done to control its spread. Regarding laboratory detection of Brucella species, culturing is the gold standard but is unfeasible owing to its lengthy process. Serological testing may be ideal in reduced-resource settings; however, confirmation requires repeated testing. Surveillance studies in the EMR require a combination of capacity-building activities and awareness-raising activities. The veterinary sector in the EMR can benefit from training on advanced laboratory detection methods. Moreover, updates to animal vaccination strategies are necessary to reduce the loss of livestock. Applying the One Health concept by building multisectoral teams is necessary for the successful implementation of programs aimed at controlling the spread of zoonotic diseases such as brucellosis. Among the lessons learned were the importance of border activity in controlling case numbers within individual nations and creating a demand for multinational or regional collaboration on the issue of brucellosis. Establishing a sustainable communication channel and a culture of collaboration between public health, animal control, and environmental bodies within the government were unanimously championed by workshop facilitators and participants as the way forward to deal with brucellosis and zoonoses. The next steps should include an analysis of the socioeconomic assessment and burden of the disease, as it can be a powerful tool to influence the development of national action plans in EMR countries. Properly designed surveillance programs, which involve the animal health sector as well as the human health sector and use appropriate methods for testing, can provide a more accurate picture of the epidemiology of brucellosis and aid in creating more specific action plans to target and respond to it. The workshop participants made different recommendations including conducting regular awareness sessions for farmers on the benefits of vaccinating their herds and for community members on safe practices for dairy production and consumption. The use of vaccination against *Brucella abortus* in cattle should be covered by animal vaccination programs as it has been isolated in Jordan and Iraq. In any laboratory handling Brucella samples, proper implementation of biosafety and biosecurity guidelines is necessary to prevent laboratory-acquired infections. Due to the scarcity of biosafety level III laboratories in the EMR, serological detection methods may be preferred, as they are less dangerous than culturing methods.

Identifying Barriers and Strategies to Integrate and Use Noncommunicable Diseases Data Sources

The generation of high-quality health data is a prerequisite for evidence-based decision-making and national planning [10]. An effective health information system (HIS) provides accurate and timely information on health indicators to guide national health system management efforts [11]. The goal of this workshop was to review the fundamentals of HIS and to identify barriers and strategies for the integration and use of noncommunicable diseases data sources, using Jordan as a case study. The workshop highlighted the current efforts in Jordan toward establishing an HIS that caters to noncommunicable diseases. The comprehensive HIS assessment conducted by the World Health Organization (WHO) in 2016 and 2018 was also highlighted. The discussions raised several barriers and challenges toward achieving a comprehensive and consolidated national HIS. One of these barriers is the multisectoral approach. In Jordan, the health care system includes the public sector, private sector, Royal Medical Services, King Hussein Cancer Center, and university hospitals [12], and data exchange between the different health sectors is the ultimate goal of attaining a national HIS. In contrast, the multisectoral approach of the health care system creates difficulties in integrating the HIS, considering the complexities in obtaining approval from the concerned parties to integrate their HIS. There is a lack of useful data, where unnecessary data are sometimes collected, creating technical and financial burdens in translating the data into information. In addition, a low workforce capacity in developing, deploying, and running the HIS is another barrier [13]. Finally, financial constraints and lack of sustainable health system financing prevent the continuation of such an HIS. There was consensus on the importance of a national HIS that caters to noncommunicable diseases in Jordan. Several recommendations were made during the discussion, including the need to develop a noncommunicable disease strategy that identifies national priorities for data collection and establishes a unified mechanism for all sectors. Case definitions of the indicators and targets for noncommunicable diseases should be agreed upon, unified between data sources, and used by the HIS. The WHO's [14] HIS assessment recommendations should be considered, and action should be taken with regard to establishing committees of representatives from involved sectors; developing a prototype of the dashboard model; developing HIS policy; continuing with Survey, Count, Optimize, Review, Enable assessment annually to monitor the progress of the work; and finally disseminating the data to all stakeholders. Human resource development is key to ensuring that the collected data are analyzed and translated into information that can be used by policy makers for action. A mechanism to govern patient data should be available, with transparency in how the data are collected, analyzed, used, and shared to ensure that patient confidentiality is maintained.

Roundtables

Several topics were discussed in the roundtable discussions including the role of FETPs in responding to COVID-19; institutionalization of rapid response to public health emergencies in EMR countries; integration of early warning and response (EWAR) with event-based and indicator-based surveillance; sustaining the International Health Regulations (IHR) in Iraq; health systems resilience; strengthening the "One Health" approach; supporting public health research capacity, quality, and productivity in a diverse region such as the EMR; the anticipated future of public health post COVID-19; and COVID-19 vaccines and routine immunization synergies and drawbacks.

Role of FETPs in Responding to COVID-19: Lessons Learned and Challenges

Established in 1980 by the US Centers for Disease Control and Prevention, the FETPs are competency-based training programs aimed at enhancing the epidemiological capacity of the public health workforce [15]. FETP fellows and graduates in the EMR have contributed significantly to the control of many epidemics in the past and continue to contribute to emerging public health threats, including COVID-19 [16].

FETP graduates worldwide and in the EMR were well engaged in the response to the pandemic, including conducting surveillance activities; conducting screening interviews at POEs; data collection, management, and analysis; risk communications; and other activities. However, it is also important to focus and learn from mistakes, shortcomings, and failures. The major challenges faced during the pandemic dealt with logistics, establishing solid surveillance systems, meeting country-specific needs, and vaccine acceptance. FETP residents, both graduates and in training, gained experience improvising and innovating their resources and abilities to combat the challenges brought by the novel coronavirus. The COVID-19 pandemic confronted field epidemiologists with tough challenges, but it also taught them valuable lessons that will better equip them to be well prepared for future outbreaks and pandemics. This roundtable highlighted the need for continuous technical and financial support to FETPs, in addition to the need to institutionalize FETPs and establish new FETPs in other countries.

Institutionalization of Rapid Response to Public Health Emergencies in EMR Countries

The IHR dictates the need for state parties to establish the capacity to respond promptly and effectively to public health risks [4]. Consequently, RRTs, the multidisciplinary teams who respond to public health events, are essential to contain the harmful effects of emergencies and coordinate responses in fragile situations such as the EMR. Institutionalization of the rapid response process allows the deployment process to become part of the national system and facilitates a timely and effective response to emergencies [17].

Although the setup of RRTs varies from one country to another, there have been various efforts to build and sustain the capacities of RRTs in the region. It is important to link the RRT, FETP, and EOC structures across the health system and invest in the career development of RRTs and FETP graduates and alumni to better retain and mobilize them. To leverage resources at the regional level and better coordinate workforce mobilization, more institutionalization needs to take place nationally. Moreover, coordinating different capacity-building efforts is imperative for standardizing the curriculum and competencies for RRT training. More work is needed to develop key performance indicators for RRTs and to document the challenges and opportunities in their countries.

Integration of EWAR With Event-Based Surveillance and Indicator-Based Surveillance

Countries, particularly low-income and middle-income countries, must have EWAR systems incorporated in their surveillance system as required by the IHR that was adopted in 2005. EWAR aids in detecting and responding rapidly to signals and alerts coming from both formal and informal sources and within and outside the health sector to rapidly mobilize required resources in a flexible and responsive manner. The EMR has been one of the regions that is most affected by humanitarian emergencies, including armed conflicts and the influx of displaced people. The EMR currently has 8 active Early Warning Alert and Response Network systems in 7 countries experiencing protracted humanitarian emergencies: Afghanistan, Iraq, Libya, Somalia, Sudan, Syria, and Yemen [18]. Studies have shown that EWAR performance in the EMR has been optimal when looking at the timeliness and completeness of reporting and verification of alert systems. However, the population coverage was low for most, and the Early Warning Alert and Response Network's main focus of outbreak detection was weakened by the increasing number of other diseases. Currently, the WHO is working on multiple levels of surveillance with the Ministries of Health in the EMR. One of the main challenges is the political commitment toward the surveillance systems in the region's countries as well as the fragmentation of data in the EMR. The WHO aims to finalize all tools, systems, SOPs, and guidelines and secure the support of the involved partners and stakeholders to integrate surveillance systems into the current Ministries of Health systems by 2025. Multiple EMR countries have worked on both event-based surveillance and integrated disease surveillance and response with the WHO. The region needs to have 1 surveillance system with a unified data collection point, where all information can be located to describe the country's situation. These systems should have IT infrastructure in the country to collect the required information.

Health Systems Resilience

During crises, resilient health systems can effectively adapt in response to evolving situations and reduce vulnerability across and beyond the systems. It is a key factor in coping with a crisis such as the economic crisis and the COVID-19 pandemic [19]. Key findings presented during this session included building resilient health systems—as a priority for all member states, adequate investments in health for socioeconomic development, adequate investments in health emergency preparedness, integrated approach to health security and universal health coverage (UHC), the importance of building and strengthening the primary health care (PHC) foundation, investing in the essential public health functions (EPHFs), including all hazard emergency risk management capacities, applying the whole-of-society approach, and attention to vulnerable and marginalized groups. Continuous support is needed for FETP graduates to work toward strengthening surveillance systems, investigating outbreaks, and participating in regional and global efforts to respond to COVID-19. Lessons learned from the current situation in the EMR to strengthen both pandemic preparedness and health systems include the importance of investing in EPHFs, including those required for all-hazards emergency risk management; institutionalized mechanisms for whole-of-society engagement; strengthening the PHC approach for health security and UHC; and promoting enabling environments for research, innovation, and learning.

Sustaining the IHR in Iraq—Enhancing Multisectoral Coordination in the Face of Conflict

Implementation of the IHR in the EMR comes with unique challenges because of the lack of necessary funding, expertise, and infrastructure needed to develop the capacities for disease surveillance, as stipulated by the IHR [20,21]. In contrast, the advent of the COVID-19 pandemic highlighted that improved compliance with the IHR presents an opportunity to reduce the costs of life and improve the economy [22]. Moreover, the pandemic further emphasized the necessity for international aid and cooperation for the successful implementation of the IHR in low- to middle-income countries [21]. In this regard, Iraq has worked actively to build a solid foundation to establish the necessary tools, facilities, and procedures to comply with the IHR and protect national and international health security. Since 2017, the Government of Iraq (GOI) and Kurdistan Regional Government (KRG) have partnered with Georgetown University to implement the regulations and enhance coordination between the relevant sectors within the country. Great strides have been achieved with regard to IHR compliance through the various governmental bodies in Iraq since the start of the collaboration with Georgetown University in 2017. It is evident that the continuous work to build the core capacities, appoint IHR focal points, and build multisectorial networks has been advantageous in monitoring, reporting, and responding to communicable disease cases across all the governorates of the GOI and KRG. Nonetheless, the COVID-19 pandemic exposed certain weaknesses of IHR compliance that still need to be addressed, especially regarding rapid communication between all the stakeholders and with the WHO networks, in addition to the need for capacity development, particularly at the POEs and for the zoonotic disease surveillance teams. The panelists recommended unification of all the processes regarding IHR implementation in the KRG, strengthening communication between IHR focal points of the KRG and the GOI and sustaining the relationships between them, and increasing funding for more basic scientific research on viral pathogens. In addition, increasing awareness of the contributions of veterinary laboratories in disease outbreaks, providing better definitions of the roles and responsibilities of the veterinary laboratories of the GOI and KRG, and improving communication between the Central Public Health Laboratories and veterinary laboratories are all necessary for the improved application of the One Health concept. Biorisk capacity-building activities are needed in the KRG to improve preparedness for future PHEOCs. There is also a need for efficient reporting by focal points of various POEs in the country. This requires technology, training, and personnel support for the POEs. Building sustainable networks between all ministries, focal points, and the WHO for IHR compliance is important for continued improvements.

Strengthening the "One Health" Approach

"One Health" is an integrated unifying approach that aims to sustainably balance and optimize the health of people, domestic and wild animals, and the wider environment [22]. Its area of work includes food safety, control of zoonoses, laboratory services, neglected tropical diseases, environmental health, and combating antibiotic resistance. This roundtable discussion offered an opportunity to present the dynamics of emerging infectious disease pathogens; their impact on global health security; and the integrated solutions, definitions, principles, and institutionalization of One Health. Institutionalization and governance of the One Health strategy, securing political commitment, influencing policy changes, promoting multisectoral collaboration, community engagement for breaking silos, cultural changes for working together, and a better understanding of the interconnectedness and interdependence of human-animal-ecosystems are essential to strengthen the role of One Health. This roundtable provided recommendations regarding strengthening One Health capacity building based on the 5 principles of equity, parity, equilibrium, stewardship, and transdisciplinary. The engagement of nongovernmental organizations, the private sector, and other relevant players will strengthen the role of One Health. Epidemiological data and laboratory information should be shared across sectors to effectively detect, respond to, and prevent outbreaks of zoonoses and food safety problems. Joint responses to health threats and improving surveillance systems, early detection, notification, and management of wildlife diseases should be implemented by government officials, researchers, and workers across sectors at the local, national, regional, and global levels.

The Anticipated Future of Public Health Post COVID-19

COVID-19 has highlighted the need for better governance, more robust health systems and capacities, and the need to shift the paradigm toward public health and preventive medicine. This sheds light on the importance of coordination and collaboration among countries and stakeholders in different multilateral and global initiatives. Although the focus has rightly been on the immediate response to the virus, it is important to consider what comes next and ensure that lessons learned are followed. Many aspects need to be revised, including collaboration among countries, world trade regulations, budget allocation, partnership with the private sector, and health inequalities. Governments should build on their experiences and sustain the positive impacts of COVID-19 on public health by promoting and facilitating the adoption of lifestyles to reduce environmental pollution, revising tobacco control policies to build on the success of smoking reduction, and facilitating patients with chronic diseases to adopt regular self-care and healthy lifestyles. Governments should also consider aligning international strategies, including partnerships with the private sector; investment in digital technologies to strengthen pandemic management and future preparedness for other infectious diseases; and improving digital inclusivity by providing all society segments, especially the unprivileged, with access to digital skills and appropriate infrastructure. Governments should recognize the root causes of health inequalities and commit to short and medium road map remedy strategies. They should build trust among nations to enhance global collaboration and strengthen a wide network of global public health institutions and laboratories to prepare for any future outbreaks or pandemics by timely sharing of information, viral specimens, and genomic sequences; allocate appropriate research and development funds for the development of vaccines, diagnostics, and therapeutics; and establish regional centers of excellence for the rapid manufacturing of vaccines when needed. The appreciation of health professionals should not be viewed as a short-term response. Leaders must apply incentives to improve workforce retention and empower health professionals at all levels with the skills and equipment required to deal with future public health challenges. We need to enhance and apply complex systems modeling, especially in epidemiology and behavioral science, and strengthen surveillance systems for viruses, especially in birds and animals.

Supporting Public Health Research Capacity, Quality, and Productivity in a Diverse Region

Public health research plays a critical role in strengthening health systems. However, public health research productivity in the EMR remains below the world average [23]. Many challenges and barriers face public health research capacity, quality, and productivity in the EMR, including lack of funding, problems with data availability, language barriers for publishing, lack of guidelines and regulations, and inadequate research skills and competencies. Researchers conducting research in conflict contexts face more challenges compared with those in many other settings. However, there have been success stories in the EMR regarding research publication and its positive and effective impact on policy makers and decision makers. More research should be conducted in conflict areas to investigate the needs of people living in such areas and, consequently, the appropriate responses to their needs. Official guidelines on data sharing should be developed and should be clear and consistent with all public health data. The guidelines must find a balance between making data available and accessible to researchers and safeguarding privacy. Universities in the EMR must make research skills and competencies part of their undergraduate and postgraduate curricula. Research institutions are encouraged to focus on developing a research capacity educational program such as "train the trainer," where institutions adopting such a program must be flexible and willing to revise the plan if faced with barriers and challenges. Initiatives to facilitate the publication of research in the EMR must be implemented, such as hiring professional copy editors to read the manuscript before submission to the journal. Knowledge transfer frameworks and programs should be developed and implemented for collaborative knowledge transfer between researchers, policy makers, and other relevant stakeholders to facilitate the linkage between science and policy.

COVID-19 Vaccines and Routine Immunization Synergies and Drawbacks

Disruptions in health services because of the COVID-19 pandemic have strained health systems, resulting in 22.7 million children missing out on vaccination in 2020 [24]. Most of these children live in communities affected by conflict, underserved remote places, or informal or slum settings facing multiple deprivations including limited access to basic health and key social services [14]. Although access to immunization was hindered in 2020 because of the imposed curfews and deferred or postponed supplementary outreach activities, the Expanded Program on Immunization (EPI) was further challenged in 2021 by bearing the additional constraint of introducing the COVID-19 vaccine and implementing the vaccine uptake, often at the cost of routine pediatric vaccine-preventable diseases in most countries. In the EMR, this disruption varied across countries with respect to their socioeconomic status on the one hand and their pre–COVID-19 system infrastructure on the other hand. Although some countries in the region are struggling with the drastic coverage drop in essential vaccines or the below-target COVID-19 vaccine uptake, other countries succeeded in maintaining their vaccine-preventable disease coverage rates while achieving the WHO target for COVID-19 vaccines. For instance, in Oman, acknowledging that any coverage drop is significant and might jeopardize decades of effort urged the timely monitoring of drops and the timely intervention at lower levels. In Iraq, consolidating national efforts to compensate for the drop and planning and executing the COVID-19 vaccine deployment plan using the available infrastructure was deemed successful. Lessons learned from the region reveal that preparedness is the fundamental pillar for successfully maintaining the functionality of EPI and delivering the COVID-19 vaccine. COVID-19 vaccination highlighted the availability of opportunities in terms of funding, technical and capacity building, and social mobilization. It is of utmost importance to benefit from these investments to strengthen and maintain routine immunization in the future. Embracing that "every child counts" is the fundamental driver to consider when going over the lessons learned, challenges, and success stories of countries. Sharing experience and expertise among countries is essential to accelerate efforts to compensate for the drop in vaccine coverage. In addition, structural adjustments can be implemented at the country level in terms of preparedness to avoid dropping in future disruptions. The introduction of the COVID-19 vaccine highlighted the shift in EPI service delivery from those aged <18 years to vaccine delivery covering all age groups. As this paved the way for other new vaccines that might be introduced in the future to various age groups, it is important to translate all the lessons learned and investments into sustainability in terms of quality and not just quantity of EPI service delivery.

Keynote Speakers' Sessions

EPHFs and the UHC Challenge in the EMR

EPHFs are important for achieving the public health goal of improving, promoting, protecting, and restoring population health. UHC aims for all people and communities to have access to quality health services without financial hardship, which in turn improves health security and resilience to crisis.

In 2013, EMR countries endorsed UHC as a priority and developed a regional framework for action to advance UHC. The framework includes strategic actions for countries to achieve UHC and enhance financial risk protection to reduce out-of-pocket spending and financial hardship [25]. Several countries in the EMR are involved in the process of assessing their public health system strengths and weaknesses. To date, 2 countries, namely, Qatar and Morocco, have undertaken the formal assessment of the EPHFs using the EMR assessment tool [26,27]. The COVID-19 pandemic provides an opportunity for strengthening EPHFs and reinforcing health security that should not be missed. COVID-19 has highlighted the urgent need for a global commitment to make health systems resilient against all public health threats to sustain the progress toward health security and UHC. The pandemic has emphasized the lack of preparedness around the world and the need for stronger health systems to address pertinent gaps. Assessments of health systems can help reveal these gaps, but assessments need to be owned by the countries and engage all stakeholders. Strengthening EPHFs is the most cost-effective and sustainable way of achieving UHC.

Evidence, Experience, and Expertise: Lessons From the US Coronavirus Disease Public Health Response

The COVID-19 outbreak control requires synergic and multisectoral pillars: continuous support of epidemiological and genomic surveillance that is crucial for the detection and identification of circulating and potential new emerging variants; enhancing surveillance in specific congregates of high-risk groups to timely detect and stop the spread of the outbreak; and increasing vaccine demand and booster uptake paralleled with the clarification of rumors and misinformation. Maintaining adequate epidemiological surveillance is crucial for learning about the effectiveness of the COVID-19 vaccines against the new variants. Conducting prompt investigations of outbreaks in highly vaccinated areas or populations in addition to adequate genomic surveillance are key priorities to learn about virus dynamics and mutations and to provide knowledge that would benefit vaccine production. In addition, the use of telemedicine was beneficial during the lockdown, as it allowed people with comorbidities to seek medical care. It is important to build on this success to strengthen the infrastructure of telemedicine in the near future.

Learning From COVID-19: What It Would Take to Be Better Prepared

The political instability and fragile health systems in some EMR countries have hampered the effectiveness and efficiency of the strategies adopted to combat the COVID-19 pandemic [28]. Over time, this region has witnessed many outbreaks, including Middle East Respiratory Syndrome, cholera, polio, and increasing vector-borne and zoonotic outbreaks [29]. The region is also burdened by the refugee crisis and numerous internally displaced people. Multiple factors affected transmission, including conflict, demographics, timeliness of travel and social restrictions, migrant workers, and mass gatherings or pilgrimages. Countries that responded well to COVID-19 had a high level of political commitment, with multisectoral coordination. In addition, existing infrastructures including polio teams and regional laboratories were quick to mobilize and build on polio and influenza infrastructure. There have been several innovations, including applications, telemedicine, hotlines, e-clinics, and solar-powered oxygen generators. Most countries have banned waterpipe smoking and established mental health hotlines. However, the pandemic has also highlighted the weak epidemiological capacity of the region, the fragmented surveillance systems, and the lack of trust. Different areas of improvement were also highlighted. At the country level, there is a need to invest in building human capacities including epidemiologists, emergency responders, community health workers, health economists, communication specialists, and, most crucially, health leaders; strengthen health systems and work toward UHC and health security; work toward community engagement and community trust; and develop and update a multisectoral emergency preparedness plan. At the regional level, certain countries have greater capacities than others do in the region, and there is a need for more cooperation, solidarity, and support to effectively control the spread of the pandemic. Rich countries should ensure vaccine sharing, equity, and distribution with low-income countries. Moreover, countries should implement twinning programs in which human resources are shared across countries.

Reshaping Public Health After the Pandemic Era: The Agenda for the Next Decade—Are We Ready?

In 2020, the WHO conducted a global pulse survey to understand the impact of COVID-19 on the health system. Almost 90% of the 105 engaged countries reported an interruption to a different type of services that ranged from routine and elective service delivery to critical care, especially in low- and middle-income countries. There are many sources of disruptions. Financial constraints, supply chain disruptions, redirection of services to care for patients with COVID-19, and unavailability of the workforce are some of the factors affecting accessibility to essential health care services [30]. The findings demonstrated that investment in PHC is essential to mitigate the risks of future pandemics and to maintain the accessibility and delivery of essential health services during emergencies [31]. Investment in the health workforce is another dimension for a successful response, which includes training, mobilization, and redistribution to sustain high-quality essential health service delivery. Therefore, the COVID-19 pandemic serves as an opportunity for countries to reshape their public health to improve public health security for future pandemics. Integrating public health into primary care and investment in public health workforce capacity building is an essential approach for reshaping public health. A total of 6 models were identified by the WHO technical series on primary care called “Closing the Gaps Between Public Health and Primary Care Through Integration” to attain the integration of public health into PHC. This, in return, will focus the services on the population needs and achieve the person-centered approach. These models can be applied either individually or in combination depending on the flexibility of the health systems [32].

COVID-19–Resilient PHC: Experience and Challenge in United Nations Relief and Works Agency for Palestine Refugees in the Near East

The United Nations Relief and Works Agency for Palestine Refugees in the Near East (UNRWA) provides assistance and protection to Palestine refugees in Jordan, Lebanon, Syria, the Gaza Strip, and the West Bank, including East Jerusalem. The UNRWA has adopted multiple mechanisms to address the various health care needs. The adopted mechanism allowed for the effective sustainability of services for UNRWA PHC clinic services during the pandemic. In addition, using innovative and electronic services allowed the community access to essential health care services when needed. For this, health care staff must adopt available protection measures such as vaccination and infection protection and control practices. In addition, embracing the new normal for PHC UNRWA clinics is important to ensure continuity of services and inclusion of vulnerable community groups. Many actions can be taken to promote the sustainability of PHC services during the COVID-19 pandemic or any future outbreak. These actions include the need to improve the UNRWA system to include a resilient family medicine approach and services that adopt innovative digital systems and programs. There is a need to focus on strengthening the capacity of human resources by providing them with continuous education and clear career paths that ensure their development. In addition, COVID-19 brought organizations together, and partnerships proved to play a vital role in the ability to respond to outbreaks and emergencies. In addition, focusing on staff vaccination and protection measures and ensuring that UNRWA staff are well protected is the key to ensuring service sustainability. Supporting UNRWA through donations is another action that can be taken to sustain and enhance PHC at UNRWA clinics.

The Cohesion of Society During and After a Pandemic: How Does This Translate in the EMR Public

Social cohesion refers to the strength of relationships and the sense of solidarity among members of a community [33]. A group of researchers from the Bertelsmann Foundation studied social cohesion during the COVID-19 pandemic in 2020 to examine the fabric of society during the 3 waves of the pandemic in Germany. The results showed that social cohesion remained stable during the first wave, leading to greater visibility of solidarity in some areas. People in the middle and higher socioeconomic categories were more satisfied and confident in the pandemic response measures than those in the other categories. In the second half of the year, concerns about the future increased among all groups surveyed. In terms of division, middle-aged and low-educated respondents living in unsafe circumstances felt that society is highly divided. This was accompanied by lower levels of trust, a more significant rejection of pandemic response measures, and a growing fear that societal consensus is unattainable. Solidarity is a value that should be learned. These values are not always at the forefront of political choices when building healthy societies. The concept of solidarity is based on the idea that people can unite their differences because they have a common interest in not being divided and conquered. It is important to enhance global cohesion within the current inequalities between low-income and high-income countries, and it is recommended to enhance 2-way communication with public health professionals. Both sides must exchange information and decisions. Governments are responsible for educating and training their health workforce and improving their job conditions to minimize brain drains.

Conclusions

The seventh EMPHNET regional conference was dedicated to exploring ways to achieve public health resilience. To this effect, EMPHNET adopted the theme of *Towards Public Health Resilience in the EMR: Breaking Barriers*. Conference sessions provided highly promising opportunities to explore ways to achieve such goals in the EMR. The conference sessions shed light on the latest scientific findings, important lessons learned, and discussions on ways in which current barriers can be broken down through coordination and collaboration.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Abstract book.
[PDF File (Adobe PDF File), 2556 KB - ijmr_v12i1e36356_app1.pdf]

References

1. Bhaskar S, Rastogi A, Menon KV, Kunheri B, Balakrishnan S, Howick J. Call for action to address equity and justice divide during COVID-19. Front Psychiatry 2020 Dec 3;11:559905 [FREE Full text] [doi: 10.3389/fpsyt.2020.559905] [Medline: 33343410]
2. WHO's 7 policy recommendations on building resilient health systems. World Health Organization. URL: https://www.who.int/news/item/19-10-2021-who-s-7-policy-recommendations-on-building-resilient-health-systems [accessed 2021-10-19]
3. Handbook for Public Health Capacity-building at ground crossings and cross-border collaboration. World Health Organization. 2020. URL: https://www.who.int/publications/i/item/handbook-for-public-health-capacity-building-at-ground-crossings-and-cross-border-collaboration [accessed 2023-02-13]
4. International Health Regulations (2005) Third Edition. Geneva: World Health Organization; 2016.
5. Al Nsour M, Bashier H, Al Serouri A, Malik E, Khader Y, Saeed K, et al. The role of the global health development/eastern mediterranean public health network and the eastern Mediterranean field epidemiology training programs in preparedness for COVID-19. JMIR Public Health Surveill 2020 Mar 27;6(1):e18503 [FREE Full text] [doi: 10.2196/18503] [Medline: 32217506]
6. Griffith M, Ochirpurev A, Yamagishi T, Nishiki S, Jantsansengee B, Matsui T, et al. An approach to building Field Epidemiology Training Programme (FETP) trainees' capacities as educators. Western Pac Surveill Response J 2018 Sep 30;9(3):1-3 [FREE Full text] [doi: 10.5365/wpsar.2018.9.1.010] [Medline: 30377543]
7. Campbell C, Silver I, Sherbino J, Cate OT, Holmboe ES. Competency-based continuing professional development. Medical Teacher 2010 Jul 27;32(8):657-662. [doi: 10.3109/0142159x.2010.500708]
8. Seleem MN, Boyle SM, Sriranganathan N. Brucellosis: a re-emerging zoonosis. Vet Microbiol 2010 Jan 27;140(3-4):392-398. [doi: 10.1016/j.vetmic.2009.06.021] [Medline: 19604656]
9. Rubach MP, Halliday JE, Cleaveland S, Crump JA. Brucellosis in low-income and middle-income countries. Current Opinion Infect Dis 2013;26(5):404-412. [doi: 10.1097/qco.0b013e3283638104]
10. Sorenson C, Chalkidou K. Reflections on the evolution of health technology assessment in Europe. Health Econ Policy Law 2012 Jan 05;7(1):25-45. [doi: 10.1017/s1744133111000296]
11. Toolkit on monitoring health systems strengthening: health information systems. World Health Organization. 2008. URL: https://docplayer.net/9176924-Toolkit-on-monitoring-health-systems-strengthening-health-information-systems.html [accessed 2021-12-30]
12. The Higher Health Council. The National Strategy for Health Sector in Jordan 2015- 2019. 2015. URL: https://tinyurl.com/ms7kf5ry [accessed 2021-12-30]
13. Jabareen H, Khader Y, Taweel A. Health information systems in Jordan and Palestine: the need for health informatics training. East Mediterr Health J 2020 Nov 11;26(11):1323-1330. [doi: 10.26719/emhj.20.036] [Medline: 33226099]
14. COVID-19 pandemic leads to major backsliding on childhood vaccinations, new WHO, UNICEF data shows. World Health Organization. URL: https://www.who.int/news/item/15-07-2021-covid-19-pandemic-leads-to-major-backsliding-on-childhood-vaccinations-new-who-unicef-data-shows [accessed 2023-02-13]
15. Field epidemiology training programs. The Eastern Mediterranean Public Health Network (EMPHNET). URL: https://emphnet.net/en/our-work/the-center-of-excellence-for-applied-epidemiology/field-epidemiology-training-programs/ [accessed 2023-11-02]

16. FETP activities in response to coronavirus disease 19 (COVID-19). TEPHINET. URL: https://www.tephinet.org/fetp-activities-in-response-to-coronavirus-disease-19-covid-19 [accessed 2023-02-13]
17. Global Health Security Agenda homepage. Global Health Security Agenda. URL: https://www.ghsagenda.org/packages [accessed 2023-02-13]
18. Mala P, Abubakar A, Takeuchi A, Buliva E, Husain F, Malik MR, et al. Structure, function and performance of Early Warning Alert and Response Network (EWARN) in emergencies in the Eastern Mediterranean Region. Int J Infect Dis 2021 Apr;105:194-198 [FREE Full text] [doi: 10.1016/j.ijid.2021.02.002] [Medline: 33556613]
19. Thomas S, Sagan A, Larkin J, Cylus J, Figueras J, Karanikolos M. Strengthening health systems resilience: key concepts and strategies. POLICY BRIEF 36. URL: https://apps.who.int/iris/bitstream/handle/10665/332441/Policy-brief%2036-1997-8073-eng.pdf?sequence=1&isAllowed=y [accessed 2021-12-22]
20. Wilson K, von Tigerstrom B, McDougall C. Protecting global health security through the International Health Regulations: requirements and challenges. CMAJ 2008 Jul 01;179(1):44-48 [FREE Full text] [doi: 10.1503/cmaj.080516] [Medline: 18556329]
21. Kandel N, Chungong S, Omaar A, Xing J. Health security capacities in the context of COVID-19 outbreak: an analysis of International Health Regulations annual report data from 182 countries. Lancet 2020 Mar;395(10229):1047-1053. [doi: 10.1016/s0140-6736(20)30553-5]
22. One health. World Health Organization. URL: https://www.who.int/news-room/questions-and-answers/item/one-health [accessed 2023-02-13]
23. Tadmouri GO, Mandil A, Rashidian A. Biomedical and health research geography in the Eastern Mediterranean Region. East Mediterr Health J 2019 Nov 04;25(10):728-743 [FREE Full text] [doi: 10.26719/emhj.19.082] [Medline: 31774139]
24. WHO/UNICEF. Progresses and challenges with sustaining and advancing immunization coverage during the COVID-19 pandemic. World Health Organization. 2021 Jul 15. URL: https://www.who.int/publications/i/item/progresses-and-challenges-with-sustaining-and-advancing-immunization-coverage-during-the-covid-19-pandemic [accessed 2023-02-13]
25. Universal health coverage. World Health Organization. URL: http://www.emro.who.int/about-who/regional-director/universal-health-coverage.html [accessed 2023-02-13]
26. Alwan A, Shideed O, Siddiqi S. Essential public health functions: the experience of the Eastern Mediterranean Region. East Mediterr Health J 2016 Dec 12;22(9):694-700 [FREE Full text] [doi: 10.26719/2016.22.9.694] [Medline: 27966772]
27. Assessment of Essential Public Health Functions in Countries of the Eastern Mediterranean Region. Geneva: World Health Organization; 2017.
28. Brennan R, Hajjeh R, Al-Mandhari A. Responding to health emergencies in the Eastern Mediterranean region in times of conflict. Lancet 2022 Apr;399(10332):e20-e22. [doi: 10.1016/s0140-6736(20)30069-6]
29. Mostafavi E, Ghasemian A, Abdinasir A, Nematollahi Mahani SA, Rawaf S, Salehi Vaziri M, et al. Emerging and re-emerging infectious diseases in the WHO eastern Mediterranean region, 2001-2018. Int J Health Policy Manag 2021 Mar 06;11(8):1286-1300 [FREE Full text] [doi: 10.34172/ijhpm.2021.13] [Medline: 33904695]
30. Beyond COVID-19: a whole of health look at impacts during the pandemic response. Center For Global Development. URL: https://www.cgdev.org/sites/default/files/PP177-Beyond-COVID-scoping-paper.pdf [accessed 2022-01-01]
31. Maintaining essential health services: operational guidance for the COVID-19 context: interim guidance. World Health Organization. 2020 Jun 1. URL: https://www.who.int/publications/i/item/WHO-2019-nCoV-essential_health_services-2020.2 [accessed 2022-01-05]
32. Ho K, Al-Shorjabji N, Brown E, Zelmer J, Gabor N, Maeder A, et al. Applying the resilient health system framework for universal health coverage. Stud Health Technol Inform 2016;231:54-62. [Medline: 27782016]
33. Kawachi I, Berkman LF. Social cohesion, social capital, and health. In: Social Epidemiology. New York: Oxford University Press; 2000:174.

Abbreviations

CPD: continuing professional development
EMPHNET: Eastern Mediterranean Public Health Network
EMR: Eastern Mediterranean Region
EOC: emergency operations center
EPHF: essential public health function
EPI: Expanded Program on Immunization
EWAR: early warning and response
FETP: Field Epidemiology Training Program
GOI: Government of Iraq
HIS: health information system
IHR: International Health Regulations
KRG: Kurdistan Regional Government

PHC: primary health care
PHEOC: public health emergency operations center
POE: points of entry
RRT: rapid response team
SOP: standard operating procedure
UHC: universal health coverage
UNRWA: United Nations Relief and Works Agency for Palestine Refugees in the Near East
WHO: World Health Organization

Edited by A Mavragani; submitted 12.01.22; peer-reviewed by S Doan, A Kibballi Madhukeshwar, P Nguyen, S Afifi, R Guo; comments to author 06.10.22; revised version received 16.10.22; accepted 17.01.23; published 09.03.23.

Please cite as:
The Eastern Mediterranean Public Health Network (EMPHNET)
Toward Public Health Resilience in the Eastern Mediterranean Region: Findings From the Seventh Eastern Mediterranean Public Health Network Regional Conference
Interact J Med Res 2023;12:e36356
URL: https://www.i-jmr.org/2023/1/e36356
doi:10.2196/36356
PMID:36892927

Viewpoint

Neck Collar Assessment for People Living With Motor Neuron Disease: Are Current Outcome Measures Suitable?

Samuel D J Spears[1,2], BSc, MSc; Yusuf F Abdulle[2,3], BSc, MSc; Dionisios Korovilas[4], MSc; Ryo Torii[5], PhD; Deepak M Kalaskar[1], PhD; Nikhil Sharma[2,3], PhD

[1]Institute of Orthopaedics and Musculoskeletal Science, Division of Surgery and Interventional Sciences, University College London, London, United Kingdom

[2]National Hospital for Neurology and Neurosurgery, London, United Kingdom

[3]Department for Clinical and Movement Neurosciences, University College London, London, United Kingdom

[4]R2 Factory, Rolls-Royce, London, United Kingdom

[5]Department of Mechanical Engineering, University College London, London, United Kingdom

Corresponding Author:
Deepak M Kalaskar, PhD
Institute of Orthopaedics and Musculoskeletal Science
Division of Surgery and Interventional Sciences
University College London
Royal Free Hospital
10 Pond Street
London, NW3 2PS
United Kingdom
Phone: 44 207 794 0500
Email: d.kalaskar@ucl.ac.uk

Abstract

A majority of people living with motor neuron disease (MND) experience weakness of the neck and as a result, experience head drop. This exacerbates problems with everyday activities (eating, talking, breathing, etc). Neck collars are often used to support head drop; however, these are typically designed for prehospitalization settings to manage and brace the cervical region of the spine. As a result, it has been recorded that people living with MND often reject these collars for a variety of reasons but most notably because they are too restricting. The current standardized outcome measures (most notably restricting cervical range of motion) used for neck collars are summarized herein along with whether they are suitable for a bespoke neck collar specifically designed for people living with MND.

(Interact J Med Res 2023;12:e43274) doi:10.2196/43274

KEYWORDS

motor neuron disease; outcome measures; neck collar; bespoke orthoses; 3D scanning

Introduction

Motor neuron disease (MND) is a neurodegenerative disorder that contributes to weakness in the limbs and respiratory and bulbar muscle strength. The disease is irreversible and leads to a fatal outcome typically due to respiratory failure. People living with MND often develop neck weakness. They are unable to keep their head upright, resulting in a clinical syndrome known as a "head drop." This exacerbates issues with swallowing, breathing, communicating, eating, and drinking. To aid in the management of these symptoms, neck collars are typically used. We will investigate the current issues associated with collars used by people living with MND.

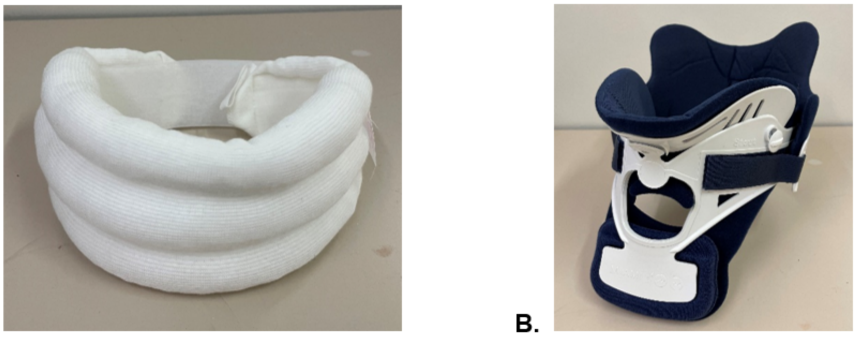
Neck collars have been well established for immobilization [1]. The primary functions include restriction of cervical spine motion, provision of spinal stability, and reduction of pain [2]. Initially, neck orthoses were designed for prehospitalization settings but are now increasingly used to reduce pain and mimic a sense of security for the patient, and to improve postsurgical outcomes [3-5]. Neck collars are used in nontrauma situations; for example, in neurodegenerative diseases such as MND. Despite the redeployment of neck collars in the management of MND, many are rejected by people living with MND due to their restriction on the cervical range of motion (CROM), which can result in unintended outcomes [6,7]. For example, possible muscle atrophy due to complete immobilization of the head accelerates the degradation of muscle tissue [8]. Current neck collars are both an uncomfortable experience and risk-exacerbating issues for people living with MND [7].

The difference in the needs of people living with MND using neck collars and the current aims of standardized collars have resulted in a low uptake of collars by people living with MND [7]. This suggests that outcome measures currently used to assess current neck collars may not be suitable for the design of a new bespoke collar for people living with MND. Therefore, creation of new assessment criteria, specifically for people living with MND, may be necessary. The current use of neck collars will be explored in trauma and MND, as well as outcome measures used to assess its efficacy, and based on the findings, we shall suggest a new way to assess the design of a new bespoke neck collar for people living with MND.

Trauma Neck Collars and Requirements

Neck collars can be classified into 2 types: soft and rigid (Figure 1). Soft collars are designed to be minimally restrictive and to provide the user with a more natural range of motion. These are commonly prescribed to support patients with neck pain and whiplash injuries and are typically made from a thick foam or rubber covered in fabric. Rigid collars are used to restrict range of motion as much as possible, typically applied in severe neck injuries, where there may be a suspected fracture or to stabilize the neck after surgery. These usually consist of a plastic outer shell and a padded inner liner.

Figure 1. Soft and rigid cervical orthoses: (A) Soft orthosis (Hereford) and (B) rigid orthosis (Miami J).



Neck collars are typically used to immobilize the spine for suspected spinal injury in prehospital settings as outlined by the National Institute for Health and Care Excellence (NICE) and the Joint Royal Colleges Ambulance Liaison Committee [9,10]. It is recommended by the Joint Royal Colleges Ambulance Liaison Committee and advanced trauma life support that a semirigid collar should be deployed when [9,11] a high-risk factor for cervical spine injury is identified and indicated by the Canadian C-spine rule, and a low-risk factor for cervical spine injury is identified and indicated by the Canadian C-spine rule and the person is unable to actively rotate his/her neck 45° left and right, whereby the collar is only used to stabilize and restrict motion of the cervical and upper thoracic region to prevent further complications from arising.

MND Neck Collar Requirements

For people living with MND, the NICE guideline for MND (NG42) states that a person experiencing muscle problems ought to be referred to orthotic services as soon as possible and for orthotics to be supplied [12]. The Motor Neurone Disease Association's (MNDA's) neck support information sheet outlines that people living with MND can experience different levels of discomfort related to immobility. Thus, it is important that any collar fitted must not create pressure points [13]. A person living with MND may be assessed for a collar by a physiotherapist, occupational therapist, or an orthotist. For many, it will be necessary to try a number of collars on, as problems associated with neck weakness vary due to disease progression, and with current "off the shelf collars," it is unlikely that one will address all these problems [13].

The MNDA's *Head supports for motor neurone disease* information sheet [13] states that (1) neck weakness is only part of the problem and is frequently associated with weakness of the shoulder girdle and long back extensor muscles; (2) many people with MND experience swallowing problems as a result of bulbar weakness, and a collar with an anterior area cut away may make swallowing easier; (3) forehead bands give freedom around the chin, mouth, and throat, making it easier to eat, drink, and speak; and (4) practical, easy-to-develop solutions, such as a roll of foam under the chin with a Velcro fastening, can offer some relief in certain circumstances.

For people living with MND requiring wheelchairs and experiencing neck weakness, it is suggested that the preferred position for the person is being tilted with the head, back, and neck supported; therefore, the interaction between the collar and chair must not interfere with one another [13]. It is clear that the requirements for people living with MND experiencing neck weakness differ from prehospitalization applications of neck collars, where MND applications are geared more toward support rather than restriction as needed for trauma.

Current Neck Collars for People Living With MND

The current head supports suggested by the MNDA are Soft collar, Wheelchair head supports, HeadUp Collar (Sheffield Support Snood), Hereford, Headmaster, Miami J, and Hensinger [13]. However, the most commonly prescribed collars for people living with MND include Aspen Vista, Philadelphia, Headmaster, HeadUp, Miami J, and Hereford (Figure 2).

Many of the collars used by people living with MND are rigid with the Hereford and Headmaster being soft and semirigid, respectively. The Headmaster collar (Figure 3) offers a unique design, with a chin rest supported by a semirigid frame and a strap around the neck to hold the orthosis in place. This acts to prevent head drop in the forward plane and does not support the head in other directions, and is often used in conjunction with a headrest for wheelchair users. Baxter et al [14] conducted a survey that investigated MND participants' experience with existing neck collars, which identified the following themes: "Difficulty fitting," "lack of physical support," "overly restrictive," "uncomfortable," and "unsuitable."

Figure 2. Motor Neurone Disease Association–recommended collars. (A) HeadUp collar, (B) Hereford, (C) Headmaster, (D) Burnett vacuum neck and head supports, (E) Hensinger, and (F) Miami J cervical collar [13].

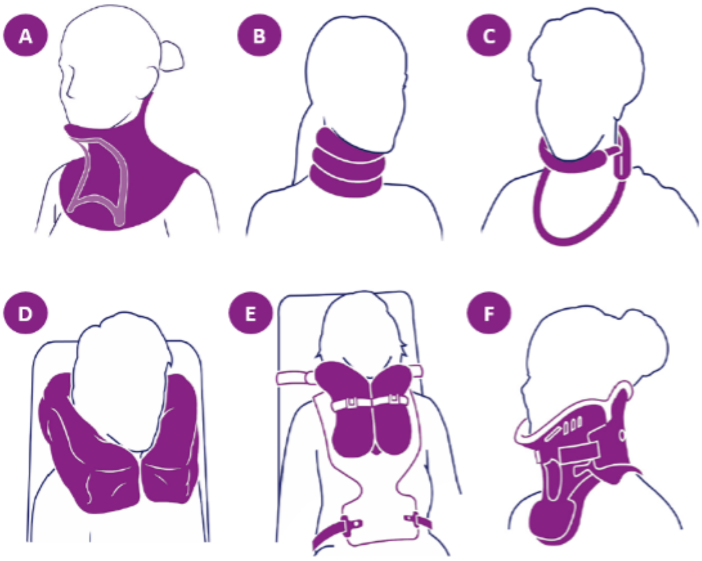


Figure 3. Headmaster collar.

First Orthosis Designed for People Living With MND: HeadUp

The only commercially available collar that has been specifically designed for people living with MND is HeadUp (TalarMade), otherwise known as the Sheffield Support Snood. The HeadUp collar was not assessed by its ability to restrict CROM and was quantitatively measured by assessment of ratio movement coupling (used to normalize the movements of angular velocities in the various planes), to establish control in performing head movements, and angular velocity, to establish whether the collar would compensate (support) the head movements without limiting natural movement velocity [14]. The collar was qualitatively evaluated via interviews or questionnaires against the following criteria:

- Level of support and range of motion
- Appearance
- Fitting the collar
- Breathing, eating, and Swallowing
- Perspiration

- Usage
- Overall satisfaction

While the consensus for the collar was positive, feedback from the interviews conducted by Baxter et al [14] highlighted that eating with the collar on presented issues, with 2 (of 16) participants stating they "can't possibly eat with it on" and "It makes my swallow harder as it presses on my Adam's apple," respectively. Regarding its fit, 2 participants reported that they had issues with the fitting of the collar, 3 reported that they would prefer to use their previous collars, and 1 stated that none of the available collars were satisfactory. Some initial training and practice were required by carers to fit the collar properly. Regarding the questions related to appearance and perspiration, the most frequent answer was neither positive nor negative. It was reported that when the usage of the HeadUp collar was queried among the same cohort of patients, there was no significant difference in the number of hours the collar was used compared to that among participants with a previous collar [14].

This study showed that when designing a collar specifically for people living with MND with neck weakness, the main requirement for the collar was to support head drop without interfering in other daily activities. As such, it may be difficult to design a collar that will be accepted by all, highlighting the importance of direct patient input in the design and outlining requirements for the collar. This can be attributed to differences in disease progression among patients; some report neck weakness only in the sagittal plane while others report weakness in the transverse plane. Also, the degree of neck weakness is due to disease progression, varying from no weakness to severe with the latter, implying that patients are unable to lift their head up whatsoever [14]. Therefore, outcome measures to capture the efficacy of a collar need to allow for a high variance of symptoms, usage, and application.

Bespoke Collar for People Living With MND

There currently is no fully bespoke neck collars designed for people living with MND, with the closest being the HeadUp, offering custom adjustments to the collar via adjustable supportive strips. A possible alternative is the development of a 3D-printed collar. 3D printing technology allows for the creation of custom-fit, comfortable, and functional orthotic devices. Another major benefit of 3D printing is that it can greatly reduce the lead time and cost associated with traditional methods of manufacturing orthotic devices and at a fraction of the cost of traditional methods [15,16]. The development of a new 3D-printed collar will aim to be fully bespoke, incorporating patient's anatomy captured via 3D scanning in the design process for each collar. It will also aim to be novel with the inclusion of force sensors to monitor the forces and pressures experienced by the user as this will provide insight into the progression of "head drop" and associated discomfort and whether a new collar is required to be printed to reflect these changes. This novel design will investigate the feasibility of using additive manufacturing and 3D scanning as a viable way to deliver a solution to improve the quality of life for those living with MND and experiencing head drop. 3D scanning can be used to create a detailed model of the patient's anatomy, which can be used to design a custom-fitted orthosis. This technology can also be used to create a digital model of the orthosis, which can be used to test its fit and function before it is manufactured. Finally, by using 3D printing and 3D scanning to design a custom neck collar, outcome measures used to assess the custom collars will be able to capture unique feedback based on each wearer's "use case," providing a more accurate efficacy analysis.

Collar Efficacy Assessment

The efficacy of neck collars has been suggested by several publications with regard to their ability to limit CROM [17]. As neck collars are typically designed for prehospital trauma settings to brace and manage the spine after a spinal injury, surgery, or degenerative changes. These collars are aimed at protecting and providing stability by reducing cervical motion. There are currently a variety of different products available on the market, and a summary of the various studies investigating the effectiveness and outcome measures of these collars is provided in Multimedia Appendix 1 [2,4,5,18-32].

The majority (14/18) of the studies compared sagittal (flexion and extension), transverse (lateral bending), and axial (rotation) planes of motion with and then without an orthosis to see the difference it makes in different planes. Different measurement systems were used to capture the data including goniometric, electromagnetic, optoelectronic, and video fluoroscopy. The subjects recruited for these studies were healthy adults (age range 18-67 years) with no previous history of cervical discomfort or weakness, previous spinal procedures, or pathological condition. All of the studies mentioned, apart from James et al [18] and Schneider et al [19], investigated flexion and extension, lateral bending, and axial rotation movements in a seated position. James et al [18] investigated movements in the supine position and Schneider et al [19] in an upright position.

Only a few studies investigated the impacts of neck orthoses using measures other than CROM. Tescher et al [20] and Plaisier et al [21] highlighted pressure ulcers as a consequence of wearing a collar for extended periods. Worsley et al [22] reported that elevated contact pressures were due to collar design at the device-skin interface with an observed inflammatory response to these increased pressures. Bell et al [23] reported that ill-fitted neck orthoses were unable to provide appropriate restriction of CROM and that orthoses that are too large or too small may cause neck impairment or increase the risk of complications. The lack of variety in collar types and their sizes has meant that there is an increased risk for patients to receive an unsuitable neck collar. Miller et al [5] observed that the efficacy of an orthosis may be reduced if the neck height position is not set correctly and may cause skin-related issues and potential hyperextension.

Collar comfort is also a key factor, as perceived comfort will affect patient compliance to wear and use the collar. Several studies [4,19,22,24] investigated collar comfort in healthy volunteers, with participants asked to rate perceived comfort on a ranking score. A summary of these studies can be found in Multimedia Appendix 2. There currently are no studies comparing collar comfort with people living with MND; therefore, these studies may prove more favorable for collars that immobilize head movement as immobilization may be considered better for aid in pain management.

Other studies suggest that the deployment of neck collars in trauma situations may not provide as much support as previously thought, with some cases actually increasing the risk of medical complications [1]. Extended use of ill-fitting neck collars can lead to increased intracranial and cerebrospinal fluid pressure, causing complications such as pressure ulcers and delirium, and with prolonged collar use, ventilator-associated pneumonia [33]. Therefore, collars should only be deployed for patients with unstable spines, which is difficult to identify in emergency situations [1].

Efficacy in neck collars has traditionally been assessed for trauma and general spine management applications, which emphasizes the need for restriction of head movement, whereas collar needs for people living with MND focus more on support rather than restriction. Applying outcome measures that purely assess CROM restriction are not suitable for collars aimed for people living with MND. However, measuring metrics such as pressure caused by extended periods of use with a collar may prove beneficial in assessing not only the perceived collar comfort by wearers but also good clinical measures that look to prevent pressure ulcers and increased intracranial and cerebrospinal fluid pressure.

Outcome Measures in Bespoke Ankle Foot Orthoses

The most common orthoses that can be made bespoke are ankle foot orthoses (AFOs) [34]. Foot and ankle problems have meant that older adults display a decreased ability to undertake daily tasks, showing cases of worsening balance, gait, increase in fall risks, and poor health-related quality of life, leading to the development of bespoke AFOs [35]. To better understand and define outcome measures that would be suitable to assess the efficacy of a bespoke neck collar, the relationship between the requirement of bespoke AFOs and the outcome measures used to assess them is investigated.

The Canadian Agency for Drugs and Technologies reported that bespoke AFOs are more effective than prefabricated orthoses when using biomechanical assessments as an objective outcome measure, such as dynamic balance, pressure relief, and load redistribution across plantar regions [36-39].

Heinemann et al [40] performed a survey with orthotists and physiotherapists to establish perspectives of quality-of-life care indicators, for people with bespoke AFOs. In this survey, 60% of participants (461 orthotists and 153 physiotherapists) stated that patient-reported outcome measures were preferred for aspects relating to quality-of-life topics, whereas clinicians were the preferred source for performance-based measures. It was further reported that the current standard assessment instruments are considered “good” by most respondents. The standard assessment measurements in the survey were predominately performance-based. However, it was noted that was there was a conflict between current standard outcome measures and outcome measures preferred by survey respondents, showing a preference for more patient-reported outcome measures to be reported. A separate survey with 257 physiotherapists found that commonly used outcome measures used for bespoke AFOs were a mixture between self-reported outcome measures and performance-based measures: pain assessments, functional tests, and range of motion [41].

A pilot study conducted by Aprile et al [42] investigated the effects of a custom AFO. The outcome measures used were a mixture of performance and patient-reported outcome measures with performance measures consisting of walking performance, stabilometric assessment, and disability; patient-orientated tools consisting of the short-form 36-item questionnaire, the North American Spine Society questionnaire, and the visual analog scale.

Shale [43] reported that patient experiences may be a passable indicator for clinical quality; however, clinical quality indicators may not translate to patient satisfaction. This can be due to complex associations among care assessments, expectations of care, patient knowledge, and objective measures that represent benefit. Wolf et al [44] further agrees that patient satisfaction does not equal patient expectations as expectations vary on a case-by-case basis; therefore, it is important not to alienate other quantitative outcome measures.

Finally, upon understanding how outcome measures are used to assess bespoke AFOs, it is clear that while patient-reported measures aid in understanding quality-of-life metrics for the patient, performance-based measures are still required to provide a clinical and quantitative insight into device performance. Therefore, to assess the design of a bespoke 3D-printed neck collar for people living with MND, there should be both a mixture of patient-reported and performance-based measures to capture both quality-of-life metrics and clinical metrics.

Discussion

The purpose of this viewpoint article is to describe outcome measures that are currently used to evaluate neck collars and to highlight that current outcome measures for collars are not suitable when applied to the design of a new bespoke collar for people living with MND. The HeadUp collar was the first neck orthosis that was designed specifically for people living with MND to aid in head drop. It focused on performance-based outcome measures to determine efficacy as well as some patent-outcome reported measures.

People living with MND often experience varying levels of progression with the disease; therefore, collar needs differ among patients. It is probable that for people living with MND, the efficacy of the neck orthoses should include patient-reported outcome measures to capture these needs. This highlights that current standardized outcome measures, primarily focusing on restriction for neck orthoses, are not applicable for people living with MND. A combination of new patient-based and performance-based outcome measures for collars designed to suit the individual needs of people living with MND are needed. Patient-focused outcome measures would be suitable in tandem with performance-based outcome measures, which are unable to report patient satisfaction and experience.

For the design of a new bespoke neck collar whose aims are to support and aid in the management of head drop symptoms in people living with MND, the outcome measures used to assess the efficacy should capture individual participants' responses compared with needs, whereby the collar is assessed on an individual use case compared with a previous collar or no collar (if they have not used one). This would capture participants' intended use for a collar—with collar use varying case by case—and effectively determine whether or not the collar meets expectations by the user. To do this, it is suggested that patient-reported assessments would be effective in capturing this feedback, by using visual analogue scales and questionnaires along with performance-based measures such as pressure caused by the collar. This would help to not only quantify patient feedback with pressure-associated discomfort but also reduce the risk of pressure-related injuries.

Conclusions

In conclusion, the current outcome measures used to assess the efficacy of neck collars are not suitable for a bespoke neck collar design. Bespoke orthoses, in general, should take care when outlining their outcome measures as the application and its end user will vary accordingly. Therefore, further work should be conducted to investigate the relationship between the variance experienced by patient expectations and outcome measures used for orthoses.

Acknowledgments

Authors would like to thank EPSRC-DTP and UCL Institute of Healthcare Engineering for supporting PhD studentship. This study represents independent research supported by the National Institute for Health and Care Research (NIHR) Biomedical Research Centre at The Royal Marsden NHS Foundation Trust and the Institute of Cancer Research, London. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Conflicts of Interest

NS is a co-founder and holds shares in BioCorteX Ltd.

Multimedia Appendix 1

Studies comparing methods and outcome measures for cervical collars (search conducted on Web of Science using "cervical collars" OR "neck collars" AND "outcome measures").

[PDF File (Adobe PDF File), 105 KB - ijmr_v12i1e43274_app1.pdf]

Multimedia Appendix 2

Averaged collar comfort scores from [4,18,21,23] (100% being most comfortable). N = No. of studies, n = Sum of No. of participants.

[PNG File , 39 KB - ijmr_v12i1e43274_app2.png]

References

1. Horodyski M, DiPaola CP, Conrad BP, Rechtine GR. Cervical collars are insufficient for immobilizing an unstable cervical spine injury. J Emerg Med 2011 Nov;41(5):513-519. [doi: 10.1016/j.jemermed.2011.02.001] [Medline: 21397431]
2. Miller CP, Bible JE, Jegede KA, Whang PG, Grauer JN. Soft and rigid collars provide similar restriction in cervical range of motion during fifteen activities of daily living. Spine 2010;35(13):1271-1278. [doi: 10.1097/brs.0b013e3181c0ddad]
3. Stone M, Tubridy C, Curran R. The effect of rigid cervical collars on internal jugular vein dimensions. Acad Emerg Med 2010 Jan;17(1):100-102 [FREE Full text] [doi: 10.1111/j.1553-2712.2009.00624.x] [Medline: 20015105]
4. Karason S, Reynisson K, Sigvaldason K, Sigurdsson GH. Evaluation of clinical efficacy and safety of cervical trauma collars: differences in immobilization, effect on jugular venous pressure and patient comfort. Scand J Trauma Resusc Emerg Med 2014 Jun 06;22(1):37 [FREE Full text] [doi: 10.1186/1757-7241-22-37] [Medline: 24906207]
5. Miller CP, Bible JE, Jegede KA, Whang PG, Grauer JN. The effect of rigid cervical collar height on full, active, and functional range of motion during fifteen activities of daily living. Spine 2010;35(26):E1546-E1552. [doi: 10.1097/brs.0b013e3181cf6f73]
6. Meadows J, Armijo-Olivo S, Magee D. Cervical Spine. In: Magee DJ, Zachazewski JE, Quillen WS, Manske RC, editors. Pathology and Intervention in Musculoskeletal Rehabilitation. Amsterdam: Elsevier; 2016:63-118.
7. Reed H, Langley J, Stanton A, Heron N, Clarke Z, Judge S, et al. Head-Up; An interdisciplinary, participatory and co-design process informing the development of a novel head and neck support for people living with progressive neck muscle weakness. J Med Eng Technol 2014 Oct 09;39(7):404-410. [doi: 10.3109/03091902.2015.1088092] [Medline: 26453038]
8. Prates A. Self-adjusting orthoses design. Mater Sci 2014.
9. National Clinical Guideline Centre (UK). Spinal Injury: Assessment and Initial Management. London: National Institute for Health and Care Excellence NICE; 2016.

10. Brown S, Kumar D, Millins M, Mark J. UK ambulance services clinical practice guidelines. Bridgwater: Class Professional Publishing; 2016.
11. American College of Surgeons, Committee on Trauma. Spine and spinal cord trauma. In: ATLS (Advanced Trauma Life Support) Manual. Chicago, IL: American College of Surgeons; 2012:174-205.
12. National Clinical Guideline Centre (UK). Motor neurone disease: assessment and management. London: National Institute for Health and Care Excellence (UK); 2016.
13. Head supports for motor neurone disease. Motor Neurone Disease Association. URL: https://www.mndassociation.org/app/uploads/Head-supports-information.pdf [accessed 2023-02-10]
14. Baxter S, Reed H, Clarke Z, Judge S, Heron N, Mccarthy A, et al. Evaluating a novel cervical orthosis, the Sheffield Support Snood, in patients with amyotrophic lateral sclerosis/motor neuron disease with neck weakness. Amyotroph Lateral Scler Frontotemporal Degener 2016 Feb 26;17(5-6):436-442 [FREE Full text] [doi: 10.3109/21678421.2016.1148170] [Medline: 26915274]
15. Boolos M, Corbin S, Herrmann A, Regez B. 3D printed orthotic leg brace with movement assist. Annals of 3D Printed Medicine 2022 Aug;7:100062. [doi: 10.1016/j.stlm.2022.100062]
16. Yoo H, Lee S, Kim J, Park C, Lee B. Development of 3D-printed myoelectric hand orthosis for patients with spinal cord injury. J Neuroeng Rehabil 2019 Dec 30;16(1):162 [FREE Full text] [doi: 10.1186/s12984-019-0633-6] [Medline: 31888695]
17. Maschmann C, Jeppesen E, Rubin MA, Barfod C. New clinical guidelines on the spinal stabilisation of adult trauma patients - consensus and evidence based. Scand J Trauma Resusc Emerg Med 2019 Aug 19;27(1):77 [FREE Full text] [doi: 10.1186/s13049-019-0655-x] [Medline: 31426850]
18. James CY, Riemann BL, Munkasy BA, Joyner AB. Comparison of cervical spine motion during application among 4 rigid immobilization collars. J Athl Train 2004 Jun;39(2):138-145 [FREE Full text] [Medline: 15173864]
19. Schneider AM, Hipp JA, Nguyen L, Reitman CA. Reduction in head and intervertebral motion provided by 7 contemporary cervical orthoses in 45 individuals. Spine (Phila Pa 1976) 2007 Jan 01;32(1):E1-E6. [doi: 10.1097/01.brs.0000251019.24917.44] [Medline: 17202874]
20. Tescher AN, Rindflesch AB, Youdas JW, Jacobson TM, Downer LL, Miers AG, et al. Range-of-motion restriction and craniofacial tissue-interface pressure from four cervical collars. J Trauma 2007 Nov;63(5):1120-1126. [doi: 10.1097/TA.0b013e3180487d0f] [Medline: 17993960]
21. Plaisier B, Gabram SGA, Schwartz RJ, Jacobs LM. Prospective evaluation of craniofacial pressure in four different cervical orthoses. J Trauma 1994 Nov;37(5):714-720. [doi: 10.1097/00005373-199411000-00004] [Medline: 7966467]
22. Worsley PR, Stanger ND, Horrell AK, Bader DL. Investigating the effects of cervical collar design and fit on the biomechanical and biomarker reaction at the skin. MDER 2018 Mar;Volume 11:87-94. [doi: 10.2147/mder.s149419]
23. Bell KM, Frazier EC, Shively CM, Hartman RA, Ulibarri JC, Lee JY, et al. Assessing range of motion to evaluate the adverse effects of ill-fitting cervical orthoses. Spine J 2009 Mar;9(3):225-231. [doi: 10.1016/j.spinee.2008.03.010] [Medline: 18504164]
24. Langley J, Pancani S, Kilner K, Reed H, Stanton A, Heron N, et al. A comfort assessment of existing cervical orthoses. Ergonomics 2018 Feb 28;61(2):329-338 [FREE Full text] [doi: 10.1080/00140139.2017.1353137] [Medline: 28697682]
25. Barati K, Arazpour M, Vameghi R, Abdoli A, Farmani F. The Effect of Soft and Rigid Cervical Collars on Head and Neck Immobilization in Healthy Subjects. Asian Spine J 2017 Jun;11(3):390-395 [FREE Full text] [doi: 10.4184/asj.2017.11.3.390] [Medline: 28670406]
26. Evans NR, Hooper G, Edwards R, Whatling G, Sparkes V, Holt C, et al. A 3D motion analysis study comparing the effectiveness of cervical spine orthoses at restricting spinal motion through physiological ranges. Eur Spine J 2013 Jan 4;22(S1):10-15. [doi: 10.1007/s00586-012-2641-0]
27. Whitcroft KL, Massouh L, Amirfeyz R, Bannister GC. A comparison of neck movement in the soft cervical collar and rigid cervical brace in healthy subjects. J Manipulative Physiol Ther 2011 Feb;34(2):119-122. [doi: 10.1016/j.jmpt.2010.12.007] [Medline: 21334544]
28. Hostler D, Colburn D, Seitz SR. A comparison of three cervical immobilization devices. Prehosp Emerg Care 2009;13(2):256-260. [doi: 10.1080/10903120802706195] [Medline: 19291567]
29. Zhang S, Wortley M, Clowers K, Krusenklaus JH. Evaluation of efficacy and 3D kinematic characteristics of cervical orthoses. Clin Biomech (Bristol, Avon) 2005 Mar;20(3):264-269. [doi: 10.1016/j.clinbiomech.2004.09.015] [Medline: 15698698]
30. Gavin TM, Carandang G, Havey R, Flanagan P, Ghanayem A, Patwardhan AG. Biomechanical analysis of cervical orthoses in flexion and extension: a comparison of cervical collars and cervical thoracic orthoses. J Rehabil Res Dev 2003;40(6):527-537 [FREE Full text] [doi: 10.1682/jrrd.2003.11.0527] [Medline: 15077665]
31. Rosen PB, McSwain NE, Arata M, Stahl S, Mercer D. Comparison of two new immobilization collars. Ann Emerg Med 1992 Oct;21(10):1189-1195. [doi: 10.1016/s0196-0644(05)81744-5] [Medline: 1416295]
32. Aker PD, Randoll M, Rheault C, O'Connor, S. Restriction of neck flexion using soft cervical collars: a preliminary study. J Can Chiropr Assoc 1991 Sep;35(3):139-145 [FREE Full text]

33. Dunham CM, Brocker BP, Collier BD, Gemmel DJ. Risks associated with magnetic resonance imaging and cervical collar in comatose, blunt trauma patients with negative comprehensive cervical spine computed tomography and no apparent spinal deficit. Crit Care 2008;12(4):R89 [FREE Full text] [doi: 10.1186/cc6957] [Medline: 18625041]
34. Choo YJ, Chang MC. Commonly used types and recent development of ankle-foot orthosis: a narrative review. Healthcare (Basel) 2021 Aug 13;9(8):1046 [FREE Full text] [doi: 10.3390/healthcare9081046] [Medline: 34442183]
35. Wang C, Goel R, Rahemi H, Zhang Q, Lepow B, Najafi B. Effectiveness of daily use of bilateral custom-made ankle-foot orthoses on balance, fear of falling, and physical activity in older adults: a randomized controlled trial. Gerontology 2019 Nov 30;65(3):299-307 [FREE Full text] [doi: 10.1159/000494114] [Medline: 30504728]
36. Abbasi F, Bahramizadeh M, Hadadi M. Comparison of the effect of foot orthoses on Star Excursion Balance Test performance in patients with chronic ankle instability. Prosthet Orthot Int 2019 Feb;43(1):6-11. [doi: 10.1177/0309364618792718] [Medline: 30101681]
37. Redmond AC, Landorf KB, Keenan A. Contoured, prefabricated foot orthoses demonstrate comparable mechanical properties to contoured, customised foot orthoses: a plantar pressure study. J Foot Ankle Res 2009 Jun 16;2(1):20 [FREE Full text] [doi: 10.1186/1757-1146-2-20] [Medline: 19531262]
38. Caravaggi P, Giangrande A, Lullini G, Padula G, Berti L, Leardini A. In shoe pressure measurements during different motor tasks while wearing safety shoes: the effect of custom made insoles vs. prefabricated and off-the-shelf. Gait Posture 2016 Oct;50:232-238. [doi: 10.1016/j.gaitpost.2016.09.013] [Medline: 27662483]
39. Bus SA, Ulbrecht JS, Cavanagh PR. Pressure relief and load redistribution by custom-made insoles in diabetic patients with neuropathy and foot deformity. Clin Biomech (Bristol, Avon) 2004 Jul;19(6):629-638. [doi: 10.1016/j.clinbiomech.2004.02.010] [Medline: 15234488]
40. Heinemann AW, Fatone S, LaVela SL, Slater BCS, Deutsch A, Peterson M, et al. Orthotists' and physical therapists' perspectives on quality of care indicators for persons with custom ankle-foot orthoses. Assist Technol 2021 Jul 04;33(4):206-216. [doi: 10.1080/10400435.2019.1610814] [Medline: 31091177]
41. Grieve R, Palmer S. Physiotherapy for plantar fasciitis: a UK-wide survey of current practice. Physiotherapy 2017 Jun;103(2):193-200. [doi: 10.1016/j.physio.2016.02.002] [Medline: 27156704]
42. Aprile I, Bordieri C, Gilardi A, Lainieri Milazzo M, Russo G, De Santis F, et al. Balance and walking involvement in facioscapulohumeral dystrophy: a pilot study on the effects of custom lower limb orthoses. Eur J Phys Rehabil Med 2013 Apr;49(2):169-178 [FREE Full text] [Medline: 23138679]
43. Shale S. Patient experience as an indicator of clinical quality in emergency care. Clinical Governance: An Intl J 2013 Oct 14;18(4):285-292. [doi: 10.1108/cgij-03-2012-0008]
44. Wolf JA, Niederhauser V, Marshburn D, LaVela SL. Reexamining "Defining Patient Experience": the human experience in healthcare. Patient Exp J 2021 Apr 28;8(1):16-29. [doi: 10.35680/2372-0247.1594]

Abbreviations

AFO: ankle foot orthosis
CROM: cervical range of motion
MND: motor neuron disease
MNDA: Motor Neurone Disease Association
NICE: National Institute for Health and Care Excellence

Edited by T Leung; submitted 07.10.22; peer-reviewed by R Pathak, L Room; comments to author 31.01.23; revised version received 03.02.23; accepted 06.02.23; published 14.03.23.

Please cite as:
Spears SDJ, Abdulle YF, Korovilas D, Torii R, Kalaskar DM, Sharma N
Neck Collar Assessment for People Living With Motor Neuron Disease: Are Current Outcome Measures Suitable?
Interact J Med Res 2023;12:e43274
URL: https://www.i-jmr.org/2023/1/e43274
doi:10.2196/43274
PMID:36917160

Viewpoint

The Impact of Digital Health on Smoking Cessation

Raquel Cobos-Campos[1], PhD; Jose Aurelio Cordero-Guevara[1], PhD; Antxon Apiñaniz[1,2,3,4], MSc; Arantza Sáez de Lafuente[1], MSc; Cristina Bermúdez Ampudia[1], MSc; Julene Argaluza Escudero[1], MSc; Iraida Pérez Llanos[1,2], MSc; Naiara Parraza Diez[1,4], PhD

[1]Epidemiology and Public Health Group, Bioaraba Health Research Institute, Vitoria-Gasteiz, Spain

[2]Osakidetza Basque Health Service, Vitoria-Gasteiz, Spain

[3]Department of Preventive Medicine and Public Health, University of the Basque Country, Vitoria-Gasteiz, Spain

[4]Research Network on Chronicity, Primary Care and Health Promotion, Madrid, Spain

Corresponding Author:
Raquel Cobos-Campos, PhD
Epidemiology and Public Health Group
Bioaraba Health Research Institute
José Atxotegi s/n
Vitoria-Gasteiz, 01009
Spain
Phone: 34 680953032
Email: raquelcobos1976@gmail.com

Abstract

Background: Smartphones have become useful tools for medicine, with the use of specific apps making it possible to bring health care closer to inaccessible areas, continuously monitor a patient's pathology at any time and place, promote healthy habits, and ultimately improve patients' quality of life and the efficiency of the health care system. Since 2020, the use of smartphones has reached unprecedented levels. There are more than 350,000 health apps, according to a 2021 IQVIA Institute report, that address, among other things, the management of patient appointments; communication among different services or professionals; the promotion of lifestyle changes related to adopting healthy habits; and the monitoring of different pathologies and chronic conditions, including smoking cessation. The number of mobile apps for quitting smoking is high. As early as 2017, a total of 177 unique smoking cessation–relevant apps were identified in the iPhone App Store, 139 were identified in Google Play, 70 were identified in the BlackBerry app store, and 55 were identified in the Windows Phone Store, but very few have adequate scientific support. It seems clear that efforts are needed to assess the quality of these apps, as well as their effectiveness in different population groups, to have tools that offer added value to standard practices.

Objective: This viewpoint aims to highlight the benefits of mobile health (mHealth) and its potential as an adjuvant tool in health care.

Methods: A review of literature and other data sources was performed in order to show the current status of mobile apps that can offer support for smoking cessation. For this purpose, the PubMed, Embase, and Cochrane databases were explored between May and November 2022.

Results: In terms of smoking cessation, mHealth has become a powerful coadjuvant tool that allows health workers to perform exhaustive follow-ups for the process of quitting tobacco and provide support anytime and anywhere. mHealth tools are effective for different groups of smokers (eg, pregnant women, patients with chronic obstructive pulmonary disease, patients with mental illness, and the general population) and are cost-effective, generating savings for the health system. However, there are some patient characteristics that can predict the success of using mobile apps in the smoking cessation process, such as the lower age of patients, dependence on tobacco, the number of quit attempts, and the previous use of mobile apps, among others. Therefore, it is preferable to offer these tools to patients with a higher probability of quitting tobacco.

Conclusions: mHealth is a promising tool for helping smokers in the smoking cessation process. There is a need for well-designed clinical studies and economic evaluations to jointly assess the effectiveness of new interventions in different population groups, as well as their impact on health care resources.

(Interact J Med Res 2023;12:e41182) doi:10.2196/41182

KEYWORDS

smoking cessation; smoking; cessation; smoker; quit; care delivery; service delivery; health technology; mHealth; mobile applications; mobile health; digital health; mobile app; health app; smartphone; health service; eHealth; trend

Introduction

The health service sector is one of the most complex existing sectors [1], and although we are witnessing how new technologies are changing industries, business models, and markets in a disruptive way within just a few years or months, this complexity explains why the health sector is slower when it comes to adapting to this evolving environment [2]. The data we handle in health care are sensitive data that are subject to exhaustive data protection regulation, which makes accessing and expanding health care technology difficult. Nevertheless, strategies for the use of information and communication technologies in the health sector have been gaining ground, and there is now a majority consensus on the fundamental role of these technologies in improving the efficiency and accessibility of the health system [2]. Technological advances are changing all aspects of society, improving and speeding up processes with the aim of improving people's quality of life [2].

The health care sector is no stranger to these changes, and many technological innovations are making increasingly useful services and tools available [3]. Among these innovations, big data and artificial intelligence (AI) have become promising tools for the management of chronic diseases [4] and enable the use of innovative and promising diagnostic and therapeutic applications [5]. AI is understood as a working methodology for compiling an enormous amount of information (known as *big data*) in order to subsequently use powerful computer programs to try to obtain data on decision-making elements in all areas of life [5].

Web 3.0 is the next generation of internet technology that relies on the use of AI to process data and create a personalized user experience [6]. Given the large amount of information and metadata that are being generated and made publicly available, it is believed that Web 3.0 technologies (eg, machine learning, AI, Internet of Things, and natural language processing) will allow computer agents to automatically link any kind of data from any system to build inferences from those data [6].

In addition to the above, another current trend is the use of mobile health (mHealth), which is defined by the World Health Organization as the use of mobile devices, such as smartphones and patient monitoring devices, for medical practice and public health [7].

Smartphones have become useful tools for medicine, with the use of specific apps making it possible to bring health care closer to inaccessible areas, continuously monitor a patient's pathology at any time and place, promote healthy habits, and ultimately improve the quality and efficiency of the health care system [3].

This improvement of health care process quality is the result of several factors, such as the remote monitoring of patients, which makes it possible to predict potential problems early and take the necessary measures within a sufficient time frame, thereby reducing the number of unnecessary consultations and hospitalizations. This allows practitioners to focus on investing their time in solving important health problems that cannot be solved remotely [3].

One of the United Nations Sustainable Development Goals for 2030 is "to ensure healthy lives and promote well-being for all at all ages" [8]. However, the World Health Organization states that universal health coverage will not be achieved without the support of eHealth [9].

There are already more mobile devices than people in the world today [10]. According to the Global System Mobile Association, there are more than 9.5 billion mobile connections [11], while the worldwide population consists of more than 7.9 billion people [12].

Since 2020, the use of smartphones has reached unprecedented levels. In 2021, this use grew by 30% when compared to that in 2020. Moreover, according to App Annie's *State of Mobile 2022* report [13], 230,000 new apps were downloaded in 2021—a 5% increase from 2020. The amount of health-related mobile apps is starting to reach considerable numbers. There are currently more than 350,000 health apps, according to a 2021 report by the IQVIA Institute [14], aimed at, among other things, the promotion of lifestyle changes related to adopting healthy habits, the monitoring of different pathologies, the management of patient appointments, and communication among different services or professionals [15]. It is estimated that around 30% of health apps are targeted toward health care professionals and 70% are targeted toward the general population [15]. Self-care through mobile devices is another growing field [16].

Of these apps, very few have the necessary quality that should be demanded from such tools, which has resulted in greater noise and difficulty in selecting apps that can add value to people's lives [15,17-19]. The Andalusian Health Quality Agency developed extensive guidance on the criteria [20] that a good health app should meet. Some of these criteria are as follows: *relevance* (it is clearly defined what the app is for, what its objectives are, and who it is aimed at), *testing* (the app has been tested beforehand on the target audience), *transparency* (authors, funding sources, and conflicts of interest are clearly identified), *content and sources* (the health app is based on reliable sources and available scientific evidence and specifies when the information was last updated), and *risk management* (risks that may be associated with the use of the app are identified). It is necessary to involve end users (both health care professionals and patients) in app design to ensure greater quality and usability [20] and, once the app is designed, evaluate app efficacy. There is now ample evidence about the utility of mHealth in different contexts, such as increasing the rate of consultation attendance [21], promoting safer sex [22], monitoring patients with diabetes, managing low back pain [23,24], and treating smoking dependence. The National Institute for Health and Care Excellence (NICE) considers digital and mHealth interventions as options for helping people stop smoking and adjuncts to existing services [25]. The NICE also advises that text message–based interventions that use tailored messages may be more effective than other digital health and mHealth interventions [25].

Smoking is one of the main causes of global morbidity and mortality and a risk factor of a high number of chronic diseases, such as cancer, cardiovascular disease, and chronic obstructive pulmonary disease (COPD), among others. The life span of smokers is, on average, about 10 years shorter than that of nonsmokers. However, quitting smoking can increase life expectancy, and the number of years of life gained depends on the age at which a person quits smoking [26]. Tobacco kills more than 8 million people per year, of whom more than 7 million are direct users and about 1.2 million are nonsmokers exposed to secondhand smoke [27]. Furthermore, the adverse health consequences of tobacco are well known and have major economic implications [28]. The World Bank estimated that high-income countries spend 6% to 15% of their total health expenditure on tobacco-related diseases [29]. Smokers have higher rates of absenteeism and longer absences from work than those of nonsmokers due to the higher prevalence of tobacco-related diseases among smokers [28].

There are currently different approaches to smoking cessation treatment, such as more intensive or less intensive motivational counseling–based interventions and pharmacological therapy [30]. The mobile telephony boom and, in recent years, the increase in the number of mobile apps, with a penetration of 100% in the world population, have provided new tools for helping both professionals and patients in the management of different pathologies, including smoking cessation.

According to the Survey on Alcohol and Other Drugs in Spain (Encuesta Sobre Alcohol y Otras Drogas en España 2019-2020), the average age of smoking initiation is 16.6 years [31]. With 50% and 29% of the population accessing their first smartphone at 11 to 12 years of age and between 13 and 14 years of age, respectively, mHealth is a great alternative tool for preventing, reducing, or quitting smoking in these age groups [32].

There are a large number of smokers who prefer to not use drugs to quit smoking. It is therefore clear that it is necessary to carry out interventions that do not imply the need for pharmacological treatment and to have data on efficacy and efficiency that support their generalization to the smoking population. Other reviews that address this topic have been performed [33-35]. However, this viewpoint paper is not so much a review in itself, but rather an update on the state of digital health in general and its impact on smoking cessation programs. The aim of this viewpoint paper is to provide readers with an overview of the usefulness of digital health and, in particular, mHealth as adjuvant tools in smoking cessation programs.

Evidence of mHealth Focused on Smoking Cessation

There is enough scientific evidence about the great potential of using mHealth as a complement of usual treatment in smoking cessation.

Whittaker et al [33] reported a pooled relative risk of smoking cessation of 1.69 in a systematic review of 12 clinical trials that evaluated the efficacy of mobile phone–based interventions. Chen et al [34] concluded that interventions based on the internet, software, mobile phones, or other electronic tools increase the likelihood of quitting tobacco when compared to no intervention or the use of generic self-help material. Dahne et al [36] assessed asynchronous smoking cessation e-visits that were performed proactively through the electronic health records of adult smokers who were treated within primary care. After 3 months, e-visit participants, when compared with usual treatment participants, were 4.13 (95% CI 1.06-16.10; *P*=.04) times more likely to have reduced their number of cigarettes smoked per day by at least 50%. In a recent systematic review that evaluated the efficacy of digital interventions in randomized clinical trial studies of smoking cessation, 19 trials (15,472 participants) were included in the analysis, and the overall abstinence rate (percentage of participants who did not smoke during a follow-up period of at least 3 months) at the end point was 17.8% (95% CI 17%-18.7%); the authors concluded that digital health had a clear positive effect when compared to self-help guidelines or no intervention [35].

By analyzing results among different groups of patients, overall, studies have demonstrated that smoking cessation apps are feasible for use among people diagnosed with mental illness, especially those with a high score on the System Usability Scale [37]—a reliable tool for measuring the ease of use of a wide variety of products and services, including hardware, software, mobile devices, websites, and applications. However, it appears clearly that apps designed specifically for patients with schizophrenia or other mental illnesses may be more accessible and user-friendly [38], emphasizing the importance of end user involvement in app development.

In another systematic review that assessed the efficacy of mobile phone–based behavioral interventions in pregnancy to promote maternal and fetal health in high-income countries, the authors concluded that the utilization of mobile phone–based health behavior interventions in pregnancy demonstrates some correlation with positive beliefs, behaviors, and health outcomes [39].

These types of tools have also been proven to be effective in managing COPD. In a review published recently, the authors concluded that pharmacotherapy combined with behavioral interventions that are delivered via mHealth may be an effective, safe, accessible, and cost-effective strategy for helping smokers with COPD quit smoking [40].

Smoking cessation has become a ubiquitous intervention approach for which user engagement can be readily measured. Nearly 500 English-language smartphone apps for smoking cessation have been downloaded more than 33 million times since 2012 (R Nelson, Sensor Tower Inc, email, April 15, 2020). Higher user engagement in smartphone interventions for smoking cessation is predictive of cessation outcomes [41,42]. However, there are certain characteristics that either predict the time when an app will be used or predict that an app will not be used. The act of smoking up to one-half pack per day, the act of smoking the first cigarette within 5 minutes after waking, a higher mean acceptance of internal physical sensations, female sex, minority race (people of color), Hispanic ethnicity, and a history of smoking for 10 or more years are related to longer periods of app use [43].

Our research group obtained similar results in a randomized clinical trial conducted with 320 motivated smoking cessation patients and evaluated the effectiveness of a combined program (motivational counseling and reinforcement messages sent to mobile phones) versus motivational counseling alone (OR 2.329) at 12 months after baseline [44].

These results allowed the transfer of the combined program to clinical practice after transforming the messaging program into a corporate app, which is available in all health centers of the Basque Public Health System (*Vive sin tabaco* app [Figure 1]). In parallel, a cost-effectiveness study was carried out to justify this transfer; the incremental cost-effectiveness ratio was calculated, with cost savings (from a societal perspective) of €5398 (US $5885.98) and €3290 (US $3587.42) per quality-adjusted life year gained for men and women, respectively [45]. In addition, a further cohort study was carried out with 92 patients who initiated a quit attempt with the *Vive sin tabaco* app, which showed smoking cessation results that were very similar to those of the previous clinical trial (14.1% vs 16.5% at 12 months) [46].

Quitting smoking can be a difficult challenge that sometimes requires many attempts before success is achieved. Nicotine dependence is a complex disorder [47]. However, the earlier smoking cessation occurs, the higher the number of life years regained [26]; therefore, it is of vital importance to encourage young people to not take up smoking and ensure that they internalize the benefits of not smoking. mHealth tools for smoking cessation have great potential for this age group.

Smoking cessation treatment is not only clinically effective but also cost-effective. Health advice is considered one of the most cost-effective interventions in the treatment of smoking [48]; however, the changes promoted by health advice do not last long [49]. Therefore, it is necessary to establish reinforcement mechanisms, among which are information and communication technologies and, more specifically, mHealth, for which there is ample evidence in the treatment of smoking.

Figure 1. The *Vive sin tabaco* app.

Discussion

Principal Findings

Mobile technology has changed the way we live, work, and communicate. The use of mobile technologies to support the achievement of mHealth goals is an emerging and rapidly developing field that has the potential to play a key role in transforming health care to increase the quality and efficiency of care, and the mission of this field is to complement rather than replace traditional health care [50]. Health developments mainly include apps aimed directly or indirectly at maintaining or improving people's healthy behaviors, quality of life, and well-being [50].

Health care is transforming. Health care costs are rising, as health care must cope with the demand for increasingly personalized and long-term care. Moreover, it is estimated that the use of mobile apps could improve the efficiency of patient care and reduce the time spent accessing and analyzing information by up to 30% [51]. In fact, a study that was recently published in the *Journal of Medical Economics* concluded that patients who use digital health tools can reduce their monthly medical costs by around 22% [52].

Health apps are becoming technological tools with great potential for improving the way chronic diseases are managed. If they are well designed and focus on the needs of patients, they could more effectively facilitate the management of health care resources and communication between professionals and patients, thereby enhancing the active role of the population in their self-care [15].

Mobile apps have great potential to support patients in health care and encourage healthy behavioral changes. However, it is the features of apps that determine patients' attitudes toward the use of apps, which in turn determine the success of apps [53]. Therefore, it is necessary that end users are involved at the beginning of the design process in order to increase the usability of apps.

Research has shown that demographics and personality characteristics are associated with the adoption and use of mobile apps. Income and level of education correlate positively with mobile phone use, whereas age correlates negatively with it [53]. The procurement of mHealth tools by older people is limited [54], and almost half (43%) of those aged over 70 years stop using them within the first 14 days [55] mainly due to the complexity of the tools [55], the limited health knowledge of users (ie, the knowledge required to fully understand the data) [55], and the cost of the technology [55]. This further emphasizes the need to design simple, end user–oriented tools and to involve end users in their design in order to obtain tools that fully meet end users' expectations.

The number of quit attempts, nicotine dependence, the previous use of digital aids to quit smoking, and the Fagerstrom test score correlate with smokers' attitudes toward the use of a smoking cessation app. However, different studies have found no significant relationship between demographic characteristics and attitudes toward or intentions to use a smoking cessation app [56,57]. Further, 77.5% of smokers who have used a mobile app to quit smoking have never checked the credibility of the developer or publisher of the health app [57].

It is clear that it is necessary to invest money and effort into having useful tools in health care that meet the expectations of end users and complement health care to increase the quality and efficiency of care. Mobile telephony was born to bring people together, but its purpose is now much greater than that.

Conclusions

It seems clear that mHealth is a valuable tool that can provide support to both health professionals and patients in the complex process of smoking cessation. However, it is likely that several mHealth user characteristics predict the likelihood of the success of smoking cessation apps, such as age, tobacco dependence, and the number of cigarettes smoked per day, among others. Therefore, it would make sense to offer these apps to, for example, younger smokers, those who are more dependent on tobacco, and those who smoke more than half a pack of cigarettes per day.

Conflicts of Interest

None declared.

References

1. Institute of Medicine and National Academy of Engineering. 6 next steps: Aligning policies with leadership opportunities. In: Engineering a Learning Healthcare System: A Look at the Future: Workshop Summary. Washington, DC: The National Academies Press; 2011:271-282.
2. Federación de Empresas de Tecnologías Sanitarias. Hacia la transformación digital del sector de la salud. Sociedad Española de Informática de la Salud. URL: https://seis.es/wp-content/uploads/2018/02/LA-TRANSFORMACION-DIGITAL-DEL-SECTOR-SALUD-EN-ESPA%C3%91A.pdf [accessed 2022-07-05]
3. Ambit-BST. mHealth: todo lo que debes saber sobre la salud móvil. Ambit-BST. 2021 Feb 11. URL: https://www.ambit-bst.com/blog/mhealth-todo-lo-que-debes-saber-sobre-la-salud-m%C3%B3vil [accessed 2022-07-05]
4. Majnarić LT, Babič F, O'Sullivan S, Holzinger A. AI and big data in healthcare: Towards a more comprehensive research framework for multimorbidity. J Clin Med 2021 Mar 14;10(4):766 [FREE Full text] [doi: 10.3390/jcm10040766] [Medline: 33672914]
5. Kedra J, Gossec L. Big data and artificial intelligence: Will they change our practice? Joint Bone Spine 2020 Mar;87(2):107-109. [doi: 10.1016/j.jbspin.2019.09.001] [Medline: 31520738]
6. Curchoe CL, Malmsten J, Bormann C, Shafiee H, Farias AFS, Mendizabal G, et al. Predictive modeling in reproductive medicine: Where will the future of artificial intelligence research take us? Fertil Steril 2020 Dec;114(5):934-940 [FREE Full text] [doi: 10.1016/j.fertnstert.2020.10.040] [Medline: 33160516]
7. The Competitive Intelligence Unit. mHealth: Garantizar una vida sana y bienestar para todos. Squarespace. URL: https://static1.squarespace.com/static/587fdc951b10e30ca5380172/t/619e96eb4e24b11376dcf83f/1637783276593/The+CIU+-+WP+Series+2021-III+MHealth+v12+ESP.pdf [accessed 2022-07-05]
8. Department of Economic and Social Affairs. Goal 3. United Nations. URL: https://sdgs.un.org/goals/goal3 [accessed 2022-07-05]
9. Global Observatory for eHealth. Global diffusion of eHealth: Making universal health coverage achievable. World Health Organization. 2016. URL: https://bit.ly/3xBGsH5 [accessed 2022-07-05]
10. Fernández R. Número de líneas de telefonía móvil por cada 100 habitantes en España 2016-2022. Statista. 2022 Oct 13. URL: https://es.statista.com/estadisticas/477127/tasa-penetracion-telefonia-movil-mensual-espana/ [accessed 2022-07-05]
11. GSMA. About us. GSMA. URL: https://www.gsma.com/aboutus/ [accessed 2022-03-07]
12. Worldometer. Población mundial. Worldometer. URL: https://www.worldometers.info/es/poblacion-mundial/ [accessed 2022-07-05]

13. data.ai. State of mobile 2022. data.ai. URL: https://www.data.ai/en/go/state-of-mobile-2022/ [accessed 2022-07-15]
14. Olsen E. Digital health apps balloon to more than 350,000 available on the market, according to IQVIA report. MobiHealthNews. 2021 Aug 04. URL: https://www.mobihealthnews.com/news/digital-health-apps-balloon-more-350000-available-market-according-iqvia-report [accessed 2022-07-15]
15. Iborra CR. Las aplicaciones móviles de salud como herramientas de apoyo a la autogestión de cuidados del paciente crónico. Universidad Autónoma de Madrid. 2019 May. URL: https://repositorio.uam.es/bitstream/handle/10486/687937/rey_iborra_cristinatfg.pdf?sequence=1 [accessed 2022-07-05]
16. Fundación Telefónica. Sociedad Digital en España 2020-2021. Fundación Telefónica. 2021 May. URL: https://www.fundaciontelefonica.com/cultura-digital/publicaciones/sociedad-digital-en-espana-2020-2021/730/#close [accessed 2022-07-05]
17. Salud Conectada. Introducción, normativa y certificación de las Apps de Salud. Salud Conectada. URL: https://www.saludconectada.com/2-1-introduccion-normativa-certificacion-las-apps-salud/ [accessed 2022-07-05]
18. Grau I, Kostov B, Gallego JA, Iii FG, Fernández-Luque L, Sisó-Almirall A. [Assessment method for mobile health applications in Spanish: The iSYScore index]. Semergen 2016;42(8):575-583. [doi: 10.1016/j.semerg.2015.12.001] [Medline: 26879598]
19. PICKASO Team. Informe: El Uso de las Apps en España y en el Mundo en 2018. PICKASO. 2018 Jul 23. URL: https://pickaso.com/2018/uso-apps-espana-y-mundo-2018 [accessed 2022-03-07]
20. Agencia de Calidad Sanitaria de Andalucía. Recomendaciones. Estrategia de calidad y seguridad en aplicaciones móviles para la salud. URL: http://www.calidadappsalud.com/recomendaciones/ [accessed 2022-12-05]
21. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: Systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
22. Gold J, Aitken CK, Dixon HG, Lim MSC, Gouillou M, Spelman T, et al. A randomised controlled trial using mobile advertising to promote safer sex and sun safety to young people. Health Educ Res 2011 Oct;26(5):782-794. [doi: 10.1093/her/cyr020] [Medline: 21447750]
23. Larsen ME, Turner J, Farmer A, Neil A, Tarassenko L. Telemedicine-supported insulin optimisation in primary care. J Telemed Telecare 2010;16(8):433-440. [doi: 10.1258/jtt.2010.100103] [Medline: 20841384]
24. Rintala A, Rantalainen R, Kaksonen A, Luomajoki H, Kauranen K. mHealth apps for low back pain self-management: Scoping review. JMIR Mhealth Uhealth 2022 Aug 26;10(8):e39682 [FREE Full text] [doi: 10.2196/39682] [Medline: 36018713]
25. National Institute for Health and Care Excellence. Tobacco: preventing uptake, promoting quitting and treating dependence. National Institute for Health and Care Excellence. URL: https://www.nice.org.uk/guidance/ng209 [accessed 2023-01-31]
26. Jha P, Ramasundarahettige C, Landsman V, Rostron B, Thun M, Anderson RN, et al. 21st-century hazards of smoking and benefits of cessation in the United States. N Engl J Med 2013 Jan 24;368(4):341-350 [FREE Full text] [doi: 10.1056/NEJMsa1211128] [Medline: 23343063]
27. World Health Organization. Tabaco. World Health Organization. 2022 May 25. URL: https://www.who.int/es/news-room/fact-sheets/detail/tobacco [accessed 2022-07-05]
28. Suárez-Bonel MP, Villaverde-Royo MV, Nerín I, Sanz-Andrés C, Mezquida-Arno J, Córdoba-García R. Health care costs and work absenteeism in smokers: study in an urban community. Arch Bronconeumol 2015 Dec;51(12):615-620. [doi: 10.1016/j.arbres.2015.05.001] [Medline: 26198012]
29. The W. Curbing the epidemic: governments and the economics of tobacco control. The World Bank. Tob Control 1999;8(2):196-201 [FREE Full text] [doi: 10.1136/tc.8.2.196] [Medline: 10478406]
30. Banegas JR, Díez-Gañán L, Bañuelos-Marco B, González-Enríquez J, Villar-Álvarez F, Martín-Moreno JM, et al. [Smoking-attributable deaths in Spain, 2006]. Med Clin (Barc) 2011 Feb 12;136(3):97-102. [doi: 10.1016/j.medcli.2010.03.039] [Medline: 20980030]
31. Owens T. At what age did you get your first smartphone? Statista. 2022 Jun 13. URL: https://www.statista.com/statistics/1076293/age-obtaining-first-smartphone-teenagers-france/ [accessed 2022-07-05]
32. Ministerio de Sanidad. La Encuesta sobre alcohol y otras drogas en España, EDADES. Plan Nacional sobre Drogas. URL: https://pnsd.sanidad.gob.es/profesionales/sistemasInformacion/sistemaInformacion/encuestas_EDADES.htm [accessed 2022-07-05]
33. Whittaker R, McRobbie H, Bullen C, Borland R, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2012 Nov 14;11:CD006611. [doi: 10.1002/14651858.CD006611.pub3] [Medline: 23152238]
34. Chen YF, Madan J, Welton N, Yahaya I, Aveyard P, Bauld L, et al. Effectiveness and cost-effectiveness of computer and other electronic aids for smoking cessation: a systematic review and network meta-analysis. Health Technol Assess 2012;16(38):1-205, iii-v [FREE Full text] [doi: 10.3310/hta16380] [Medline: 23046909]
35. Sha L, Yang X, Deng R, Wang W, Tao Y, Cao H, et al. Automated digital interventions and smoking cessation: Systematic review and meta-analysis relating efficiency to a psychological theory of intervention perspective. J Med Internet Res 2022 Nov 16;24(11):e38206 [FREE Full text] [doi: 10.2196/38206] [Medline: 36383408]

36. Dahne J, Player M, Carpenter MJ, Ford DW, Diaz VA. Evaluation of a proactive smoking cessation electronic visit to extend the reach of evidence-based cessation treatment via primary care. Telemed J E Health 2021 Mar;27(3):347-354 [FREE Full text] [doi: 10.1089/tmj.2020.0167] [Medline: 33085578]
37. Usability.gov. System Usability Scale (SUS). Usability.gov. URL: https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html [accessed 2022-12-29]
38. Sawyer C, Hassan L, Guinart D, Agulleiro LM, Firth J. Smoking cessation apps for people with schizophrenia: How feasible are m-Health approaches? Behav Sci (Basel) 2022 Aug 01;12(8):265 [FREE Full text] [doi: 10.3390/bs12080265] [Medline: 36004836]
39. Hussain T, Smith P, Yee LM. Mobile phone-based behavioral interventions in pregnancy to promote maternal and fetal health in high-income countries: Systematic review. JMIR Mhealth Uhealth 2020 May 28;8(5):e15111 [FREE Full text] [doi: 10.2196/15111] [Medline: 32463373]
40. Feng L, Lv X, Wang Y, Chu S, Dai Z, Jing H, et al. Developments in smoking cessation interventions for patients with chronic obstructive pulmonary disease in the past 5 years: a scoping review. Expert Rev Respir Med 2022 Jul;16(7):749-764. [doi: 10.1080/17476348.2022.2108797] [Medline: 35916493]
41. Browne J, Halverson TF, Vilardaga R. Engagement with a digital therapeutic for smoking cessation designed for persons with psychiatric illness fully mediates smoking outcomes in a pilot randomized controlled trial. Transl Behav Med 2021 Sep 15;11(9):1717-1725 [FREE Full text] [doi: 10.1093/tbm/ibab100] [Medline: 34347865]
42. Bricker JB, Mull KE, Kientz JA, Vilardaga R, Mercer LD, Akioka KJ, et al. Randomized, controlled pilot trial of a smartphone app for smoking cessation using acceptance and commitment therapy. Drug Alcohol Depend 2014 Oct 01;143:87-94 [FREE Full text] [doi: 10.1016/j.drugalcdep.2014.07.006] [Medline: 25085225]
43. Bricker JB, Mull KE, Santiago-Torres M, Miao Z, Perski O, Di C. Smoking cessation smartphone app use over time: Predicting 12-month cessation outcomes in a 2-arm randomized trial. J Med Internet Res 2022 Aug 18;24(8):e39208 [FREE Full text] [doi: 10.2196/39208] [Medline: 35831180]
44. Cobos-Campos R, de Larrinoa AAF, de Lafuente Moriñigo AS, Diez NP, Barandiaran FA. Effectiveness of text messaging as an adjuvant to health advice in smoking cessation programs in primary care. A randomized clinical trial. Nicotine Tob Res 2017 Aug 01;19(8):901-907. [doi: 10.1093/ntr/ntw300] [Medline: 27838659]
45. Cobos-Campos R, Mar J, Apiñaniz A, de Lafuente AS, Parraza N, Aizpuru F, et al. Cost-effectiveness analysis of text messaging to support health advice for smoking cessation. Cost Eff Resour Alloc 2021 Feb 15;19(1):9 [FREE Full text] [doi: 10.1186/s12962-021-00262-y] [Medline: 33588885]
46. Cobos-Campos R, Apiñaniz A, de Lafuente AS, Parraza N. Development, validation and transfer to clinical practice of a mobile application for the treatment of smoking. Aten Primaria 2022 Jul;54(7):102363 [FREE Full text] [doi: 10.1016/j.aprim.2022.102363] [Medline: 35636019]
47. Arias AC. Dependencia de Nicotina: aproximacion a su manejo farmacologico. Rev Colomb Psiquiatr 2002 Mar;31(1):67-72 [FREE Full text]
48. López MJT, Pérez AMB, Barceló IB, Ortíz JMB, Galvis JG, Pérez MG, Grupo de Atención al Tabaquismo (GAT) de SmuMFyC. Actitud de los profesionales de Atención Primaria frente al tabaco. Semergen 2008 Mar;34(3):113-118. [doi: 10.1016/S1138-3593(08)71862-8]
49. Cummings SR, Rubin SM, Oster G. The cost-effectiveness of counseling smokers to quit. JAMA 1989 Jan 06;261(1):75-79. [doi: 10.1001/jama.1989.03420010085038] [Medline: 2491762]
50. Gazdecki A. 9 mobile technology trends For 2017 (infographic). Bizness Apps. URL: https://www.biznessapps.com/blog/mobile-technology-trends/ [accessed 2022-07-05]
51. Haskins BL, Lesperance D, Gibbons P, Boudreaux ED. A systematic review of smartphone applications for smoking cessation. Transl Behav Med 2017 Jun;7(2):292-299 [FREE Full text] [doi: 10.1007/s13142-017-0492-2] [Medline: 28527027]
52. Whaley CM, Bollyky JB, Lu W, Painter S, Schneider J, Zhao Z, et al. Reduced medical spending associated with increased use of a remote diabetes management program and lower mean blood glucose values. J Med Econ 2019 Sep;22(9):869-877 [FREE Full text] [doi: 10.1080/13696998.2019.1609483] [Medline: 31012392]
53. Chevalking SKL, Allouch SB, Brusse-Keizer M, Postel MG, Pieterse ME. Identification of users for a smoking cessation mobile app: Quantitative study. J Med Internet Res 2018 Apr 09;20(4):e118 [FREE Full text] [doi: 10.2196/jmir.7606] [Medline: 29631988]
54. Peek STM, Luijkx KG, Vrijhoef HJM, Nieboer ME, Aarts S, van der Voort CS, et al. Understanding changes and stability in the long-term use of technologies by seniors who are aging in place: a dynamical framework. BMC Geriatr 2019 Aug 28;19(1):236 [FREE Full text] [doi: 10.1186/s12877-019-1241-9] [Medline: 31462214]
55. Puri A, Kim B, Nguyen O, Stolee P, Tung J, Lee J. User acceptance of wrist-worn activity trackers among community-dwelling older adults: Mixed method study. JMIR Mhealth Uhealth 2017 Nov 15;5(11):e173 [FREE Full text] [doi: 10.2196/mhealth.8211] [Medline: 29141837]
56. Kruse CS, Mileski M, Moreno J. Mobile health solutions for the aging population: A systematic narrative analysis. J Telemed Telecare 2017 May;23(4):439-451 [FREE Full text] [doi: 10.1177/1357633X16649790] [Medline: 27255207]

57. BinDhim NF, McGeechan K, Trevena L. Who uses smoking cessation apps? A feasibility study across three countries via smartphones. JMIR Mhealth Uhealth 2014 Feb 06;2(1):e4 [FREE Full text] [doi: 10.2196/mhealth.2841] [Medline: 25098439]

Abbreviations

AI: artificial intelligence
COPD: chronic obstructive pulmonary disease
mHealth: mobile health
NICE: National Institute for Health and Care Excellence

Edited by A Mavragani; submitted 19.07.22; peer-reviewed by N Kute, A Coca, M Desimone; comments to author 22.11.22; revised version received 05.12.22; accepted 03.01.23; published 15.03.23.

Please cite as:
Cobos-Campos R, Cordero-Guevara JA, Apiñaniz A, de Lafuente AS, Bermúdez Ampudia C, Argaluza Escudero J, Pérez Llanos I, Parraza Diez N
The Impact of Digital Health on Smoking Cessation
Interact J Med Res 2023;12:e41182
URL: https://www.i-jmr.org/2023/1/e41182
doi:10.2196/41182
PMID:36920468

Viewpoint

Strengthening the One Health Approach in the Eastern Mediterranean Region

Ekhlas Hailat[1], MSc, DVM; Mirwais Amiri[1], MPH, MD; Nitish Debnath[2], MSci, DVM, PhD; Mahmudur Rahman[1], MBBS, MPHM, PhD; Md Nurul Islam[1], MVM, DVM; Zahida Fatima[3], DVM, PhD; Yousef Khader[4], BDS, MSci, MSPH, SCD; Mohannad Al Nsour[1], MD, MSci, PhD

[1]Global Health Development | Eastern Mediterranean Public Health Network, Amman, Jordan

[2]Development Alternatives Incorporated, Dhaka, Bangladesh

[3]Animal Sciences Division, Pakistan Agricultural Research Council, Islamabad, Pakistan

[4]Department of Public Health, Faculty of Medicine, Jordan University of Science and Technology, Irbid, Jordan

Corresponding Author:
Mirwais Amiri, MPH, MD
Global Health Development | Eastern Mediterranean Public Health Network
Shmeisani, Abdallah Ben Abbas Street
Bldg No 42
Amman, 11196
Jordan
Phone: 962 792985349
Email: mamiri@globalhealthdev.org

Abstract

One Health aims to use a multidisciplinary approach to combat health threats at animal, human, and environmental health interfaces. Among its broad focus areas are issues related to food safety, the control of zoonoses, laboratory services, neglected tropical diseases, environmental health, biosafety and biosecurity, and combatting antimicrobial resistance. A roundtable session was conducted on November 18, 2021, as part of the Eastern Mediterranean Public Health Network's (EMPHNET) seventh regional conference to highlight what role Global Health Development (GHD)|EMPHNET can play to strengthen the One Health approach. This viewpoint summarizes the findings of the roundtable discussion to highlight the experts' viewpoints on strengthening the One Health approach, including the extent of zoonotic diseases and the dynamics of pathogens and emerging diseases; the occurrence of antimicrobial-resistant pathogens as a silent pandemic; issues surrounding the globalization of trade and food safety; the importance of integrated solutions as a new norm; issues around the institutionalization and governance toward effective operationalization of the One Health approach in the region; and how the One Health approach can be operationalized at global, regional, and local levels. The panel concluded that One Health is an integrated unifying approach that aims to sustainably balance and optimize the health of people, animals, and ecosystems, and provided recommendations to strengthen the One Health approach. It also discussed how GHD|EMPHNET can play its role in transferring the concept of One Health from theory to practice via a solid operationalization road map guide at the Eastern Mediterranean region level. The five broad priority areas of this operational guide include (1) establishing and strengthening a governance architecture, legal framework, and policy and advocacy structure for One Health operationalization in the region; (2) fostering coordination, communication, and collaboration for One Health actions across the region and beyond; (3) building the workforce capacity for effective One Health operationalization in the region; (4) supporting regional platforms for timely, effective, and efficient data sharing and exchange on all One Health–related issues; and (5) supporting risk communication, behavior change communication, and community engagement efforts in the region.

(Interact J Med Res 2023;12:e41190) doi:10.2196/41190

KEYWORDS

One Health; operationalization; zoonosis; antimicrobial resistance, Eastern Mediterranean region countries

Introduction

One Health has been defined by the World Health Organization (WHO) as “an approach to designing and implementing programs, policies, legislation, and research in which multiple sectors communicate and work together to achieve better public health outcomes” [1]. The importance of One Health is ever apparent in recent decades, as rapidly increasing human populations in the 21st century have led to encroachment into new geographic areas. Marked changes in climate and land use such as deforestation and intensive farming practices have resulted in more people living in close contact with domestic and wild animals [2]. Recent reports have revealed that more than 25% of original forest cover has been lost, and 75% of terrestrial environments and 66% of marine environments were severely altered by human interventions [3]. This close contact between animals and their environments creates increasing opportunities for the spilling over of pathogens between animals and people, and the rise of new diseases (ie, emerging infectious diseases [EIDs]). Furthermore, the movement of people, animals, and animal products has grown due to advances in international travel and trade, allowing these EIDs to spread easily across borders and around the world [4].

For effective prevention, detection, and response to EIDs or zoonotic outbreaks, communication, coordination, and collaboration among experts from all relevant fields, including public, animal, and environmental health professionals, working closely to share data and expertise is needed [1]. The WHO, Food and Agriculture Organization of the United Nations (FAO), and the World Organisation for Animal Health, formerly the Office International des Epizooties (OIE), have led the charge in promoting multisectoral responses to these issues and other public health threats at the human-animal-ecosystem interface [1,5,6].

In this regard, Global Health Development (GHD)|Eastern Mediterranean Public Health Network (EMPHNET) strongly believes in the effective role of One Health in responses and actions at the animal-human-ecosystem interface, especially targeting emerging and endemic zoonoses, and commends the role of FAO-OIE-WHO Tripartite to create and support One Health programs. A roundtable session was conducted on November 18, 2021, as part of EMPHNET’s seventh regional conference to highlight what role GHD|EMPHNET can play to strengthen the One Health approach. This viewpoint summarizes the findings of the roundtable discussion to highlight the drivers, integrated solutions, and success stories regarding the implementation of One Health, and highlight the role that GHD|EMPHNET can play in transferring the concept of One Health from theory to practice.

Roundtable Panel Discussion

The panel members discussed the extent of zoonotic diseases and the dynamics of pathogens and emerging diseases. It was highlighted that nearly 75% of all new or EIDs affecting humans at the beginning of the 21st century are of zoonotic origin [4]. Of those, 71.8% are reported to have genetic origins from wildlife, indicating increasing spillover in recent years [7]. Examples of the effects of increasing interconnectedness and the global impact of these diseases are the HIV/AIDS, severe acute respiratory syndrome (SARS), the H5N1 strain of avian influenza, and the 2009 H1N1 influenza virus pandemics. The speed by which these diseases emerge and spread causes serious economic and developmental concerns, in addition to their effects on public health. The emergence of these diseases had been concentrated in certain “hot spot” areas, like Central Africa, South and Southeast Asia, and Latin America, where compounding factors contribute to disease spread and highlight the need for the improvement of disease detection and response capacities in these countries [8].

Several infectious viruses have emerged or re-emerged from wildlife, generating serious threats to the global health and the global economy. Ebola and Marburg hemorrhagic fevers, Lassa fever, dengue fever, yellow fever, West Nile fever, Zika, and chikungunya vector-borne diseases, swine flu, Middle East respiratory syndrome (MERS), and the recent COVID-19 are additional examples of zoonoses that have spread internationally, causing significant impact and creating a need for rapid intervention from scientists and public health professionals [9]. In fact, evidence suggests that SARS, MERS, and COVID-19 must serve as a wake-up call to be better prepared when facing the coming onslaught of the pathogen [10].

The panel also discussed the occurrence of antimicrobial-resistant pathogens as a silent pandemic. Antimicrobial resistance (AMR) occurs when bacteria, viruses, fungi, and parasites no longer respond to medications, making infections more difficult to treat [11]. AMR is a prime example of a global public health threat, which requires urgent multisectoral action [11-13]. In fact, the WHO has declared that AMR is one of the top 10 global public health threats facing humanity, citing the misuse and overuse of antibiotics as the main drivers in the development of microbial drug resistance [13]. In addition, AMR poses a significant threat to world economies, increasing mortality and disability rates, increasing longer hospital stays, and creating a need for new drug developments. The most alarming aspect of AMR is that the reduced effectiveness of antibiotics and other antimicrobials may create a future in which major surgeries and cancer chemotherapies are considered too risky [13]. While antibiotic misuse in medicine has been addressed increasingly in recent years, abuse in the agricultural sector is massively neglected and more extensive. Indeed, the Food and Drug Administration reports revealed that more than 20 million pounds of antibiotics were sold for use in livestock farms in 2014 [14] and about 80% of medically important antibiotics are regularly fed to livestock in some countries [15].

Another important topic discussed during the panel revolved around the globalization of trade and food safety. It was highlighted that globalization of trade plays an important role in disease spread and food safety, posing a challenge to the public health sector [16]. FAO promotes One Health with a focus on food safety and security, sustainable agriculture, AMR, nutrition, animal and plant health, fisheries, and livelihoods. Ensuring a One Health approach is essential for progress to anticipate (early warning), prevent, detect, and control responses to diseases that spread between animals and humans; tackle

AMR; ensure food safety; prevent environment-related human and animal health threats; and combat many other health challenges arising at human-animal-environmental interfaces. Good practices from farm to fork represent a One Health approach to food safety [17].

The panel also emphasized the importance of integrated solutions as a new norm. It was highlighted that humans, animals, and environments are ever intertwined in the current globalized landscape. Therefore, a multidisciplinary approach is necessary to address any resultant emergence of zoonotic events. One Health presents a shift in the way we think about human and animal health and offers a new direction to tackle these issues, but successful implementation of interventions requires multisectoral collaboration, communication, and coordination, as well as integration. The silo mentality of certain institutions can impede the progress toward an integrated, inclusionary response and must be adjusted for effective action plans. This can be done by reframing One Health as a way to aid the smooth implementation of plans by offering a road for the cooperation of all relevant sectors and departments when handling any public health issue. When developing action plans or response programs, all concerned parties should be contacted, ideally creating a multidisciplinary team. The One Health teams can include health care providers, public health professionals, and epidemiologists representing the human health sector, and veterinarians, veterinary epidemiologists, para-veterinarians, farmers, and agriculture field workers representing animal health, in addition to ecologists and wildlife experts representing the environmental sector. One Health teams can also include law enforcement officers, social scientists, policy makers, and community members as needed to ensure effective collaboration and representation of any matter concerning the interaction at the animal-human-environment interface [18].

Another point discussed by the panel revolved around the importance of institutionalization and the governance toward effective operationalization of the One Health approach in the region. The FAO, OIE, WHO, and the United Nations Environment Program (UNEP) have advocated the new operational definition of One Health as recommended by their advisory panel, the One Health High Level Expert Panel (OHHLEP), on December 1, 2021 [19-21]. The new One Health definition developed by the OHHLEP states "One Health is an integrated, unifying approach that aims to sustainably balance and optimize the health of people, animals, and ecosystems" [18]. It is important to stress that institutionalization and governance of the One Health approach within the government systems is crucial to reflect the success of its implementation across the region. This will only be achieved if there is enough buy-in from the countries regarding the benefits of One Health application on the public, agricultural, and economic lives of their citizens and landscapes. Ultimately, the goal is to ensure that the One Health way of thinking is sustained in all governmental operations and entities working to prevent, prepare, detect, and respond to public health events and infectious diseases. This would ensure not only more efficient national action plans but also the prosperity of nations and meeting of sustainable developmental goals [21,22].

Finally, the panel discussed how the One Health approach can be operationalized at global, regional, and local levels. It was discussed that the Tripartite Zoonoses Guide (TZG), jointly developed by the FAO, OIE, and WHO to support countries in adopting a One Health approach to address zoonotic outbreaks, provides recommendations, options, and best practices, which can be used to assist countries in achieving sustainable methods for dealing with diseases with spillover potential. The TZG provides an operational guide for countries to develop the necessary capacities for preparedness for zoonotic events and efficient flow of information among concerned parties, even in low-resource settings [23]. Fortunately, One Health was adopted by the Group of Seven countries, Group of 20 countries, and World Health assemblies to reform public health, which may facilitate their implementation globally [23].

Recommendations and Key Action or Follow-up Areas

As a result of the panel discussion and the subsequent questions raised by the participants during the questions and answers session, the most essential One Health considerations and their implications specifically pertaining to the countries across the region were identified. Recommendations to strengthen the One Health approach include institutionalization and governance of the One Health strategy, securing political commitment and influencing policy changes, developing a One Health legal framework, establishing an effective coordination mechanism and promoting multisectoral collaboration, community engagement for breaking silos, and better understanding of the interconnectedness and interdependence of human-animal-ecosystem interfaces. On the other hand, it was emphasized that epidemiological data and laboratory information should be shared across sectors to ensure effective detection of and response to health threats. Joint responses to health threats should be implemented by trained One Health workforce across sectors at the local, national, regional, and global levels.

As a success story on the active follow-up of the panel session, an important immediate next step was to collect and synthesize all important discussion points and develop an operational guide specifically geared toward the One Health priorities of the countries across the region. Thus, in April 2022, GHD|EMPHNET developed a technical guide entitled "Operationalization of the One Health Approach in the Eastern Mediterranean Region" to serve as a road map for One Health operationalization at the regional level [24]. The guide took into consideration the operational definition of One Health [19-21] and the most important themes that emerged from the panel discussion. Thus, for the effective operationalization of the One Health approach across the region, the following five broad priority areas were outlined: (1) establishing and strengthening a governance architecture, legal framework, and policy and advocacy structure for One Health operationalization in the region; (2) fostering coordination, communication, and collaboration for One Health actions across the region and beyond; (3) building the workforce capacity for effective One Health operationalization in the region; (4) supporting regional platforms for timely, effective, and efficient data sharing and exchange on all One Health–related issues; and (5) supporting risk communication, behavior change communication, and community engagement efforts in the region. Each of these five broader strategic areas contains additional subcomponents to allow for the development of country-specific implementation action plans.

As a further follow-up of the panel discussion, GHD|EMPHNET will establish a regional committee on One Health. This committee will assume two main roles or dual functions: (1) ensure effective regional communication and coordination by serving primarily as a *liaison* between the countries in the region and relevant global entities involved in the One Health efforts and (2) ensure effective regional collaboration and capacity development by functioning as a *technical advisory, support, and oversight body* across the region to facilitate the One Health Quadripartite work in the region and ensure that all stakeholders across the region are actively engaged/involved in the One Health response. Further, GHD|EMPHNET will work with the countries and relevant One Health stakeholders in the region, using its documented road map [24], to create operational work plans for the countries within and overall operational framework for the region.

Conclusions

As promoted by the Quadripartite platform of the FAO, World Organisation for Animal Health, WHO, and UNEP, One Health is an integrated unifying approach that aims to sustainably balance and optimize the health of people, animals, and ecosystems. The health of humans, domestic and wild animals, plants, and the wider environment (including ecosystems) are closely linked and interdependent. To face new health challenges that emerge at the human-animal-environment interface, collaboration, coordination, communication, and concerted action between different sectors are needed, in addition to institutionalization and governance of the One Health approach. However, many countries lack the capacity to implement such collaboration, and international organizations, nongovernmental organizations, and private sectors can help these countries. In this context, GHD|EMPHNET can play an effective role in promoting and transferring the concept of One Health from theory to practice through its developed technical guide for the operationalization of the One Health approach at the regional level.

Conflicts of Interest

None declared.

References

1. One Health. World Health Organization. 2021. URL: https://www.who.int/news-room/questions-and-answers/item/one-health [accessed 2023-03-14]
2. One Health basics. Centers for Disease Control and Prevention. 2018. URL: https://www.cdc.gov/onehealth/basics/index.html [accessed 2023-03-14]
3. Zoonosis. Population Connection. 2021. URL: https://populationconnection.org/why-population/zoonosis/ [accessed 2023-03-14]
4. Everard M, Johnston P, Santillo D, Staddon C. The role of ecosystems in mitigation and management of COVID-19 and other zoonoses. Environ Sci Policy 2020 Sep;111:7-17 [FREE Full text] [doi: 10.1016/j.envsci.2020.05.017] [Medline: 32501392]
5. One Health. World Organisation for Animal Health. 2021. URL: https://www.oie.int/en/what-we-do/global-initiatives/one-health [accessed 2023-03-14]
6. EcoHealth Alliance. 2021. URL: https://www.ecohealthalliance.org/program/health-policy-initiatives [accessed 2023-03-14]
7. Plowright RK, Parrish CR, McCallum H, Hudson PJ, Ko AI, Graham AL, et al. Pathways to zoonotic spillover. Nat Rev Microbiol 2017 Aug;15(8):502-510 [FREE Full text] [doi: 10.1038/nrmicro.2017.45] [Medline: 28555073]
8. Mackenzie JS, Jeggo M. The One Health approach-why is it so important? Trop Med Infect Dis 2019 May 31;4(2):31 [FREE Full text] [doi: 10.3390/tropicalmed4020088] [Medline: 31159338]
9. Trovato M, Sartorius R, D'Apice L, Manco R, De Berardinis P. Viral emerging diseases: challenges in developing vaccination strategies. Front Immunol 2020;11:2130 [FREE Full text] [doi: 10.3389/fimmu.2020.02130] [Medline: 33013898]
10. Sinclair JR. Importance of a One Health approach in advancing global health security and the Sustainable Development Goals. Rev Sci Tech 2019 May;38(1):145-154. [doi: 10.20506/rst.38.1.2949] [Medline: 31564744]
11. Antimicrobial resistance. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance [accessed 2023-03-14]
12. Kreuder Johnson C, Hitchens PL, Smiley Evans T, Goldstein T, Thomas K, Clements A, et al. Spillover and pandemic properties of zoonotic viruses with high host plasticity. Sci Rep 2015 Oct 07;5:14830. [doi: 10.1038/srep14830] [Medline: 26445169]
13. Antimicrobial resistance. World Organisation for Animal Health. 2021. URL: https://www.woah.org/en/what-we-do/global-initiatives/antimicrobial-resistance/#ui-id-1 [accessed 2023-03-14]
14. Rep. Louise Slaughter says 80% of antibiotics are fed to livestock. PolitiFact. URL: https://www.politifact.com/factchecks/2013/oct/15/louise-slaughter/rep-louise-slaughter-says-80-antibiotics-are-fed-l/ [accessed 2023-03-14]

15. Antibiotics in our food system. FoodPrint. 2021. URL: https://foodprint.org/issues/antibiotics-in-our-food-system/ [accessed 2023-03-14]
16. Antimicrobial resistance. World Health Organization. 2021. URL: https://www.who.int/news-room/fact-sheets/detail/antimicrobial-resistance [accessed 2023-03-14]
17. Food safety and globalization of trade in food: a challenge to the public health sector. World Health Organization. 1998. URL: https://apps.who.int/iris/handle/10665/63880 [accessed 2022-06-29]
18. World Livestock 2011. Livestock in food security. Food and Agriculture Organization of the United Nations. URL: https://www.fao.org/policy-support/tools-and-publications/resources-details/en/c/1262785/ [accessed 2022-06-29]
19. Joint tripartite and UNEP statement on definition of "One Health". UN Environment Programme. 2021. URL: https://www.unep.org/news-and-stories/statements/joint-tripartite-and-unep-statement-definition-one-health [accessed 2022-02-19]
20. Tripartite and UNEP support OHHLEP's definition of "One Health". World Organisation for Animal Health. 2021. URL: https://www.oie.int/en/tripartite-and-unep-support-ohhleps-definition-of-one-health/ [accessed 2022-02-27]
21. Tripartite and UNEP support OHHLEP's definition of "One Health": Joint Tripartite (FAO, OIE, WHO) and UNEP Statement. World Health Organization. 2021. URL: https://www.who.int/news/item/01-12-2021-tripartite-and-unep-support-ohhlep-s-definition-of-one-health [accessed 2022-02-19]
22. One Health principles and concepts. AFROHUN. 2019. URL: https://afrohun.org/wp-content/uploads/2021/01/ONE-HEALTH-PRINCIPLES-AND-CONCEPTS.pdf [accessed 2023-03-14]
23. Tripartite Zoonoses Guide: operational tools and approaches for zoonotic diseases. World Health Organization. URL: https://www.who.int/initiatives/tripartite-zoonosis-guide [accessed 2022-06-29]
24. Toward the implementation of the One Health approach in the Eastern Mediterranean Region. EMPHNET. 2022. URL: https://emphnet.net/media/hb0b2ox0/operationalization-of-the-one-health-approach-in-the-eastern-mediterranean-region.pdf [accessed 2023-03-14]

Abbreviations

AMR: antimicrobial resistance
EID: emerging infectious disease
EMPHNET: Eastern Mediterranean Public Health Network
FAO: Food and Agriculture Organization of the United Nations
GHD: Global Health Development
MERS: Middle East respiratory syndrome
OHHLEP: One Health High Level Expert Panel
OIE: Office International des Epizooties
SARS: severe acute respiratory syndrome
TZG: Tripartite Zoonoses Guide
UNEP: United Nations Environment Program
WHO: World Health Organization

Edited by A Mavragani; submitted 18.07.22; peer-reviewed by M Raimi, MM Patwary, C Zhao, A Lai, A Elidrissi; comments to author 28.12.22; revised version received 17.01.23; accepted 07.02.23; published 21.03.23.

Please cite as:
Hailat E, Amiri M, Debnath N, Rahman M, Nurul Islam M, Fatima Z, Khader Y, Al Nsour M
Strengthening the One Health Approach in the Eastern Mediterranean Region
Interact J Med Res 2023;12:e41190
URL: https://www.i-jmr.org/2023/1/e41190
doi:10.2196/41190
PMID:36943329

Viewpoint

Evolution of Health Information Sharing Between Health Care Organizations: Potential of Nonfungible Tokens

Pouyan Esmaeilzadeh[1], PhD

Department of Information Systems and Business Analytics, College of Business, Florida International University, Miami, FL, United States

Corresponding Author:
Pouyan Esmaeilzadeh, PhD
Department of Information Systems and Business Analytics
College of Business
Florida International University
Modesto A Maidique Campus
11200 SW 8th St, RB 261B
Miami, FL, 33199
United States
Phone: 1 305 348 3302
Email: pesmaeil@fiu.edu

Abstract

This study attempts to explain the development and progress of the technology used for sharing health information across health care organizations (such as hospitals and physicians' offices). First, we describe the strengths and weaknesses of traditional sharing models, health information exchange (HIE), and blockchain-based HIE. Second, the potential use of nonfungible token (NFT) protocols in HIE models is proposed as the next possible move for information-sharing initiatives in health care. In addition to some potential opportunities and distinguishing features (eg, ownability, verifiability, and incentivization), we identify the uncertainty and risks associated with the application of NFTs, such as the lack of a dedicated regulatory framework for legal ownership of digital patient data. This paper is among the first to discuss the potential of NFTs in health care. The use of NFTs in HIE networks could generate a new stream of research for future studies. This study provides practical insights into how the technological foundations of information-sharing efforts in health care have developed and diversified from earlier forms.

(Interact J Med Res 2023;12:e42685) doi:10.2196/42685

KEYWORDS

health information exchange; HIE; personal health information; PHI; blockchain; nonfungible token; NFT; evolution of technology

Introduction

People may need to visit different health care providers (such as specialists) in their lives because they may encounter various health issues. Providers need to access accurate and complete patients' past medical records to make informed treatment decisions and increase the effectiveness and efficiency of care delivery. Accessing limited or incomplete information can cause duplication of health care services, such as laboratory tests and repetition of therapy. In addition, as physicians need to search for missing information, administrative costs increase, which could cause delays in providing care and slow down the providers' workflow. Thus, it is essential for treating physicians to access, integrate, and share patients' test results and medical procedure records conducted by various providers. However, health care organizations are not necessarily affiliated and may use different systems and standards for storing patient information (such as diverse electronic health records [EHRs]). Seamless sharing of personal health information (PHI) is a demanding project in a highly fragmented US health care system [1]. Fragmented health care services may also challenge how health care providers exchange health-related data as they may use various exchange mechanisms [2]. Thus, different standards for the data storage model, data documentation process, and data transfer mechanism may be used by health care providers in the United States [3]. Health data sharing is an endeavor involving stakeholders such as data owners, data users, and regulators. Sharing health data among health care entities can yield several benefits that include improving care coordination, care quality, and patient safety while reducing mortality rates, medical errors, and health care costs [4]. The ultimate goal is to enable all health care providers to share accurate, timely, and complete medical data nationwide with other entities regardless of where the patient has been treated previously. To achieve this objective, various exchange mechanisms and sharing systems have been used in health care. These systems have some advantages and drawbacks because of the supporting technology and implementation issues. Technological advancements have been used to address the challenges associated with the preceding ones and create new generations of sharing mechanisms. The following sections explain the 3 existing information-sharing mechanisms in health care and propose a new approach to fundamentally address critical issues in mainstream sharing efforts.

Ethical Considerations

As no human participants were used, this study was exempted from obtaining approval from an institutional review board.

Information-Sharing Mechanisms in Health Care

Traditional Information-Sharing Models

The first sharing method in health care was a paper-based exchange or mail transmission. Gradually, health care providers used other conventional methods (such as phone, fax, or information carried on CDs) to exchange patients' records. Information flows among disparate health care institutions can still be managed through traditional methods such as fax, paper mailing, and phone calls. However, previous studies have reported serious issues associated with nonelectronic data exchange among providers, such as the inability to provide timely access to patients' medical records and unnecessary tests [5]. As traditional systems cannot integrate patient data into a central hub, paper-based records such as fax or mail could be lost during the treatment period. Because medical data could be sensitive, losing them can increase privacy risks. It is also inconvenient for patients to carry paper-based records or CDs from one hospital or physician's office to another. The storage of patient records in paper-based folders or CDs also leads to huge maintenance costs for large hospitals. Moreover, keeping folders may cause numerous security risks, such as the threat of natural events and disasters. Offline-based exchange requires additional work, such as copying paper-based files or storing diagnostic images on a CD or a memory stick. Thus, traditional methods cannot be expected to reduce health care costs. Finally, it is unclear to patients who can access, view, and share their paper-based medical records because no alert, alarm, or security safeguards are available to protect offline patient data from unauthorized access.

Health Information Exchange

The health care industry is currently transitioning from the offline sharing of patient health information to web-based sharing through electronic health information exchange (HIE). HIE allows for web-based transfer of medical data and patient records among health care providers and institutions, providing access to accurate and up-to-date health information across different health care settings. This enables clinicians to make more informed and effective health care decisions, ultimately improving patient outcomes [6]. The primary objective of creating ecosystems for HIE systems is to promote the secure and efficient sharing of patient data on a national scale in the United States. HIE networks facilitate interoperability between various health care entities, intending to improve the quality of care, optimize clinical workflow, provide timely access to patient records, enhance connections between different organizations, and improve overall health care efficiency [7]. There are 3 main mechanisms of HIE: direct exchange, look-up systems, and patient-centered exchanges. Direct exchange involves authorized and trusted health care providers sharing patient data directly with one another. Look-up systems use a centralized database that enables providers to send query messages and request patient records. Patient-centered HIE gives patients greater control over their health information by allowing them to collect and aggregate data from various providers and share it with other health care entities, as needed. This approach enables patients to be more active in managing their health and ensuring that their health care providers access comprehensive and accurate health information [8].

In many developed countries, HIE programs are key policy areas aimed at improving care coordination by facilitating the sharing of accurate and comprehensive health information across health care providers and organizations. HIE databases can be used for various purposes, such as health care decision-making and clinical research. However, despite the potential benefits of HIE, such as improved coordination, reduced costs, and enhanced patient safety, insufficient participation of clinicians in data exchange networks can lead to incomplete HIE databases and reduce the overall value of HIE. In addition, the use of HIE mechanisms presents several challenges. Privacy concerns and the risk of data breaches are 2 important barriers to electronic data sharing in the United States. These factors must be addressed to ensure that HIE programs can achieve their objectives and realize their full potential for improving health care outcomes [9]. Patients' concerns about the privacy and security of their information can lead to incomplete medical records in HIE databases because of information-blocking behaviors. Owing to concerns about information integrity and confidentiality, patients may be less likely to participate in data-sharing efforts. This can create a challenge for health care providers who rely on accurate and complete medical records to deliver high-quality care. To address these concerns, it is essential to implement robust privacy and security measures that can help reassure patients and build trust in HIE systems.

In addition, efforts should be made to increase patient awareness and education regarding the benefits of participating in data-sharing programs while ensuring that their privacy and security concerns are adequately addressed [10]. Incomplete information in HIE systems suggests that not all essential data sources from patients are being integrated and accumulated, potentially owing to concerns over privacy and security risks. Even patients concerned about privacy and security may not fully appreciate the benefits of data sharing and its potential impact on public health. Consequently, they may be reluctant to consent to the disclosure of their data to different health care providers. To address this issue, it is crucial to increase patient awareness and education about the potential benefits of HIE systems while ensuring that their privacy and security concerns are adequately addressed. This can help build trust and confidence in the system, which can lead to more complete and accurate data being shared across health care entities [11]. Some health care providers may choose not to participate in HIE networks because of concerns over patient privacy and security and legal implications. These providers may hesitate to share patient health information, which can hinder the success and effectiveness of HIE systems. To address these concerns, it is important to establish clear guidelines and regulations regarding the collection, use, and sharing of patient health information within HIE networks. This can help mitigate legal risks and ensure that patient privacy and security are protected, while enabling the effective sharing of health information across health care entities. In addition, increasing education and awareness among health care providers about HIE systems' benefits can help build trust and encourage greater participation in these networks [12].

The primary challenges in implementing HIE systems are often attributed to organizational, governance, and technical barriers. These include limited interoperability between different health care information systems, a lack of standardized protocols and procedures for data sharing, and difficulties in coordinating and managing various health care entities and stakeholders involved in the HIE network. Addressing these barriers requires careful planning, collaboration, and investment to develop robust technical infrastructure, governance models, and organizational frameworks that support effective HIE implementation and operation. In addition, the ongoing evaluation and monitoring of HIE systems can help identify and address any ongoing barriers or challenges [13]. The implementation of HIE systems may face several obstacles, including the absence of proper governance structures, insufficient commitment from senior leadership, uncertain return on investment from HIE investment, inadequate technological infrastructure, absence of technical standards for promoting interoperability, challenges in integrating EHR data, insufficient adoption of certified EHRs, compliance with the Health Insurance Portability and Accountability Act (HIPAA) regulations, and insufficient technical training [14].

Furthermore, inadequate collaboration from EHR vendors, limited interorganizational partnerships with other health care entities, apprehensions about patient attrition due to HIE participation, and varying consent policies across different states are further challenges health care organizations may encounter when implementing HIE systems. The literature on HIE highlights that health care providers are worried about losing patients and their associated revenue when sharing data with competing organizations [15]. Competing health care organizations may view information blocking as a way to gain a competitive advantage by controlling patient flow. To maintain their competitive position in the market, these organizations may choose to only partially exchange health information or engage in information blocking altogether. However, framing HIE agreements with partners and reaching data use agreements about HIE can be complex and difficult, which presents a considerable barrier for HIE partners [16].

In the United States, financial incentives and mandates have been provided to encourage the participation of providers and clinicians in HIE projects [17]. An example of this is the federal Meaningful Use program, which outlines the standards for implementing certified EHRs [18], promoting the smooth flow of health information to enhance collaboration and care coordination, and reducing redundant tests and diagnostics, ultimately leading to cost savings in health care [19]. In addition, the Fast Healthcare Interoperability Resources standard is a recent effort to offer adequate beneficial interoperability without the complexity associated with a comprehensive interoperability solution [20]. The advancement of interoperability and transparency in providers' efforts to achieve this is largely supported by various organizations, such as the Office of the National Coordinator for Health Information Technology within the Department of Health and Human Services, as well as the Centers for Medicare & Medicaid Services and state governments [8]. Despite the policies and incentives provided, widespread participation of providers and clinicians in HIEs cannot be guaranteed. Previous studies have shown that although some hospitals have implemented HIEs, their clinicians may not fully use them to share all types of clinical information with all health care entities, including unaffiliated ones [5].

Challenges Related to Existing Sharing Models in Health Care

In light of the literature review, the main issues with current exchange mechanisms can be categorized into 4 groups. The first challenge is that mainstream sharing models are mainly centralized and controlled by a health care organization, and they define a minor role for patients in the sharing process [21]. HIE networks mainly focus on EHRs and centralized mechanisms managed by middlemen. The second concern is the low visibility in the security and transparency of sharing mechanisms and a lack of trust in sharing procedures and technologies [22]. The threat of a data breach, a single point of failure, and unclear permission processes are some dimensions of privacy and security issues. The third reason is data quality issues such as outdated data in HIE databases, incomplete or inaccurate patient information stored in HIE databases, and mutability, as any entity participating in HIE initiatives can remove patients' medical records. The final issue is that the current models do not openly delineate data ownership, and it is not clear who is the owner of the clinical data.

Blockchain-Based *HIE*

Previous studies suggested blockchain as an alternative to mainstream HIE systems [23-25]. Blockchain is an emerging technology that enables secure and decentralized approaches to reduce the technical risks and governance challenges associated with sharing sensitive health data. Blockchain technology could be a promising platform for sharing information among stakeholders with different interests [26]. Blockchain-based platforms can be a technological solution to foster trustworthy relationships between different business entities (such as affiliated and unaffiliated health care organizations) [27]. Blockchain solutions can improve information management's authenticity, security, and confidentiality using ledgers, encryption, and distributed networks [28]. By removing intermediaries, blockchain enables data ownership and gives users more control over their data [29].

One practical use of blockchain is to share health information [30]. Previous studies have presented potential advantages of blockchain technology in response to the traditional risks associated with conventional information exchange models [31]. For instance, in the context of HIE, a permissioned blockchain-based network has been suggested as a more secure option to enable the electronic exchange of clinical data with providers [32]. Because medical records are considered sensitive data, legal, consent, and privacy concerns are the top challenges individuals may encounter in sharing their health information [33]. Existing data-sharing models mainly use central data-management mechanisms, making data ownership and controlled access more complicated for individuals. Centralized applications cannot allow multiple stakeholders to actively participate in data-sharing governance. Furthermore, because of nonautomated consent mechanisms and data access management, the custody and administration of data sharing using traditional HIE are complicated [34].

Therefore, decentralized platforms that use encrypted databases are an effective alternative that enables independent stakeholders to supervise data contributions and access [35]. As there is neither a central administration nor a third party, trust will be placed in the network and distributed ledger to collect, store, and validate data sharing among data contributors [36]. Thus, previous studies have offered blockchain-based platforms for health data transmission between patients, providers, hospitals, and research organizations [37]. Decentralized networks of distributed nodes are deemed useful for reducing the inefficiency, costs, trust, and security risks of using central data sets across different boundaries. Data transmission through blockchain platforms can enable data contributors to maintain autonomous and ongoing control of their own data [38]. In these peer-to-peer platforms, each node consists of network participants (such as patients) that collectively contribute to the process of transaction validation and store the same copies of all data-sharing records.

Types of Blockchain *HIE* Systems

In the context of health care, there are 2 main types of blockchain networks: permissioned and federated [39].

- Permissioned blockchain: a permissioned blockchain is a closed network in which access is restricted to a defined group of participants [40]. Only authorized users can participate in the network and are typically required to pass identity verification checks before they can access the blockchain. Permissioned blockchains are often used in health care to ensure data privacy and security, as they provide a higher level of control over who can access and participate in the network.
- Federated blockchain: a federated blockchain is a network where multiple independent organizations come together to participate in a shared blockchain [41]. Each organization operates its own node on the blockchain, and the nodes work together to validate transactions and maintain the integrity of the network. Federated blockchains are often used in health care to enable information sharing between different organizations, such as hospitals, clinics, and insurance companies, while maintaining some control over who can participate in the network.

The key difference between permissioned and federated blockchains is the level of control over those who can participate in the network. In a permissioned blockchain, access is tightly controlled and only authorized users can participate. In a federated blockchain, there is more flexibility in terms of who can participate; however, the network is still designed to maintain some level of control over the participants to ensure security and data privacy. Both types of blockchains have their own advantages and disadvantages, and the choice of which one to use depends on the specific needs of the health care organization and the use case at hand. For example, a health care organization that is primarily concerned with data privacy and security may choose a permissioned blockchain, whereas an organization that wants to enable information sharing between multiple entities may opt for a federated blockchain.

Several private companies have already offered blockchain-based data-sharing platforms [42]. For example, health care organizations can run their health network on the Ethereum platform to provide different providers with access to treatment information. Caregivers can review the historical interactions between medical experts and patients, which enhances the transparency of the entire medical environment.

Smart Contracts

Blockchain-HIEs can use smart contracts, programmable computer protocols that verify and execute terms based on predetermined factors. A smart contract is a self-executing contract, with the terms of agreement between the buyer and seller being directly written into lines of code. The code and agreements contained therein exist on a blockchain network, and the contract is automatically executed when certain conditions are met [43]. Smart contracts are typically written in a high-level programming language, such as Solidity for the Ethereum blockchain, and are compiled into bytecodes that can be executed on the blockchain [44]. The code is stored on the blockchain, making it tamper-proof and transparent, and it can be accessed and executed by anyone on the network. One of the key benefits of smart contracts is that they enable trustless transactions, meaning that parties can exchange value without the need for a trusted intermediary. This can reduce transaction costs, increase efficiency, and improve security and transparency. Smart contracts can also be used to automate complex business processes, reduce fraud and errors, and increase accountability [45].

In an HIE setting, smart contracts can be used to automate the sharing and exchange of health data between different entities in the health care ecosystem, such as hospitals, clinics, insurers, and patients. Some examples of how smart contracts can be applied in HIE settings are as follows:

1. Access control: smart contracts can be used to control those who have access to patient health data and under what conditions. For example, a smart contract could be programmed to only allow a patient's primary care physician to access their medical records or only a researcher to access anonymized data for a specific research study.
2. Consent management: smart contracts can be used to manage patient consent for sharing and using their health data. For example, a smart contract could be programmed to automatically grant or revoke consent based on certain

conditions, such as the completion of a clinical trial or expiration of a consent period.

3. Payment management: smart contracts can be used to automate the payment and reimbursement processes for health care services. For example, a smart contract can be programmed to automatically process insurance claims and reimburse health care providers, based on predefined rules and conditions.
4. Compliance monitoring: smart contracts can be used to monitor and enforce compliance with health care regulations and standards. For example, a smart contract could be programmed to automatically verify that a health care provider has met certain quality standards or that a patient's health data has been handled in compliance with HIPAA regulations [46].

By using smart contracts in an HIE setting, it is possible to streamline and automate many of the processes involved in exchanging and using health data, while also improving data privacy, security, and transparency. Smart contracts can also reduce the administrative burden on health care providers and increase trust among patients and other stakeholders in the health care ecosystem.

A Proposed Approach: Nonfungible Token Protocols in HIE

Nonfungible Tokens: General Definitions and Examples

In addition to mainstream information-sharing mechanisms, this study also suggests a new approach to HIE efforts. We believe that this new system can leverage the application of nonfungible tokens (NFTs) in HIE networks. Because the concept of NFT is still novel, some basic information is required before NFT-enabled HIE is explained. NFT is generally a new method of digital authentication, as this protocol can be the process or action of proving or showing something genuine or valid. So far, the primary use cases of NFT are in sports moments, collectibles, video games, digital art, music, virtual worlds, fashion, trading cards, and domain names [47]. The NFT protocol is an alternative to the US copyright system, a government body that grants producers (artists) a certificate that can prove the work (artwork) is theirs. Using NFTs, individuals do not need a third party to manage their approval process. Instead, the authentication process can be performed through the Ethereum blockchain as the work (artwork) becomes digital with a certificate (token) [48]. If people download a file (art), this does not mean that they own it. This means that they have a copy of the art that is not original.

When an artwork becomes an NFT, individuals are likely to acquire it because they want to claim ownership of a rare and unique piece of the original art. NFT protocols can also protect artists by enabling one-on-one relationships between them and fans. NFT can help artists sell their products (eg, music and painting) directly to buyers without the involvement of a middleman such as a record label company. There are several reasons why people are eager to accept NFTs instead of copying and pasting artwork. Previous studies have highlighted several reasons why people enjoy purchasing and collecting NFTs [49]. The main motives are uniqueness, greater security than physical collectibles, potential to make money, competitive aspects, entertainment, and connection to an innovative community. Thus, we can define the NFT value based on the following formula: reputation of the creator (eg, artist) + utilities offered (for instance, sending the original tangible artwork) + ownership history (who owned the NFT before and how many times it has been sold) + future value (as a rare digital product).

The Role of Speculation in the Finances of *NFTs*

Speculation plays a major role in the financial aspect of NFTs. NFTs are unique digital assets that can represent the ownership of a particular item or piece of information, and the perceived rarity and demand of an asset often determine its value. As a result, NFTs have become popular assets for investors and collectors, leading to a surge in speculative buying and selling [50]. One factor driving the speculation in NFTs is the limited supply of certain assets. For example, a rare piece of artwork or a memorable moment in sports can be converted into an NFT, and the scarcity of such assets can increase their value in the market. In addition, the hype around certain NFTs can contribute to speculative buying, as investors seek to capitalize on the perceived value of a particular asset. The speculative nature of NFTs has led to considerable price volatility, with some NFTs selling for millions of dollars, whereas others fail to attract any buyers. This unpredictability can make NFT investment risky because the market can be influenced by various factors, including changing consumer tastes and technological advancements. Despite these risks, many investors and collectors continue to view NFTs as a valuable addition to their portfolios, and the popularity of NFTs is likely to continue to grow as technology and use cases evolve.

Technical Foundations of *NFT*

Metadata

NFTs are a type of digital asset stored on a blockchain, such as Ethereum. NFTs are unique, meaning that each NFT has a distinct value and cannot be replicated or duplicated. However, it is important to note that NFTs themselves do not contain the data in question but rather a very small collection of metadata that provide information about the asset [51]. For example, an NFT representing a digital artwork might include metadata such as the artist's name, the title of the artwork, and the date of creation. The actual artwork itself would be stored elsewhere, such as on a centralized server or decentralized storage platform such as the InterPlanetary File System [52]. When an NFT is purchased, ownership rights are recorded on the blockchain, making it a transparent and immutable record of ownership. NFT can be transferred to another owner by sending it to a digital wallet address.

Copyright

Copyright issues can arise with NFTs because they provide a way to monetize digital assets that may not have been possible previously. This has led to some controversy regarding NFTs and their impact on the art world and other creative industries. One issue is that NFTs do not necessarily confer ownership of the underlying asset but rather a unique identifier that is linked to the asset [53]. This means that someone who purchases an

NFT representing a digital artwork may not actually own the copyright to that artwork and may not have the right to reproduce or distribute it without the artist's permission. Another issue is that NFTs can be used to monetize assets that were previously freely available on the Internet, such as memes or other forms of user-generated content. This has led to concerns that NFTs could be used to profit from the work of others without their consent. Overall, although NFTs offer a new way to monetize digital assets and provide a mechanism for creators to protect their work, they also raise important questions regarding ownership, copyright, and the value of digital art and other assets.

Potential Application of *NFTs* in *HIE*

This section describes the potential application of NFTs to create digital proof of ownership in HIE. NFTs are recognized as a new way of creating value in various industries; however, they are still in their infancy and are challenged by speculation and inadequate regulations [54]. NFTs, as blockchain-based cryptographic assets that denote proof of ownership for digital objects, can be used in health care to authenticate digital PHI. All test results, treatments, medications, prescriptions, and care plans were considered PHI. NFTs can be produced on permissioned or federated blockchains, which provide a digital token of ownership for PHI. NFTs assigned to PHI can reduce health care organizations' time and effort to verify critical documentation, thus refining administrative operations of information sharing. NFT-based HIE issuing certificates can eliminate the workload of record keeping, with each medical record having a unique NFT that can be checked for authenticity. Moreover, issuing certificates on the blockchain-enabled HIE makes digital records resistant to tampering, which decreases the chance of encountering fraudulent PHI.

Network Topology

This section explains the type of blockchain that would be the best network for the proposed NFT-based HIE. Permission-less blockchains are open and decentralized. As no central entity can manage membership or ban illegitimate readers or writers, any individual can join and leave the network as a reader and writer at any time [40]. Thus, the stored on-chain content is readable by all members. However, permissioned blockchains authorize a limited set of readers and writers. Thus, a central entity decides and grants members the right to participate in the write or read operations of the blockchain [55]. Readers and writers can operate in separate parallel interconnected blockchains to promote privacy. To justify the best choice between permission-less and permissioned networks, the properties of these networks suggested by previous studies [56] can be evaluated as follows: (1) public verifiability enables anyone to verify the correctness of the system's state. For example, each state transition is confirmed by miners in the Bitcoin blockchain; (2) transparency explains the amount of information that should be transparent to an observer and the extent to which every participant can access every piece of information; integrity describes the extent to which health information is protected from unauthorized modifications; (3) redundancy in a blockchain-based HIE is mainly provided through replication across writers; and (4) trust anchor has the highest authority of a blockchain-enabled HIE system to grant and revoke read and write access to the system.

Moreover, to evaluate the best blockchain option for NFT-based HIE, we can use the following evaluation framework:

- Security: the blockchain option should be secure, ensuring the privacy and confidentiality of health information.
- Scalability: the blockchain option should be able to handle a large number of transactions without compromising performance.
- Governance: the blockchain option should have a transparent and robust governance mechanism to ensure the integrity of the data stored on the blockchain.
- Accessibility: the blockchain option should be accessible to all participants in the HIE network.
- Interoperability: the blockchain option should be able to work seamlessly with other existing systems and technologies.

Permission-less blockchains (such as Bitcoin and Ethereum) have a high level of security because they use a distributed ledger system that is difficult to hack. However, they are unsuitable for HIE owing to their limited scalability and governance issues. Permission-less blockchains can handle only a limited number of transactions per second, which is insufficient for large-scale HIE networks. Permission-less blockchains are also unsuitable for handling sensitive health information because of their lack of privacy and confidentiality. In contrast, permissioned blockchains (such as Quorum and Ripple) offer better security and privacy than public blockchains and also provide a good balance between security and scalability. They are scalable and can handle a large number of transactions per second, making them suitable for HIE networks. Permissioned blockchains can provide the required level of governance for HIE networks, as they allow only authorized parties to participate in the network, maintaining the transparency and accessibility of the network. However, permissioned blockchains can be more expensive than public blockchains and may require more resources for maintenance.

On the basis of the evaluation framework, the best blockchain option for an NFT-based HIE is a permissioned blockchain owing to several factors. First, permissioned blockchains offer higher security than public blockchains, because they allow only authorized participants to join the network. This ensures that sensitive health information is protected from unauthorized access or tampering.

Second, permissioned blockchains are scalable and can handle many transactions per second, making them useful in HIE networks. This is particularly important for HIE networks because they require the ability to handle a large volume of transactions while maintaining the integrity of the data. Third, permissioned blockchains provide a transparent and robust governance mechanism that is essential for ensuring the integrity of the data stored in the blockchain. This allows for a higher level of accountability and trust among participants in the network. Fourth, permissioned blockchains offer accessibility to all participants in the HIE network, as they allow authorized users to join the network and access data. This ensures that all relevant stakeholders can access the information they need to make informed decisions. Finally, permissioned blockchains are interoperable, meaning they can work seamlessly with other existing systems and technologies. This is particularly important for HIE networks, as they must integrate various health care systems and technologies to ensure the smooth exchange of health information.

In summary, a permissioned blockchain is the best option for NFT-based HIE owing to its high level of security, scalability, governance, accessibility, and interoperability. By using a permissioned blockchain, stakeholders in the health care industry can ensure secure and efficient exchange of sensitive health information while maintaining transparency and accountability among all participants in the network.

Consensus Mechanism

The consensus mechanism is a critical aspect of blockchain technology because it enables all nodes in the network to agree on the state of the ledger and improve their fault tolerance [57]. The consensus mechanism determines how new transactions are verified and added to a blockchain. The 2 main types of consensus mechanisms used in blockchain technology are Proof of Work (PoW) and Proof of Stake (PoS). Permission-less blockchains, such as Bitcoin and Ethereum, use PoW as their consensus mechanism. In PoW, miners solve complex mathematical problems to verify transactions and add them to a blockchain. The first miner to solve this problem is rewarded with a newly minted cryptocurrency. PoW is a computationally intensive and energy-consuming process, which makes it less efficient and environmentally friendly than other consensus mechanisms.

In contrast, permissioned blockchains such as Hyperledger Fabric and Corda use PoS or other consensus mechanisms such as Practical Byzantine Fault Tolerance (PBFT) or Raft. In PoS, validators hold a stake in the network, and the probability of being chosen to verify transactions and add them to the blockchain is proportional to the size of their stake. PoS is more energy-efficient than PoW, making it a more suitable consensus mechanism for permissioned blockchains.

Furthermore, PoS consensus mechanisms are often faster and can handle more transactions per second than PoW, making them more suitable for permissioned blockchains that require a high transaction throughput. PBFT and Raft, by contrast, offer a faster consensus mechanism by allowing nodes to reach an agreement through direct communication rather than mining.

In summary, permission-less blockchains rely on PoW as their consensus mechanism, which is computationally intensive and energy-consuming. Permissioned blockchains, by contrast, use more efficient consensus mechanisms, such as PoS, PBFT, or Raft, which are faster, more energy-efficient, and more suitable for high transaction throughput. In a permissioned blockchain, the consensus mechanism is designed to be more efficient, scalable, and suitable for the specific use case of NFT-based HIE. One of the most commonly used consensus mechanisms in this permissioned blockchain could be PoS. In PoS, the validators are incentivized to behave honestly as they stand to lose their stake if they act maliciously. PoS is more energy-efficient than PoW, making it a more suitable consensus mechanism for NFT-based HIE. Because permissioned blockchains have a known set of validators, the consensus mechanism can be optimized for efficiency, throughput, and security. Another advantage of permissioned blockchains is the use of other consensus mechanisms such as PBFT or Raft. These consensus mechanisms use direct communication between nodes to reach a consensus, allowing for faster transaction times and higher transaction throughput.

In NFT-based HIE, permissioned blockchains can be designed to accommodate different types of participants, such as health care providers, insurance companies, and patients, each with their own set of permissions and access levels. This ensures that only authorized participants can access the sensitive health information stored on the blockchain. Thus, the consensus mechanism for permissioned blockchains, such as PoS, PBFT, or Raft, is designed to be more efficient, scalable, and suitable for NFT-based HIE. These consensus mechanisms provide a more energy-efficient and faster alternative to PoW and allow customized permission levels for participants in the network, ensuring that sensitive health information is accessible only to authorized parties.

Authentication Process via *NFTs*

NFTs enable patients to own their medical records. Thus, health care providers' new entries (eg, test results) can be first encoded as NFTs and then added to the blockchain. Next, the ownership certification of ownership can be sent to the patient node. This authentication protocol can increase the transparency of medical data ownership and offer new ways to claim or enact ownership. All entities in the blockchain (eg, physicians and insurers) are notified of new data entry, but they cannot access, view, and share records because they do not own them. Another characteristic of NFTs is their verifiability, which is their ability to validate asset ownership. Verifiability proposes the protection of digital assets (such as PHI) against security attacks such as tampering, denial of service, spoofing, and repudiation [58]. When a patient grants permission to a treating physician, a smart contract can share the NFT assigned to patient data with the physician to view records for consent. Sharing the NFT designated as a PHI implies that it confers some rights to the holder (for example, analyzing patient data for finding care planning), but legally, patients will remain the original owner. Therefore, a PHI can be considered a commodity that is useful information transferable between health care providers and patients.

In this system, patient data are represented by an NFT, which contains a small amount of metadata that describes the data and links them to the actual data stored in an external system. Thus, on-chain or off-chain modulation can be implemented. Some metadata on health data transfer (such as sender and recipient addresses and purpose of transfer) could be saved on-chain, and some sensitive data (such as medical records and care planning) could be stored in cloud servers, as cloud computing may play a role in the off-chain storage of health data. Off-chain blockchain systems imply computation or data structurally external to the blockchain network [59]. This explains the communication and interplay between on-chain and off-chain storage, computation, and efforts to evaluate their performance.

The main advantages of these blockchain systems are improved scalability, reduced data storage requirements, and enhanced data privacy. These features are well-suited to the needs of the health care industry because of the need to manage various types of medical records, patients, and other health-related data.

The NFT acts as a digital asset that the patient can own and control [60]. When a patient grants permission to a treating physician to access their data, this permission is recorded on the blockchain as a transaction that is validated by the network. A smart contract is then used to manage sharing of the NFT assigned to the patient's data with the treating physician. The smart contract contains a set of rules and conditions that specify the terms of the patient's consent and the conditions under which the physician is authorized to access the data. For example, the smart contract might specify that the physician is only authorized to view certain types of data for a specific period or that the physician is required to obtain further consent from the patient before sharing the data with other parties. Once the conditions of the smart contract are met, the NFT assigned to the patient's data is shared with the treating physician, who can access the actual data stored in the external system. The smart contract records the physician's access to and use of the data, providing an auditable trail of all data accesses and uses. By using smart contracts in this manner, blockchain-based HIE systems can provide patients with greater control over their health data and enable them to securely share it with authorized parties. Smart contracts also enable patients to set specific conditions and rules for using their data, ensuring that they are only accessed and used per their wishes.

One challenge is that when a patient grants permission to a treating physician to access their data, there may be a need to re-encrypt the data for the physician. This requires a considerable amount of computational effort, bandwidth, and storage, depending on the size of the data and level of encryption used. One approach to address this challenge is to use a hybrid encryption scheme that combines symmetric and asymmetric encryption [61]. In this approach, the patient encrypts their data using a symmetric encryption key, which is then encrypted by the physician's public key using an asymmetric encryption algorithm. The encrypted data and encrypted symmetric key are stored in an external system. When the physician requests access to the data, a smart contract is triggered, and the patient's private key is used to decrypt the symmetric key, which is then used to decrypt the data. This process ensures that the data remain encrypted at rest and in transit and can only be decrypted by authorized parties. To reduce the computational effort and bandwidth requirements, the data can be compressed before being encrypted and transmitted to the physician. In addition, advanced encryption algorithms such as homomorphic encryption can be used to perform computations on encrypted data, further reducing the need to decrypt the data and increasing privacy and security. It is worth noting that although re-encrypting data for physicians can be computationally intensive, it is a necessary step to ensure the privacy and security of the patient's data. Using advanced encryption techniques and optimizing the data transfer process can reduce the computational burden and make the exchange of encrypted health data more feasible in blockchain-based HIE systems.

Patient Nodes or Wallets

It should be noted that in a blockchain-based HIE system, the "patient node" refers to the part of the network that stores and manages the health data of individual patients. The assumption is not necessarily that patients themselves operate a blockchain node but rather that they have control over their own health data and can grant access to it to authorized parties. The patient node can be operated by various entities, such as health care providers, hospitals, or third-party vendors. In some cases, patients may also be able to operate their own nodes if they have the technical knowledge and resources to do so. However, even if patients do not directly operate a node, they can still benefit from the use of blockchain technology in HIE. For example, blockchain can provide patients with greater control over their health data and enable them to securely share it with health care providers and other stakeholders, as needed. Using a blockchain-based HIE system, patients can also have greater confidence that their data are being protected and used in accordance with their wishes.

In NFT-based HIE, patients can have their own nodes or wallets depending on the design of the blockchain network. However, it is important to note that the level of participation and access to the blockchain network for patients may be limited compared with other participants, such as health care providers or insurance companies. Patients can have their own nodes, which are essentially software clients that allow them to interact with a blockchain network. These nodes can be used to access their health information, verify transactions related to their health records, and grant permission to use their data in research or other applications. However, running a node requires technical expertise and resources, which may not be accessible to all patients. If patients have technical expertise and resources, running their own nodes can give them greater control over their health information and ability to participate more actively in the network. However, this option requires more technical knowledge and resources and may not be accessible to all patients.

An alternative option for patients is to use a wallet, which is a digital tool that allows them to store and manage their NFTs representing their health records. The wallet can be used to authorize access to health records and grant permission for their use in different applications. The use of a wallet is generally easier and more accessible to patients than running a node. Patient wallets are generally more accessible and user-friendly, requiring minimal technical expertise. This option provides patients with a more streamlined and convenient way to manage their health information on the blockchain network. In general, patient wallets may be a more suitable option for most patients with NFT-based HIE as they offer a balance between accessibility and control. Patients can use wallets to manage their health information and authorize access to their data, while retaining some level of control and ownership over their data.

Thus, patients can have their own nodes or wallets in NFT-based HIE, depending on various factors, such as the technical expertise of the patient, desired level of control and access to the network, and design of the blockchain network. Although running a node provides more control and access to the network, using a wallet is a more accessible option for patients who may not have technical expertise or resources to run a node. A well-designed NFT-based HIE should provide patients with a range of options for managing their health information on the blockchain network, ensuring that their data are secure, accessible, and under their control.

Incentivization

In a blockchain-based HIE system, the main challenge is motivating patients to share their medical records with other nodes. Blockchain technology has been suggested to eliminate the inefficiencies, costs, and risks associated with traditional data sharing in health care. Blockchain can also be used to authenticate genuine content [62]. However, the issue with blockchain-based HIE is finding an appropriate and meaningful incentive mechanism to use the promise of data sharing by relying on a decentralized system for data storage and management. As NFTs are nonfungible, their perceived value depends on their content, characteristics, and purpose of use. We can expect that because PHI is unique and its units are noninterchangeable with one another, it is nonfungible. Thus, patients retain ownership of an NFT assigned to PHI and collect royalties (incentives) from sharing their content. Smart contracts can provide reasonable incentives for sharing NFT-based medical records for different purposes. Smart contract terms and conditions can be set based on 2 primary purposes of HIE: health care and medical research.

The NFT assigned to a patient's PHI is often shared with other physicians for health care reasons such as receiving professional advice, diagnosis, prescription, treatment options, and care planning. In this case, blockchain-based HIE can reward data owners (patients) using recognition points. Thus, blockchain technology can support building incentives for data owners to share their data in exchange for credits encoded in smart contracts [63]. Credits are integrated into blockchain-based platforms and shared with others in the HIE network. For example, the holders of credits will receive recognition for sharing their health data that could be used to improve health quality, help physicians find customized care, reduce health risk factors, and discover the best health care practices. Receiving more points implies that the patient has been actively engaged in their health care procedures. Even gamification concepts, such as points and leveling systems, can be used to calculate engagement scores and rank patients accordingly compared with their peers in HIE networks.

In the second case, disease foundations and academic institutions may ask data owners (patients) to share the NFT assigned to their PHI for clinical research purposes. Blockchain-based HIE can incentivize patients with digital tokens to encourage them to assist in health discoveries and help drive medical innovation for the greater good of humanity. NFTs enable patients to receive royalties each time their PHI is transferred to a new research project. Thus, terms and conditions defined in smart contacts can calculate incentives and electronically reward data owners with cryptocurrencies to share the NFT of medical data for medical searches. For example, owners of NFTs who share their medical records, lifestyle data, and other health information with scientists through a secure platform are not the subjects of research, but are partners in discovering new treatments. In return, patients who share NFTs assigned to health data will receive coins, which can be exchanged with other cryptocurrencies (such as Bitcoin and Ethereum). As patients share NFTs in the network and the value of NFTs varies, incentives can be calculated based on a mix of recency, variety, and volume of medical data, as well as the frequency of sharing. One copy of NFTs exists in this decentralized platform, and patients can control their inclusion in the network and release their consent to how it is used in research. All health data are deidentified, accumulated, encrypted, and stored in the permissioned blockchain. If patients no longer want to contribute to health research, they can revoke permission and remove their NFT assigned to health data from the platform.

Challenges to Incentivization

Incentivizing users to share data for financial gain in a decentralized and anonymous environment can create challenges related to data quality. When users are incentivized to share data for financial gain, there is a risk that malicious actors will fabricate data sets to take advantage of the incentives [64]. This can result in the creation of large volumes of low-quality or fraudulent data that can be detrimental to commercial users and scientific research. One way to mitigate these risks is to design incentives to reward users for sharing high-quality data validated through independent sources. For example, rewards could be tied to data that a trusted third party, such as a research institution or a regulatory agency can verify independently.

In addition, incentives could be designed to encourage users to share data relevant to specific research or commercial applications and discourage the sharing of data that are not relevant or of poor quality. Another approach for mitigating data quality issues in a decentralized and anonymous environment is to use data validation algorithms to detect and filter out fraudulent or low-quality data. These algorithms can be designed to analyze patterns and anomalies in data to identify potential sources of fraud or errors. Using these algorithms can reduce the risk of fraudulent data and maintain the overall quality of data sets. Overall, it is important to carefully consider the design of incentives and validation mechanisms when incentivizing users to share data in a decentralized and anonymous environment. Using a combination of trusted third-party validation and sophisticated data analysis techniques can incentivize users to share high-quality data while reducing the risk of fraudulent or low-quality data.

Key Management System

Ensuring safe custody of patient keys is a critical component of any blockchain-based HIE system. One approach to address this challenge is to use a key management system (KMS) designed to securely store and manage cryptographic keys, including private keys [65]. A KMS can offer a range of features and safeguards to protect private keys, such as encryption, access control, and backup and recovery capabilities. For example, a KMS can encrypt private keys using strong cryptographic algorithms and protect them by restricting access controls that can view or modify them. In addition, a KMS can store backup copies of private keys in secure, off-site locations, which can help prevent key losses owing to hardware failures, natural disasters, or other unforeseen events. In addition to using a KMS, several other measures can be taken to ensure the safe custody of patient keys in a health care environment. For example, patients can be educated about the importance of safeguarding their private keys and providing instructions on how to do so. Health care providers can also implement policies and procedures to help patients manage their keys, such as offering secure storage options or periodically reminding patients to check the status of their keys. Ultimately, the key to ensuring the safe custody of patient keys in a blockchain-based HIE system is to balance security with usability. Although it is important to use strong security measures to protect private keys, it is also important to ensure that patients can easily access and manage their keys without undue burden or complexity.

Data Ownership and Access Control

It is worth mentioning that there is still debate about patients always being the owners of their health data [66]. In most cases, patients are considered the owners of their health data. However, ownership of health data can be a complex issue and may vary depending on the specific situation and jurisdiction. For example, in some cases, health care providers may own certain portions of a patient's health data, such as test results or clinical notes. In addition, if a patient has agreed to participate in a research study or clinical trial, ownership of their health data may be transferred to researchers conducting the study. The ownership of health data may change for a variety of reasons. One possible reason is when a patient decides to share their health data with a health care provider or another third party for a specific purpose, such as obtaining a second opinion or participating in a clinical trial. In this case, the patient may transfer ownership of their health data to the health care provider or third party for the duration of the specific purpose.

Another possible reason for a change in health data ownership is when a patient agrees to sell their health data to a third party, such as a pharmaceutical company or research organization. In this case, the patient transfers ownership of their health data to a third party in exchange for compensation. Any transfer of ownership of health data should be performed with informed consent from the patient and in compliance with applicable privacy laws and regulations. In addition, patients should be able to revoke their consent and regain ownership of their health data at any time. Even if we assume that patients own their health data, they can remain the owner but share more than one copy of a given health data set with health care researchers (in exchange for incentives) using the design principles of NFT. Thus, data ownership changes in the context of HIE can facilitate data sharing.

NFTs can be useful in tracking the ownership and provenance of digital health information, but they do not inherently provide privacy or secure access. In addition to ownership, access control is an essential aspect of the exchange of health information. Access control determines who has the permission to view, modify, or share health information. Although an NFT can indicate ownership of a piece of digital health information, it does not automatically provide access control. Access control mechanisms must be in place to ensure that only authorized individuals or entities can access information. Several access forms are needed in NFT-based HIE to ensure proper privacy and security of PHI. These access forms are as follows:

- View access: this is the ability to view health information. View access is necessary for health care providers and patients to access their health records.
- Modify access: this is the ability to modify or update health information. Access modification is necessary for health care providers to update patient records with new information, such as diagnoses, treatments, and medications.
- Share access: the ability to share health information with other health care providers or entities. Share access is necessary for health care providers to share patient records with other providers involved in patient care, such as specialists or hospitals.
- Revoke access: this ability to revoke access to health information. Revoke access is necessary for patients to control access to their health records and to prevent unauthorized access.
- Audit access: this is the ability to audit access to health information. Audit access is necessary to track who has access to health records and monitor for unauthorized access.

These forms of access are crucial in ensuring that PHI is properly secured and only accessed by authorized individuals or entities. Although NFTs can be used to track the ownership of digital health information, access control mechanisms must be implemented to ensure the privacy and security of PHI so that only authorized individuals or entities can access it. Thus, NFT-based HIE with access control mechanisms can potentially help solve ownership issues related to health data. In traditional HIE, ownership of health data can be unclear, with different parties (such as health care providers, patients, and health systems) claiming ownership of different aspects of the data. Using NFTs to track ownership of health data can clarify who owns which pieces of data. NFTs with robust access control can be used to create a clear and transparent record of ownership, which can help prevent disputes over ownership of health data. This can potentially streamline the sharing of health information and make it easier for patients to access their own health records by ensuring the proper use and protection of PHI.

Data Storage, Security, and Cost

On the basis of on-chain or off-chain modulation, the data can be stored in or off the network. In the on-chain model, an NFT will only hold metadata for the health data, not the health data itself, because health information may be too big to be efficiently saved on chain or they could be very sensitive, which could raise privacy concerns. Blockchain technology, which underpins NFTs, has limitations in terms of scalability, and storing large amounts of data on a blockchain can be expensive and slow down the network. However, NFTs can still be useful for securely tracking and managing health data (such as data related to health data transfer between 2 health care organizations). In the off-chain model, an NFT can hold more sensitive data, such as the patient's name, medical record number, date of birth, and other relevant health-related information. Thus, patient names and other identifiers are not included in the NFT core data because of privacy concerns. On the basis of this modularity, metadata can be used to link the NFT to the actual health data stored in an external system, such as a centralized database or decentralized storage network.

Therefore, the health data must be stored in an external system. For example, health data could be a centralized system, where a single entity or organization is responsible for operating the data storage, maintaining encryption, and standard techniques for securing sensitive data and bearing costs. The choice of encryption scheme would depend on the system's specific requirements, such as the level of security required, size of the data, and system performance requirements. Some examples of encryption algorithms to secure health data stored in an external system can be advanced encryption standards, RSA encryption, elliptical curve cryptography, and blowfish [67]. Using an NFT to represent a patient's health data makes it possible to maintain a secure and tamper-proof record of the data ownership, access, and use. This can improve data privacy and security, reduce the risk of data breaches, and increase trust in the health care system.

The entity that operates the binary data storage depends on the specific implementation of the system. In a centralized system, a single entity or organization may be responsible for operating storage. By contrast, in a decentralized system, storage and cost may be distributed among multiple nodes in a blockchain network. In either case, it is essential to ensure that the entity operating the storage has proper security measures in place to prevent unauthorized access and protect data from cyber threats. Regarding the cost related to storing large binary data, the responsible party depends on the specific implementation of the system. In a centralized system, the entity operating the storage unit is typically responsible for bearing costs. In a decentralized system, the cost may be distributed among multiple nodes in the blockchain network. The nodes that store the data may be incentivized by rewards or other compensations. Ultimately, the responsible parties and the cost structure must be determined based on the specific use case and implementation.

Will *NFTs* Act as New Standards?

Despite the mentioned flaws of traditional HIE systems, they have been tested and tried, and many adhere to strict regulatory requirements, which is not the case for the novel, blockchain-based HIE. Indeed, health care organizations often use different standards, making information sharing more complex [68]. However, it is also important to note that using NFTs in blockchain-based HIE systems does not necessarily imply the creation of a new standard. Rather, NFTs can be viewed as tools for facilitating information sharing across existing standards and systems. By creating a common mechanism for representing and accessing patient data, NFTs can enable health care organizations to exchange data more easily and efficiently, without necessarily requiring them to adopt a new standard. One of the main advantages of blockchain-based HIE systems is that they are designed to operate in a decentralized and interoperable manner, which means that they can work with various standards and systems. By leveraging the power of blockchain technology and NFTs, health care organizations can create a more unified and standardized approach to data sharing without necessarily forcing them to adopt a single, rigid standard. There may still be challenges associated with integrating different standards and systems, and there will likely be a need for ongoing collaboration and cooperation among health care organizations to ensure that data are exchanged accurately and securely. However, by using NFTs in blockchain-based HIE systems, health care organizations can take an important step toward creating a more efficient and effective health care ecosystem that can better meet the needs of patients and providers alike.

How Health Data Are Exchanged Using *NFTs:* Steps and Processes

NFTs are unique digital assets that represent ownership of a particular item or piece of information. In health care, NFTs can be used in HIE models to secure information exchange between different health care providers. The process of exchanging information using NFTs in HIE models typically involves the following steps:

1. Creation of NFTs: health care providers create NFTs that represent specific pieces of patient information, such as medical records, test results, or imaging studies.
2. Authentication of NFTs: before exchanging information, NFTs are authenticated to ensure that they represent valid and accurate information. This authentication process can include verifying the identity of the health care provider who created the NFT and checking the integrity of the data represented by the NFT.
3. Transfer of NFTs: once authenticated, NFTs are transferred securely between health care providers using blockchain technology. The blockchain ensures that the transfer of the NFT is immutable and tamper-proof, which helps maintain the privacy and security of patient information.
4. Verification of NFT ownership: when a health care provider receives an NFT, they verify the ownership of the NFT to ensure that they have the right to access the patient information represented by the NFT. This verification process involves checking the digital signature associated with the NFT or consulting a blockchain ledger to confirm ownership of the NFT.
5. Accessing patient information: once ownership of the NFT is verified, the health care provider can access the patient information represented by the NFT. This information can be used to inform patient care and treatment decisions.

Overall, using NFTs in HIE models can help ensure secure and efficient information exchange between health care providers, while protecting patient privacy and data security.

The Architecture of *NFT*-Based *HIE*

The NFT-based HIE mechanism consists of several key components:

1. Health care providers and patients: health care providers (such as doctors, hospitals, clinics, and pharmacies) create and access EHRs for their patients. Patients can also access their own EHRs and share them with health care providers.
2. EHRs: EHRs are electronic records that contain patient health information, including medical history, diagnoses, treatments, and medications. These records are stored in a

secure and decentralized manner using the blockchain technology.

3. NFTs: NFTs are unique digital tokens used to represent ownership of digital assets. In the context of NFT-based HIE, NFTs are used to represent ownership of patient EHRs.
4. Smart contracts: smart contracts are self-executing programs that run on a blockchain. In the context of NFT-based HIE, smart contracts are used to automate the process of sharing patient EHRs. Smart contracts define the rules and conditions for sharing EHRs, and ensure that these rules are followed.
5. Data sharing: when a health care provider requests access to a patient's EHR, the patient can grant permission by transferring ownership of the NFT representing their EHR to the health care provider's wallet. The smart contract is then executed and the health care provider can access the patient's EHR.
6. Audit trail: the blockchain maintains a transparent and immutable audit trail of all EHR transactions, providing a secure and reliable record of who accessed what information and when.

In the first schematic diagram (Figure 1), the NFT-based HIE network comprises various participants, including health care providers, insurance companies, patients, and the blockchain network. Each health care provider has its own node connected to the NFT-based HIE network, enabling it to interact with the blockchain and access patient health records. The insurance company also has its own node connected to the network, enabling it to verify insurance claims and payment transactions. Patients have their own digital wallets or nodes connected to the network, which they can use to manage their health records and grant access to health care providers or insurance companies. When a patient visits a health care provider, the provider creates a new NFT representing the patient's health record and adds it to the blockchain network. The NFT contains a unique identifier that links it to the patient's identity and other relevant information, such as the type of medical treatment, date, and health care provider.

The health care provider can then access the patient's health records through the blockchain network using their own nodes. The insurance company can also verify transactions related to the health care claim and process the payment through its own node that is connected to the network. Thus, this network topology and mechanism enable secure and efficient health information sharing between different NFT-based HIE network participants, while ensuring data privacy, security, and ownership.

The second diagram (Figure 2) schematically shows the operation of an NFT-based HIE network. The patient's medical data are stored in their wallet as an NFT, containing a unique ID and all their health information. When health care providers need access to this information, they request it from the patient's wallet through the HIE network. The HIE network uses a smart contract to manage NFTs and access control for health care providers. The provider wallet also contains an NFT that identifies them as health care providers and allows them to access the patient's health information. Once the provider has verified their identity and permissions, they can access the patient's health information from the patient's wallet. The provider can then update the patient's health information and send the updated data back to the patient's wallet via the HIE network. All transactions between the patient's wallet, provider's wallet, and HIE network are recorded on the blockchain as secure and private transactions. The blockchain also contains a medical data registry, which stores medical data and associated NFTs, and enables secure and private access to patient health information. The patient's HIE record is a permanent, tamper-proof record of all of their health information. The record is stored as an NFT on the blockchain and is accessible only to authorized health care providers with patient permission. Overall, an NFT-based HIE network provides a secure, private, and decentralized way for patients to control and share their health information, while also ensuring that health care providers have access to accurate and up-to-date medical data.

Figure 1. Schematic diagram of relationships between entities in a nonfungible token (NFT)–based network. HIE: health information exchange.

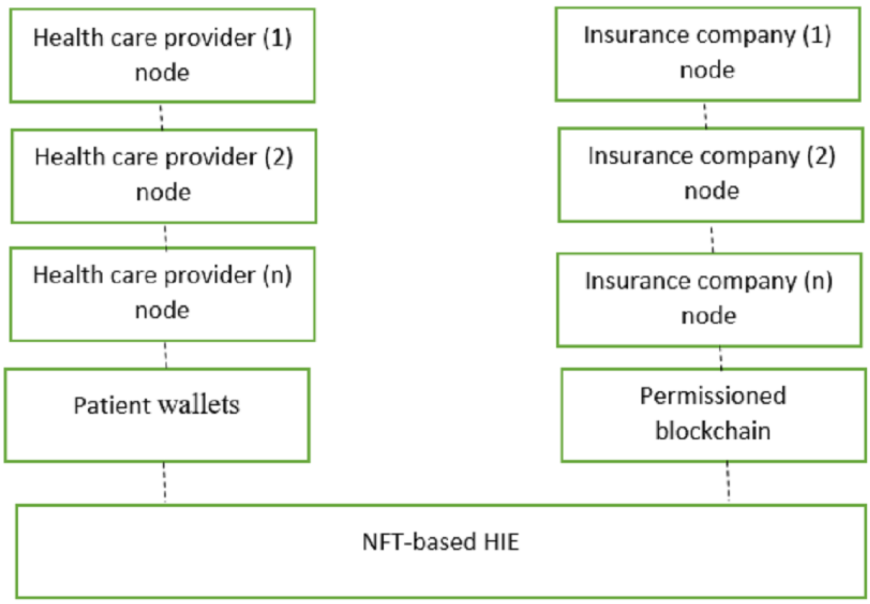
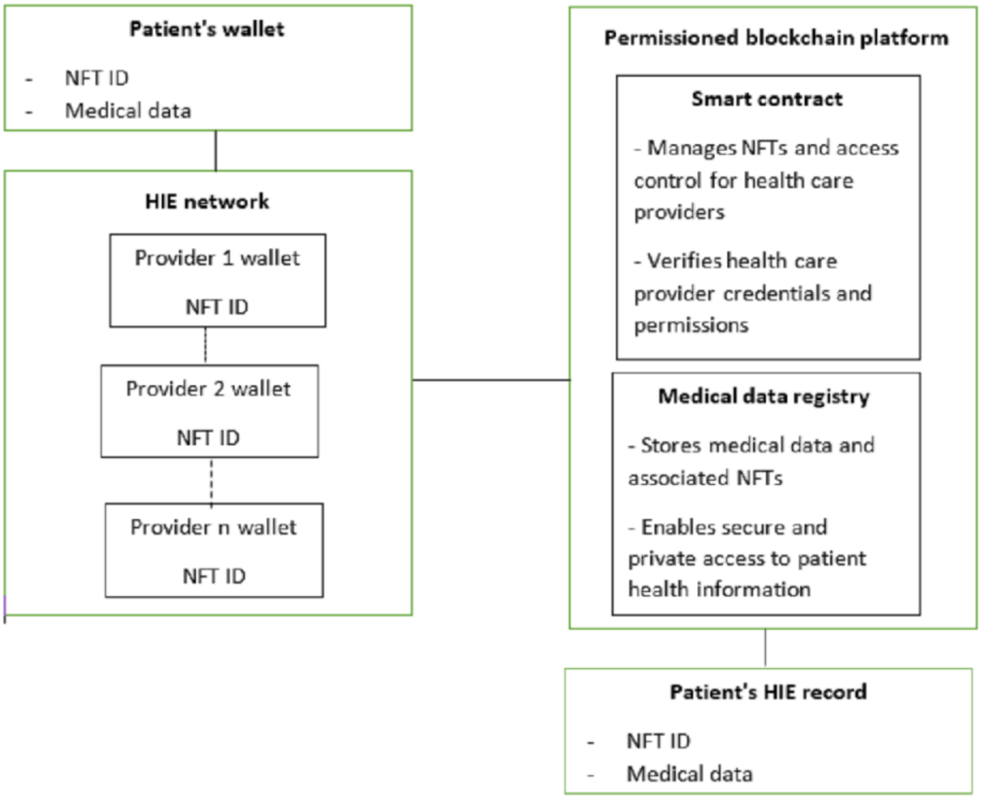


Figure 2. Schematic diagram of how a nonfungible token (NFT)–based health information exchange network works.



Costs and Data Volumes Affecting the Implementation Process

The costs and data volumes can significantly affect the design and implementation of NFTs in blockchain-based HIE systems:

- Costs: the cost of implementing NFTs in blockchain-based HIE systems can vary depending on the system's complexity, blockchain technology used, and number of participants involved. As NFTs are unique digital assets, the cost of creating and storing them can be high, particularly for large volumes of data. The cost of creating and managing NFTs can also increase as the number of parties involved in the HIE system increases.
- Data volumes: the amount of data being exchanged via NFTs in blockchain-based HIE systems can significantly affect system design and performance. As data volumes increase, the HIE system may need to be designed to handle increased traffic, potentially requiring additional computing power and storage capacity. In addition, as data volumes increase, the system's security mechanisms must be scalable to ensure that the data are not compromised.

Several strategies can be used to address the challenges posed by costs and data volumes:.

- Optimization of system design: system designers can optimize the design of blockchain-based HIE systems to reduce costs and improve performance. This can include designing the system to scale dynamically, using cost-efficient blockchain technologies, and minimizing the amount of data exchanged via NFTs.
- Data compression and aggregation: to reduce costs associated with NFT creation and storage, data can be compressed and aggregated to reduce the size of the NFT. This can be done by extracting only essential data from patient records, which can help reduce the cost and complexity of creating and managing NFTs.
- Collaborative models: by implementing a collaborative model for HIE, the cost and complexity of managing NFTs can be reduced. In a collaborative model, health care providers can share the costs associated with NFT creation and management, potentially leading to lower costs for all the parties involved.

Thus, cost and data volume considerations must be carefully considered in the design and implementation of NFT-based HIE systems to ensure that the system is efficient, secure, and scalable.

Concerns About Using *NFTs* in *HIE* Systems

As with any emerging technology, there are criticisms and concerns surrounding the use of NFTs in HIE [69]. Some criticisms include the following:

- Limited scope: although NFTs can potentially transform HIE by providing a secure, decentralized mechanism for exchanging health information, their scope is limited. NFTs can only be used to exchange specific pieces of information, such as medical records or test results. They cannot be used to exchange real-time data, such as patient vitals, which are critical for health care decision-making.
- Lack of interoperability: one of the key challenges in HIE is interoperability—the ability of different systems to exchange and use information. Although NFTs can provide a secure mechanism for exchanging information, they may

not be interoperable with existing HIE systems. This can limit its usefulness and adoption.

- Regulatory challenges: the use of NFTs in HIE raises regulatory challenges, including issues related to data privacy, security, and ownership [69]. The regulatory landscape for NFT-based HIE is still evolving, and it is unclear how regulators approach these challenges.
- Technical challenges: the technical challenges associated with designing and implementing NFT-based HIE systems can be noteworthy. These challenges include ensuring the scalability and performance of the system, managing the costs associated with NFT creation and management, and ensuring the security of the patient data.

Thus, although NFTs can potentially transform HIE, some challenges and limitations need to be addressed. To effectively integrate NFTs into HIE, careful consideration must be given to these challenges and potential solutions must be explored.

Challenges and Suggestions for Future Studies

The concept of NFTs is suggested to address a long-standing problem related to the proof of ownership for PHI by offering a mechanism to validate who could own the medical data in the HIE networks. However, our study is among the first attempts to highlight this opportunity, and it is far from achieving this goal, with several questions remaining regarding the legal, financial, and user aspects. The first challenge regarding the application of NFTs in HIE projects can be viewed from the perspective of regulatory considerations. Topics related to NFTs are still novel; thus, a lack of regulation may facilitate fraudulent activities and increase uncertainty regarding the use of NFTs in health care. As the NFT sector is currently prone to fraud, such as phishing activity in the digital asset domain, new dedicated regulations are required to distinguish the application of NFTs in health care. For example, a new amendment to HIPAA is required to articulate how a blockchain-based HIE in which NFT protocols are embedded can be used nationwide.

Moreover, in the United States, different states have diverse rules and regulations regarding the ownership of medical data (ranging from no clear laws to stringent regulatory frameworks). Because of various regulatory strictness, some states will likely create favorable environments that try to adopt applications of NFTs in HIE networks; other states might ban the use of NFTs outright. It would be an interesting research area for future studies to shed more light on the concept of NFTs (especially in health care) from a regulatory perspective.

The second challenge is the cost of creating NFTs. A possible barrier is the additional cost of minting NFTs. In this case, how would this impact cost and convenience, and who will bear the cost of creating and minting associated NFTs? For instance, do care providers and patients jointly contribute to creating patients' medical records, or is this responsibility for health care organizations? These questions can explain the complexity of adding NFTs to the blockchain HIE and the incremental benefit of this change. Thus, there is a lack of clarity on whether expanding NFTs' functions in health care is a financially feasible project. Minting NFTs assigned to PHI on a permissioned blockchain requires a robust technological infrastructure with stringent security safeguards. Therefore, further research is required to examine the phenomenon of NFTs and their application in HIE from a financial perspective.

Third, NFTs in health care could be promising; however, their implementation remains challenging. Different stakeholders in the health care ecosystem and layers in the NFT-based HIEs architecture require robust protocols for stakeholder collaboration and interaction. For example, most patient visits are attributed to older patients (older than 50 years of age), who may not be technology proficient, may require extensive training to understand the technology, and may need to provide access to providers for their medical records. What would happen if certain medical records (older or generated through nonparticipating providers) cannot be converted to NFTs? What will happen if a patient is incapacitated (or not in the correct mental state) and cannot grant access to medical records for urgently needed care?

The fourth challenge refers to user perception, as little is known about whether potential users of information-sharing projects in health care will accept NFT-based HIE. As the NFT concept is still new and there is a lack of public awareness about this phenomenon, many questions remain unanswered regarding the perceived viability, utility, and value of NFTs. Thus, further studies are needed to investigate how users (such as physicians and patients) may adapt to NFT technology in health care settings. For instance, researchers can examine the value of NFT-based HIE from user perspectives, such as ease of use, usefulness, cost-effectiveness, error reduction, and productivity.

A Framework for Further Investigation

As the use of NFTs in blockchain-based HIE systems continues to evolve, further research is needed on the design and implementation of these systems. This research could entail the following areas:

- Technical design considerations: there is a need for further research into the technical design considerations of integrating NFTs into blockchain-based HIE systems. This could include exploring optimal blockchain technology for HIE, designing smart contracts that govern NFT exchange and storage, and developing efficient authentication mechanisms.
- Regulatory and legal considerations: there is a need for further research into the regulatory and legal considerations of NFTs in health care. This could include exploring the legal implications of exchanging health care information via NFTs, the potential impact on patient privacy, and the role of regulatory bodies in overseeing the use of NFTs in health care.
- User acceptance and adoption: further research is needed to understand user acceptance and adoption of NFTs in health care. This could involve assessing the usability of NFT-based HIE systems, identifying barriers to adoption, and understanding the perspectives of health care providers and patients.
- Data security and privacy: there is a need for further research into the data security and privacy implications of using NFTs in health care. This could involve exploring the potential vulnerabilities of NFT-based HIE systems,

designing robust security mechanisms, and identifying potential threats to patient-data privacy.

Overall, future studies should provide insights into the design and implementation of NFT-based HIE systems that are secure, efficient, and user-friendly while also addressing regulatory and legal challenges and protecting patient data privacy.

Evolution of Information-Sharing Technology in Health Care

Table 1 summarizes the growth paths of technologies used in information-sharing initiatives in health care. The types of technology and examples of sharing mechanisms for each initiative are described. Moreover, the key challenges of each technology are highlighted, accompanied by the changes required to address those issues, which lead to the transition to the next technological advancement.

Table 1. Comparison between different information-sharing initiatives in health care.

Information-sharing initiatives	Type of technology	Example of sharing mechanisms	Challenges	Changes required
Traditional models	Conventional models (paper-based or voice-based)	Fax, mail, CD, and phone calls	1. The inability to provide timely access to patients' medical records 2. Performing unnecessary and repetitive medical tests 3. Cannot integrate patient data into a central hub 4. Chance of losing data 5. Inconvenient for patients to carry nonelectronic records 6. Space and costs of storing files 7. Security and privacy risks 8. Additional workload for physicians	1. Digitalization of medical records 2. Developing a central database to store patient data 3. Facilitating interoperability across multiple health care entities 4. Standards of data storage and transfer should be determined
HIE[a] networks	Centralized platforms (databases + emails + patient portals)	Direct exchange, look-up, and patient-centered	1. Implementation issues (organizational, financial, and governance barriers) 2. Lack of certified EHRs[b] 3. Interorganizational partnerships with unaffiliated health care organizations 4. Trust-based networks 5. Privacy concerns and risks of a data breach 6. Various patient consent policies 7. Lack of transparency on sharing procedures	1. Developing decentralized networks so that multiple stakeholders can overlook data sharing 2. More transparency in sharing patient data 3. The threat of a single point of failure should be solved 4. More stringent security measures should be applied 5. Data ownership should be clear 6. Better mechanisms for authentication and granting permission to access data should be used
Blockchain-based HIE	Decentralized platforms	Permissioned blockchain, federated blockchain, and smart contracts	1. Lack of awareness about blockchain applications in health care 2. Lack of regulations and guidelines 3. Little is known about the perceptions of potential users 4. Lack of incentives for sharing medical records	1. More organizational training and marketing strategies to promote blockchain applications in health care 2. Need for federal and state-based regulations dedicated to the use of blockchain in health care projects 3. Incentive mechanisms are required to encourage information sharing 4. Patient medical data can be treated as a nonfungible asset
NFT[c]-based HIE	Decentralized platforms	Permissioned blockchain, federated blockchain, smart contracts, and NFTs	1. NFT technology is still novel 2. Lack of dedicated regulations for NFTs 3. Lack of research on the feasibility of NFT-based HIE 4. Market traction	1. More research is required on the practicality, viability, value, and utility of using NFT technology in health care 2. Types of incentives should be studied 3. New amendments, compliance, and dedicated regulatory framework for NFTs 4. Implementation barriers to minting NFT in health care and required protocols for interactions with stakeholders should be addressed

[a]HIE: health information exchange.

[b]EHR: electronic health record.

[c]NFT: nonfungible token.

Conclusions

Overview

This study sheds light on the characteristics of emerging technologies that support health information–sharing efforts. Rapid technological advancements are accompanied by higher security risks, such as authenticity. We evaluated the potential of NFTs as a novel technology that can be leveraged in new use cases such as health care to mainly solve ownership and authenticity problems. The use of NFTs in HIE systems has the potential to revolutionize the health care industry by enabling the secure and efficient sharing of patient health information. NFT-based HIE may perform existing information exchange functions differently. The benefits of using NFTs include enhanced data security and privacy, improved interoperability, and streamlined data exchanges. We believe NFT technology can be a good fit for HIE networks because, first, NFTs are noninterchangeable. Each NFT is linked to a digital PHI that specifies the medical record’s values, ownership, and sharing rights. Second, NFTs are immutable; thus, they cannot be altered, manipulated, or forged in the information-sharing process. Third, every NFT needs to have an owner, and this is a public record that is easy for anyone to verify. In the proposed NFT-based HIE, patients are the original owners of their PHI, and other entities (such as providers) may be granted the right to check, analyze, and share such medical records based on the terms and conditions defined in a smart contract. NFTs can provide secure records of ownership and authentication in HIE networks. However, several challenges must be addressed before the widespread adoption of NFTs in HIE systems. In addition to the distinguishing features of NFTs, this technology presently faces a lack of dedicated NFT regulation due to its novelty and weakly enforced markets. For example, developing a regulatory framework to control NFT activities could help reduce the high degree of uncertainty in NFTs by forcing creators to obey specific guidelines. The level of regulatory clarity regarding NFTs can encourage more entrepreneurs to invest in different use cases (such as in health care). These challenges include the need for technical standards and infrastructure, legal and regulatory issues, and concerns regarding scalability and sustainability. Overall, although challenges need to be addressed, the benefits of using NFTs in HIE systems outweigh their drawbacks and offer promising opportunities for improving health care outcomes. Further research and development are necessary to address these challenges and fully realize the potential of NFTs in HIE systems. This study suggests that adding NFTs to HIE frameworks could be promising; however, further research is required to validate the value of this change.

Availability of Data and Materials

All the data analyzed in this study are included in this published article.

Conflicts of Interest

None declared.

References

1. Cherry JC, Dryden K, Kobb R, Hilsen P, Nedd N. Opening a window of opportunity through technology and coordination: a multisite case study. Telemed J E Health 2003;9(3):265-271. [doi: 10.1089/153056203322502650] [Medline: 14611694]
2. Esmaeilzadeh P. Patients' perceptions of different information exchange mechanisms: an exploratory study in the United States. Methods Inf Med 2020 Aug;59(4-05):162-178. [doi: 10.1055/s-0040-1721784] [Medline: 33618421]
3. Hatef E, Weiner JP, Kharrazi H. A public health perspective on using electronic health records to address social determinants of health: the potential for a national system of local community health records in the United States. Int J Med Inform 2019 Apr;124:86-89. [doi: 10.1016/j.ijmedinf.2019.01.012] [Medline: 30784431]
4. Yeung T. Local health department adoption of electronic health records and health information exchanges and its impact on population health. Int J Med Inform 2019 Aug;128:1-6. [doi: 10.1016/j.ijmedinf.2019.04.011] [Medline: 31160006]
5. Esmaeilzadeh P, Sambasivan M. Health information exchange (HIE): a literature review, assimilation pattern and a proposed classification for a new policy approach. J Biomed Inform 2016 Dec;64:74-86 [FREE Full text] [doi: 10.1016/j.jbi.2016.09.011] [Medline: 27645322]
6. Vest JR, Gamm LD. Health information exchange: persistent challenges and new strategies. J Am Med Inform Assoc 2010 May;17(3):288-294 [FREE Full text] [doi: 10.1136/jamia.2010.003673] [Medline: 20442146]
7. Menachemi N, Rahurkar S, Harle CA, Vest JR. The benefits of health information exchange: an updated systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1259-1265 [FREE Full text] [doi: 10.1093/jamia/ocy035] [Medline: 29718258]
8. Esmaeilzadeh P, Mirzaei T. Comparison of consumers' perspectives on different health information exchange (HIE) mechanisms: an experimental study. Int J Med Inform 2018 Nov;119:1-7. [doi: 10.1016/j.ijmedinf.2018.08.007] [Medline: 30342677]
9. Wright A, Soran C, Jenter CA, Volk LA, Bates DW, Simon SR. Physician attitudes toward health information exchange: results of a statewide survey. J Am Med Inform Assoc 2010 Jan;17(1):66-70 [FREE Full text] [doi: 10.1197/jamia.M3241] [Medline: 20064804]
10. Yeager VA, Walker D, Cole E, Mora AM, Diana ML. Factors related to health information exchange participation and use. J Med Syst 2014 Aug;38(8):78. [doi: 10.1007/s10916-014-0078-1] [Medline: 24957395]

11. Esmaeilzadeh P, Sambasivan M. Patients' support for health information exchange: a literature review and classification of key factors. BMC Med Inform Decis Mak 2017 Apr 04;17(1):33 [FREE Full text] [doi: 10.1186/s12911-017-0436-2] [Medline: 28376785]
12. Thorn SA, Carter MA, Bailey JE. Emergency physicians' perspectives on their use of health information exchange. Ann Emerg Med 2014 Mar;63(3):329-337. [doi: 10.1016/j.annemergmed.2013.09.024] [Medline: 24161840]
13. Feldman SS, Schooley BL, Bhavsar GP. Health information exchange implementation: lessons learned and critical success factors from a case study. JMIR Med Inform 2014 Aug 15;2(2):e19 [FREE Full text] [doi: 10.2196/medinform.3455] [Medline: 25599991]
14. Wu YH, Cristancho-Lacroix V, Fassert C, Faucounau V, de Rotrou J, Rigaud AS. The attitudes and perceptions of older adults with mild cognitive impairment toward an assistive robot. J Appl Gerontol 2016 Jan;35(1):3-17. [doi: 10.1177/0733464813515092] [Medline: 24652924]
15. Adler-Milstein J, Pfeifer E. Information blocking: is it occurring and what policy strategies can address it? Milbank Q 2017 Mar;95(1):117-135 [FREE Full text] [doi: 10.1111/1468-0009.12247] [Medline: 28266065]
16. Grossman JM, Kushner KL, November EA. Creating sustainable local health information exchanges: can barriers to stakeholder participation be overcome? Res Brief 2008 Feb(2):1-12. [Medline: 18496926]
17. Williams C, Mostashari F, Mertz K, Hogin E, Atwal P. From the office of the national coordinator: the strategy for advancing the exchange of health information. Health Aff (Millwood) 2012 Mar;31(3):527-536. [doi: 10.1377/hlthaff.2011.1314] [Medline: 22392663]
18. Roehrs A, da Costa CA, da Rosa Righi R, da Silva VF, Goldim JR, Schmidt DC. Analyzing the performance of a blockchain-based personal health record implementation. J Biomed Inform 2019 Apr;92:103140 [FREE Full text] [doi: 10.1016/j.jbi.2019.103140] [Medline: 30844481]
19. Blumenthal D, Tavenner M. The "meaningful use" regulation for electronic health records. N Engl J Med 2010 Aug 05;363(6):501-504. [doi: 10.1056/NEJMp1006114] [Medline: 20647183]
20. Braunstein ML. Healthcare in the age of interoperability: the promise of fast healthcare interoperability resources. IEEE Pulse 2018 Nov;9(6):24-27. [doi: 10.1109/MPUL.2018.2869317] [Medline: 30452344]
21. Christodoulou K, Christodoulou P, Zinonos Z, Carayannis EG, Chatzichristofis SA. Health information exchange with blockchain amid COVID-19-like pandemics. In: Proceedings of the 16th International Conference on Distributed Computing in Sensor Systems. 2020 Presented at: DCOSS' 20; May 25-27, 2020; Marina del Rey, CA, USA p. 412-417. [doi: 10.1109/dcoss49796.2020.00071]
22. Shen N, Bernier T, Sequeira L, Strauss J, Silver MP, Carter-Langford A, et al. Understanding the patient privacy perspective on health information exchange: a systematic review. Int J Med Inform 2019 May;125:1-12. [doi: 10.1016/j.ijmedinf.2019.01.014] [Medline: 30914173]
23. Esmaeilzadeh P. Benefits and concerns associated with blockchain-based health information exchange (HIE): a qualitative study from physicians' perspectives. BMC Med Inform Decis Mak 2022 Mar 28;22(1):80 [FREE Full text] [doi: 10.1186/s12911-022-01815-8] [Medline: 35346176]
24. Lee D, Song M. MEXchange: a privacy-preserving blockchain-based framework for health information exchange using ring signature and stealth address. IEEE Access 2021 Nov 25;9:158122-158139 [FREE Full text] [doi: 10.1109/access.2021.3130552]
25. Jiang S, Cao J, Wu H, Yang Y, Ma M, He J. BlocHIE: a BLOCkchain-based platform for healthcare information exchange. In: Proceedings of the 2018 International Conference on Smart Computing. 2018 Presented at: SMARTCOMP' 18; June 18-20, 2018; Taormina, Italy p. 49-56. [doi: 10.1109/smartcomp.2018.00073]
26. Abbas Y, Martinetti A, Moerman JJ, Hamberg T, van Dongen LA. Do you have confidence in how your rolling stock has been maintained? A blockchain-led knowledge-sharing platform for building trust between stakeholders. Int J Inf Manage 2020 Dec;55:102228 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2020.102228]
27. Wamba SF, Queiroz MM. Blockchain in the operations and supply chain management: benefits, challenges and future research opportunities. Int J Inf Manage 2020 Jun;52:102064 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2019.102064]
28. Lu Y. Blockchain and the related issues: a review of current research topics. J Manag Anal 2018 Sep 03;5(4):231-255 [FREE Full text] [doi: 10.1080/23270012.2018.1516523]
29. Iansiti M, Lakhani KR. The truth about blockchain. Harvard Business Review. 2017 Jan. URL: https://hbr.org/2017/01/the-truth-about-blockchain [accessed 2019-02-02]
30. Xia Q, Sifah EB, Smahi A, Amofa S, Zhang X. BBDS: blockchain-based data sharing for electronic medical records in cloud environments. Information 2017 Apr 17;8(2):44 [FREE Full text] [doi: 10.3390/info8020044]
31. Kuo TT, Kim HE, Ohno-Machado L. Blockchain distributed ledger technologies for biomedical and health care applications. J Am Med Inform Assoc 2017 Nov 01;24(6):1211-1220 [FREE Full text] [doi: 10.1093/jamia/ocx068] [Medline: 29016974]
32. Esmaeilzadeh P, Mirzaei T. The potential of blockchain technology for health information exchange: experimental study from patients' perspectives. J Med Internet Res 2019 Jun 20;21(6):e14184 [FREE Full text] [doi: 10.2196/14184] [Medline: 31223119]

33. Soni H, Grando A, Murcko A, Diaz S, Mukundan M, Idouraine N, et al. State of the art and a mixed-method personalized approach to assess patient perceptions on medical record sharing and sensitivity. J Biomed Inform 2020 Jan;101:103338 [FREE Full text] [doi: 10.1016/j.jbi.2019.103338] [Medline: 31726102]
34. Nembaware V, Johnston K, Diallo AA, Kotze MJ, Matimba A, Moodley K, et al. A framework for tiered informed consent for health genomic research in Africa. Nat Genet 2019 Nov;51(11):1566-1571 [FREE Full text] [doi: 10.1038/s41588-019-0520-x] [Medline: 31659323]
35. Toh S, Pratt N, Klungel O, Gagne JJ, Platt RW. Distributed networks of databases analyzed using common protocols and/or common data models. In: Strom BL, Kimmel SE, Hennessy S, editors. Pharmacoepidemiology. 6th edition. New York, NY, USA: Wiley-Blackwell; 2019:617-638.
36. Yue X, Wang H, Jin D, Li M, Jiang W. Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control. J Med Syst 2016 Oct;40(10):218. [doi: 10.1007/s10916-016-0574-6] [Medline: 27565509]
37. Azaria A, Ekblaw A, Vieira T, Lippman A. MedRec: using blockchain for medical data access and permission management. In: Proceedings of the 2nd International Conference on Open and Big Data. 2016 Presented at: OBD' 16; August 22-24, 2016; Vienna, Austria p. 25-30. [doi: 10.1109/obd.2016.11]
38. Philippakis AA, Azzariti DR, Beltran S, Brookes AJ, Brownstein CA, Brudno M, et al. The Matchmaker Exchange: a platform for rare disease gene discovery. Hum Mutat 2015 Oct;36(10):915-921 [FREE Full text] [doi: 10.1002/humu.22858] [Medline: 26295439]
39. Lu Y, Huang X, Zhang K, Maharjan S, Zhang Y. Communication-efficient federated learning and permissioned blockchain for digital twin edge networks. IEEE Internet Things J 2021 Feb 15;8(4):2276-2288 [FREE Full text] [doi: 10.1109/jiot.2020.3015772]
40. Helliar CV, Crawford L, Rocca L, Teodori C, Veneziani M. Permissionless and permissioned blockchain diffusion. Int J Inf Manage 2020 Oct;54:102136 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2020.102136]
41. Singh S, Rathore S, Alfarraj O, Tolba A, Yoon B. A framework for privacy-preservation of IoT healthcare data using federated learning and blockchain technology. Future Gener Comput Syst 2022 Apr;129:380-388 [FREE Full text] [doi: 10.1016/j.future.2021.11.028]
42. Olfson M, Wall MM, Blanco C. Incentivizing data sharing and collaboration in medical research-the S-index. JAMA Psychiatry 2017 Jan 01;74(1):5-6. [doi: 10.1001/jamapsychiatry.2016.2610] [Medline: 27784040]
43. Mohanta BK, Panda SS, Jena D. An overview of smart contract and use cases in blockchain technology. In: Proceedings of the 9th International Conference on Computing, Communication and Networking Technologies. 2018 Presented at: ICCCNT '18; July 10-12, 2018; Bengaluru, India p. 1-4. [doi: 10.1109/icccnt.2018.8494045]
44. Hewa T, Ylianttila M, Liyanage M. Survey on blockchain based smart contracts: applications, opportunities and challenges. J Netw Comput Appl 2021 Mar;177:102857 [FREE Full text] [doi: 10.1016/j.jnca.2020.102857]
45. Wu H, Zhong B, Li H, Guo J, Wang Y. On-site construction quality inspection using blockchain and smart contracts. J Manage Eng 2021 Nov;37(6):04021065 [FREE Full text] [doi: 10.1061/(asce)me.1943-5479.0000967]
46. Lee JS, Chew CJ, Liu JY, Chen YC, Tsai KY. Medical blockchain: data sharing and privacy preserving of EHR based on smart contract. J Inf Secur Appl 2022 Mar;65:103117 [FREE Full text] [doi: 10.1016/j.jisa.2022.103117]
47. Rehman W, Zainab H, Imran J, Bawany NZ. NFTs: applications and challenges. In: Proceedings of the 22nd International Arab Conference on Information Technology. 2021 Presented at: ACIT' 21; December 21-23, 2021; Muscat, Oman p. 1-7. [doi: 10.1109/acit53391.2021.9677260]
48. Wilson KB, Karg A, Ghaderi H. Prospecting non-fungible tokens in the digital economy: stakeholders and ecosystem, risk and opportunity. Bus Horiz 2022 Sep;65(5):657-670 [FREE Full text] [doi: 10.1016/j.bushor.2021.10.007]
49. Chohan R, Paschen J. NFT marketing: how marketers can use nonfungible tokens in their campaigns. Bus Horiz 2022;66(1):43-50 [FREE Full text] [doi: 10.1016/j.bushor.2021.12.004]
50. Belk R, Humayun M, Brouard M. Money, possessions, and ownership in the metaverse: NFTs, cryptocurrencies, Web3 and wild markets. J Bus Res 2022 Dec;153:198-205 [FREE Full text] [doi: 10.1016/j.jbusres.2022.08.031]
51. Musamih A, Salah K, Jayaraman R, Yaqoob I, Puthal D, Ellahham S. NFTs in healthcare: vision, opportunities, and challenges. IEEE Consum Electron Mag (forthcoming) 2022 Aug 04:1-11. [doi: 10.1109/mce.2022.3196480]
52. Psaras Y, Gipp B, Schubotz M, Scott W, Castro I, Tyson G, et al. Design and evaluation of IPFS: a storage layer for the decentralized web. In: Proceedings of the ACM SIGCOMM 2022 Conference. 2022 Presented at: SIGCOMM '22; August 22-26, 2022; Amsterdam, The Netherlands p. 739-752. [doi: 10.1145/3544216.3544232]
53. Guadamuz A. The treachery of images: non-fungible tokens and copyright. J Intellect Prop Law Pract 2021 Dec;16(12):1367-1385 [FREE Full text] [doi: 10.1093/jiplp/jpab152]
54. Chalmers D, Fisch C, Matthews R, Quinn W, Recker J. Beyond the bubble: will NFTs and digital proof of ownership empower creative industry entrepreneurs? J Bus Ventur Insights 2022 Jun;17:e00309 [FREE Full text] [doi: 10.1016/j.jbvi.2022.e00309]
55. Peng L, Feng W, Yan Z, Li Y, Zhou X, Shimizu S. Privacy preservation in permissionless blockchain: a survey. Digit Commun Netw 2021 Aug;7(3):295-307 [FREE Full text] [doi: 10.1016/j.dcan.2020.05.008]
56. Wüst K, Gervais A. Do you need a blockchain? In: Proceedings of the 2018 Crypto Valley Conference on Blockchain Technology. 2018 Presented at: CVCBT '18; June 20-22, 2018; Zug, Switzerland p. 45-54. [doi: 10.1109/cvcbt.2018.00011]

57. Platt M, McBurney P. Sybil in the haystack: a comprehensive review of blockchain consensus mechanisms in search of strong sybil attack resistance. Algorithms 2023 Jan 06;16(1):34 [FREE Full text] [doi: 10.3390/a16010034]
58. Wang Q, Li R, Wang Q, Chen S. Non-fungible token (NFT): overview, evaluation, opportunities and challenges. aiXiv. Preprint posted online May 16, 2021. [FREE Full text]
59. Miyachi K, Mackey TK. hOCBS: a privacy-preserving blockchain framework for healthcare data leveraging an on-chain and off-chain system design. Inf Process Manag 2021 May;58(3):102535 [FREE Full text] [doi: 10.1016/j.ipm.2021.102535]
60. Blaney JE, Middleton KE. Using NFTs to store health data: a new era or a privacy disaster. Bus Law Int 2022 Sep 01;23(3):1-17 [FREE Full text]
61. Zhang Q. An overview and analysis of hybrid encryption: the combination of symmetric encryption and asymmetric encryption. In: Proceedings of the 2nd International Conference on Computing and Data Science. 2021 Presented at: CDS '21; January 28-29, 2021; Stanford, CA, USA p. 616-622. [doi: 10.1109/cds52072.2021.00111]
62. Kietzmann J, Lee LW, McCarthy IP, Kietzmann TC. Deepfakes: trick or treat? Bus Horiz 2020 Mar;63(2):135-146 [FREE Full text] [doi: 10.1016/j.bushor.2019.11.006]
63. Henderi, Gunawan IK, Sukmana HT, Ardianto AY. Blockchain technology as a media for sharing information that generates user access rights and incentives. Blockchain Front Technol 2021 Jul 11;1(1):44-55 [FREE Full text] [doi: 10.34306/bfront.v1i01.2]
64. Riesco R, Larriva-Novo X, Villagra VA. Cybersecurity threat intelligence knowledge exchange based on blockchain: proposal of a new incentive model based on blockchain and smart contracts to foster the cyber threat and risk intelligence exchange of information. Telecommun Syst 2020;73(2):259-288 [FREE Full text] [doi: 10.1007/s11235-019-00613-4]
65. Khalil U, Malik OA, Uddin M, Chen CL. A comparative analysis on blockchain versus centralized authentication architectures for IoT-enabled smart devices in smart cities: a comprehensive review, recent advances, and future research directions. Sensors (Basel) 2022 Jul 10;22(14):5168 [FREE Full text] [doi: 10.3390/s22145168] [Medline: 35890848]
66. Liddell K, Simon DA, Lucassen A. Patient data ownership: who owns your health? J Law Biosci 2021 Oct 1;8(2):lsab023 [FREE Full text] [doi: 10.1093/jlb/lsab023] [Medline: 34611493]
67. Rehman S, Talat Bajwa N, Shah MA, Aseeri AO, Anjum A. Hybrid AES-ECC model for the security of data over cloud storage. Electronics 2021 Oct 31;10(21):2673 [FREE Full text] [doi: 10.3390/electronics10212673]
68. Holmgren AJ, Adler-Milstein J. Health information exchange in US hospitals: the current landscape and a path to improved information sharing. J Hosp Med 2017 Mar;12(3):193-198. [doi: 10.12788/jhm.2704] [Medline: 28272599]
69. Kostick-Quenet K, Mandl KD, Minssen T, Cohen IG, Gasser U, Kohane I, et al. How NFTs could transform health information exchange. Science 2022 Feb 04;375(6580):500-502. [doi: 10.1126/science.abm2004] [Medline: 35113709]

Abbreviations

EHR: electronic health record
HIE: health information exchange
HIPAA: Health Insurance Portability and Accountability Act
KMS: key management system
NFT: nonfungible token
PBFT: Practical Byzantine Fault Tolerance
PHI: personal health information
PoS: Proof of Stake
PoW: Proof of Work

Edited by A Mavragani; submitted 13.09.22; peer-reviewed by M Platt, B Sainz-de-Abajo, S Ashraf, T Moody; comments to author 21.02.23; revised version received 03.03.23; accepted 16.03.23; published 12.04.23.

Please cite as:
Esmaeilzadeh P
Evolution of Health Information Sharing Between Health Care Organizations: Potential of Nonfungible Tokens
Interact J Med Res 2023;12:e42685
URL: https://www.i-jmr.org/2023/1/e42685
doi:10.2196/42685
PMID:37043269

Viewpoint

Strategies to Bridge Equitable Implementation of Telehealth

Allison M Gustavson[1,2], DPT, PhD; Allison A Lewinski[3,4], MPH, PhD; Ellen E Fitzsimmons-Craft[5], PhD; Gloria D Coronado[6], PhD; Sarah E Linke[7], MPH, PhD; Denalee M O'Malley[8,9], PhD; Alyce S Adams[10], PhD; Russell E Glasgow[11], PhD; Lisa M Klesges[12], MS, PhD

[1]Center for Care Delivery and Outcomes Research, Minneapolis Veterans Affairs Health Care System, Minneapolis, MN, United States
[2]Department of Medicine, University of Minnesota, Minneapolis, MN, United States
[3]Center of Innovation to Accelerate Discovery and Practice Transformation, Durham Veterans Affairs Health Care System, Durham, NC, United States
[4]School of Nursing, Duke University, Durham, NC, United States
[5]Department of Psychiatry, Washington University School of Medicine, St Louis, MO, United States
[6]Kaiser Permanente Center for Health Research, Portland, OR, United States
[7]Herbert Wertheim School of Public Health and Human Longevity Science, University of California San Diego, La Jolla, CA, United States
[8]Department of Family Medicine and Community Health, Research Division, Rutgers Robert Wood Johnson Medical School, New Brunswick, NJ, United States
[9]Rutgers Cancer Institute of New Jersey, New Brunswick, NJ, United States
[10]Stanford Cancer Institute, Stanford, CA, United States
[11]Department of Family Medicine and Adult & Child Consortium for Health Outcomes Research and Delivery Science, University of Colorado School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, United States
[12]Division of Public Health Sciences, Washington University School of Medicine, St Louis, MO, United States

Corresponding Author:
Allison M Gustavson, DPT, PhD
Center for Care Delivery and Outcomes Research
Minneapolis Veterans Affairs Health Care System
1 Veterans Drive
Minneapolis, MN, 55417
United States
Phone: 1 612 467 3406
Fax: 1 612 727 5699
Email: allison.gustavson@va.gov

Abstract

During the COVID-19 pandemic, the rapid scaling of telehealth limited the extent to which proactive planning for equitable implementation was possible. The deployment of telehealth will persist in the postpandemic era, given patient preferences, advances in technologies, growing acceptance of telehealth, and the potential to overcome barriers to serve populations with limited access to high-quality in-person care. However, aspects and unintended consequences of telehealth may leave some groups underserved or unserved, and corrective implementation plans that address equitable access will be needed. The purposes of this paper are to (1) describe equitable implementation in telehealth and (2) integrate an equity lens into actionable equitable implementation.

(Interact J Med Res 2023;12:e40358) doi:10.2196/40358

KEYWORDS

implementation science; equity; telehealth; equitable implementation; digital age; post pandemic

Background

The COVID-19 pandemic catalyzed the rapid development, implementation, and scaling of telehealth, which we define for this commentary as the synchronous delivery of health care services by phone or video [1,2]. Prior to the pandemic, the implementation of telehealth was variable whereby providers, patients, and organizations could self-select to use telehealth (eg, opt-in if available). During the COVID-19 public health emergency, telehealth became the primary option for receipt of many health care services for most patients (eg, all-in). Rapid implementation often occurred with limited prior knowledge about telehealth and based on a selective sample of providers who were willing to offer it, often for only select problems that would be reimbursed, and with a selected sample of individuals who opted into using telehealth. As a result, significant gaps in equitable implementation exist given the rapid nature of the roll-out, which likely reinforces health disparities in health care access for already marginalized patient populations [3,4].

Implementation science plays a critical role in bridging the gap between the implementation of telehealth and equity. Implementation research involves understanding, evaluating, and providing strategies that enhance how much and how well telehealth is accessed, delivered, and received for the right patient at the right time [5]. This commentary addresses a gap in our understanding of *how* telehealth should be equitably implemented, adapted, and sustained to reach entire target populations (including those most in need or historically excluded) and diverse institutions (eg, high- and low-resourced institutions). It also highlights an urgent need to address unintended consequences of widespread telehealth and apply strategies to ameliorate inequity where possible. We present a broad perspective on equitable implementation of telehealth and provide discussion and recommendations through the literature, an illustrative example, and our own practical experiences. The specific purposes of this commentary are to (1) describe the importance of and nuances in equitable implementation in telehealth and (2) integrate an equity lens into actionable equitable implementation processes. Recognizing the complexities in equitable implementation of telehealth, we provide a perspective on how implementation can either exacerbate or proactively address inequity. Other authors have described inequalities in patients who have access to, use, and adhere to interventions and intervention inequalities if technology-based interventions are not equally effective for all [6]. The focus of this paper is not on the intervention but describes how adapting existing implementation frameworks have the potential to enhance equity-focused decision-making during *implementation* to facilitate equitable telehealth outcomes.

Importance of Equitable Implementation

Implementation science extends its long focus on health equity [7-12] to play an essential role in understanding, adapting, and reevaluating the integration of telehealth [5,13]. While evidence-based interventions existed for telehealth practices in selected settings, contexts, and populations before the pandemic [14-17], the urgency with which these practices were adopted during the pandemic limited more deliberate evaluation of their expanded implementation. Of particular concern was the inability to consider the often sparse evidence available and to evaluate initial conditions of inequity and other contextual factors. Many systems adopted telehealth based on resources available, lessons learned from collaboration with other systems, and practical experience. This helped with the expediency required but did not allow for the careful consideration needed to avoid unintended consequences, including the potential to create or exacerbate inequities.

To guide *equitable implementation*, it is ideal to begin with a framework that accounts for social disadvantage and injustice [8,10,12,18,19]. Implementation frameworks [11]—particularly the Health Equity Implementation Framework by Woodward et al [7] and the Consolidated Framework for Implementation Research by Damschroder et al [20]—articulate conceptual models to understand *determinants* of health equity to better adapt interventions and implementation strategies. Other frameworks have focused specifically on digital equity [21,22]; of note is the Digital Health Equity Framework (DHEF) [23,24]. The DHEF considers the multilevel, ecological impact when digital determinants of health (how digital health technologies influence equity in health) interact with intermediate health factors (eg, environment, current health status, and health-related beliefs and behaviors) [25]. While these digital and equity-focused frameworks provide a critical foundation for identifying and measuring different factors related to health equity, we are still left with *how* to integrate this knowledge into equitable implementation strategies and evaluation of telehealth.

Framework Consideration to Maximize Equitable Implementation of Telehealth

Our goal was to provide a broad perspective and guidance to health care systems and researchers on strategies to equitably use and evaluate the implementation of telehealth. We chose to illustrate our perspective by selecting and integrating exemplar frameworks that capture the rapid speed at which telehealth is being adopted and implemented across disciplines and health care settings, as well as contextual factors that might influence equitable outcomes. We also sought an integrated framework that could expand beyond understanding implementation determinants to describe *processes* by which equity or inequities are driven by the interaction with and context of the external and internal environments.

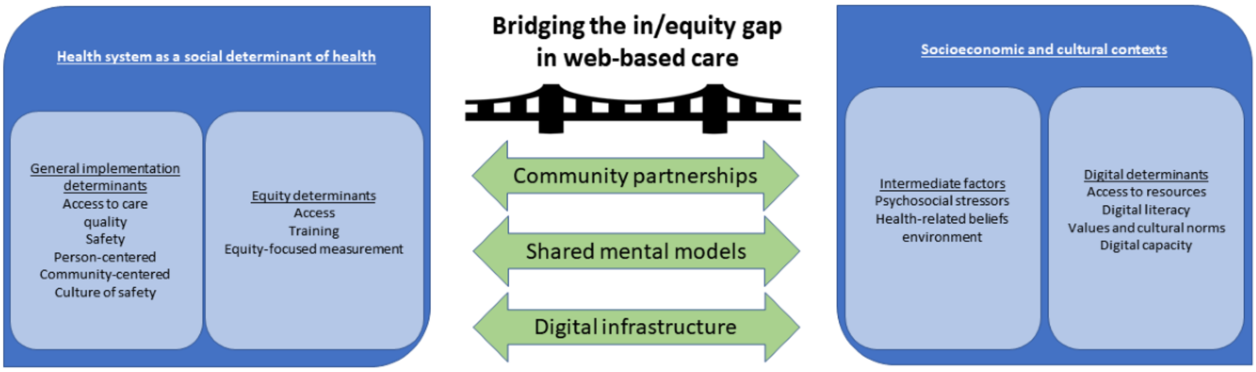
With these considerations, we integrated the EPIS (Exploration, Preparation, Implementation, and Sustainment) and DHEF to encapsulate equity within the process of implementation, where EPIS guides us in moving from concept to impact and the DHEF tells us where to focus if we want to impact equity (Figure 1) [25-28]. In addition, we wanted to focus on the iterative nature of implementation and the multiphase EPIS conceptualization fit this need especially well. The rapid cycle guidance provided by EPIS is advantageous in the case of telehealth as—for the most part—exploration and preparation phases were accelerated to rapidly move to emergency implementation during the initial stages of the pandemic and limited the ability to conduct a thorough community needs assessment.

EPIS encompasses a 4-phase implementation cycle (Exploration, Preparation, Implementation, and Sustainment) and describes implementation processes, inner and outer contextual factors, and bridging factors that facilitate the interplay between inner and outer factors through each of these phases [26]. For example, internal factors may be parsed out into organizational (eg, leadership decision-making, capacity, and resources to deliver telehealth) and individual levels (acceptability, technology skills and proficiency, and literacy). External factors may include federal or state policies and reimbursement surrounding telehealth, investment in digital infrastructure (eg, technical support and equipment), and reimbursement policies around what providers can deliver telehealth and in what instances.

The DHEF adds an equity focus to the internal and external factors in EPIS [29]. The DHEF was developed to consider the multilevel health equity factors that can reduce or exacerbate disparities in access to and receipt of digital health technologies (eg, telehealth, mobile health apps, web-based health services, and wearable technologies) [23]. Importantly, the DHEF expands upon the concept that the health system is a social determinant of health, and therefore, organizations or systems need to look beyond patient-level factors to truly lead the implementation of telehealth with equity. When integrated with EPIS, we start to regroup internal and external factors into the health system as a social determinant of health and socioeconomic and cultural contexts, respectively (Figure 1). In this way, we capture both the (1) general implementation determinants (access to care, quality, and safety) [29] and equity determinants (eg, access, training, and equity-focused measurement) within the health care system and (2) the intermediate factors (eg, psychosocial stressors, health-related beliefs) and digital determinants (eg, access to digital resources, digital health literacy, and digital capacity building) in the broader societal context.

The integration of EPIS and the DHEF informs equitable implementation in telehealth by considering who can or cannot (1) access telehealth, (2) receive telehealth, and (3) deliver telehealth and *why*. An understanding of who in the population is not reached and why they may have been excluded can lead to diverse community and health system engagement and offer contextual adaptations to the telehealth clinical practice and implementation strategy.

Figure 1. Critical bridging solutions to equitable telehealth implementation that integrate the EPIS (Exploration, Preparation, Implementation, and Sustainment) and the Digital Health Equity Framework.



Bridging Solutions to Equity Across Multilevel Contexts

Overview

EPIS contains “bridging factors,” which consider and account for the interconnectedness and bidirectional nature of movement within and across internal and external factors that shape telehealth access, receipt, and delivery [26,28,30]. Instead of exploring internal and external contexts separately, bridging factors examine the interdependence of how external forces shape health systems and vice versa. While not explicitly cited as a bridging factor in the original EPIS model, we surmise that equity is a product of critical bridging factors that tie together various levels of context (potential drivers of inequities) important to telehealth implementation. For instance, complex dynamics of oppression and injustice may be operating on multiple levels (eg, ideological, internalized, interpersonal, and institutional) that require proactive, aligned bridging strategies to overcome [31-33].

In Figure 1, we propose 3 bridging solutions whose presence supports equity of telehealth access, receipt, and delivery: community partnerships, shared mental models, and digital infrastructure. These bridging solutions connect equity concepts of the health system as a social determinant of health and socioeconomic and cultural contexts (DHEF) over the course of an implementation process (EPIS). First, fostering *community partnerships* are the backbone of successful implementation research and can create the bidirectional flow of information necessary to align telehealth goals across care systems, individuals, and national or state policies. Second, *communicating mental models* nurtures the sharing of an interrelated set of beliefs that shapes a person’s expectations for the future and how they understand the ways the world works [34,35]. When mental models are shared across systems, challenges that were seen as intractable can be resolved to achieve a common care delivery improvement goal (eg, improve access to underserved populations) [28,35]. Finally, bolstering *digital infrastructure* bridges an organization’s human and technical resources to provide telehealth with a focus on advocacy at federal or state levels to incentivize payers to reimburse telehealth and invest in digital infrastructure. In future, if rapid deployment suggests bridging factors as key considerations for the equitable implementation of telehealth. Table 1 outlines a guide to questions that promote bridging solutions for the equitable implementation of telehealth.

Table 1. Guide to questions that promote bridging solutions for equitable implementation of telehealth.

Bridging issues for future exploration	Question for future research
Population	
• Patients who do not use, or minimally use, telehealth in health care delivery • Patients who experience challenges to health and wellness due to sociodemographic factors • Patients who experience intersecting health and sociodemographic factors	• What is the impact of (*age, sex, race/ethnicity*) on telehealth reach? • What are the factors of patients who do not (*receive, sustain, engage in*) telehealth? • What (*resources such as interpreters, accessible devices*) are necessary for equitable implementation of telehealth? • What structural factors (eg, *racism*) impact the equitable use of telehealth in a population?
Setting	
• Community-based health settings • Outpatient settings (eg, community-based clinics, large health care systems, small health care settings, rural clinics) • Primary care or specialty settings	• What contextual factors influence equitable delivery of telehealth at the (*person, provider, health care system*) level? • What factors promote or hinder equitable implementation of (*telehealth modality*) at the (*setting*) level? • In what ways does (*setting*) impact the delivery and receipt of (*telehealth modality*)? • How does the implementation of telehealth differ by local or national (*policy, available infrastructure*)? • How does adaptation of (*telehealth modality*) interventions differ by setting?
Outcomes	
• Patient-centered outcomes • Acceptability of telehealth (eg, access, use, sustainment) • Patient and clinician satisfaction • Harms and unintended consequences (eg, missed diagnoses, nonengagement of certain groups or individuals)	• How are outcomes culturally or contextually defined in telehealth use? • How do patient-reported outcomes differ by (*patient characteristic, subgroup*)? • What patient-reported outcomes differ by (*patient characteristic, sub-group*)?

Case Scenario

We offer a case scenario that illustrates the implementation phase of EPIS and considerations for equitable telehealth as outlined by DHEF. We first illustrate the health system as a social determinant of health and socioeconomic and cultural contexts defined by DHEF that drive the multilevel context in which the delivery of web-based physical therapy evolved over the pandemic. During the initial weeks of the COVID-19 pandemic in the United States, in-person services provided by outpatient physical therapists were immediately discontinued. Outpatient physical therapists across a multitude of health care systems and clinic networks transitioned from almost 100% in-person visits to almost 100% web-based visits in a matter of weeks [36,37]. This shift was significant in the physical therapy profession as prepandemic restrictions in the adoption of telehealth included both internal and external factors: reimbursement challenges, lack of organizational infrastructure to support web-based platforms of care, and limited provider education and training in web-based delivery of physical therapy services hampered widespread adoption of telehealth [38-40]. Additionally, before the pandemic, the physical therapy profession was already grappling with disparities in access to in-person physical therapy due to reduced staffing capacity. Staffing issues, in turn, precipitated long waitlists and high out-of-pocket costs for patients due to restrictive insurance policies for reimbursement of specialty physical therapy care [41-44]. Therefore, the quick deployment of telehealth may have further underscored unequal access for patients already experiencing challenges in receiving physical therapy care before the pandemic.

As the pandemic has progressed, so has the phased reopening of outpatient physical therapy services for in-person care. However, the value of telehealth—for example, the potential for expanded access and decreased transportation burden—has created momentum to continue the provision of physical therapy care via web-based modalities [38,45]. An operational response in some health care systems was to create threshold goals for the percentage of patients receiving in-person physical therapy care. For example, outpatient physical therapy clinics were expected to perform, say on average, 60% of visits as in-person care by a specified date in the phased reopening. Some health care systems and clinics have used additional policies that dictate a web-based visit must be initiated before an in-person visit, thus requiring the use of telehealth for entry into a physical therapy care pathway. In addition, it is important to note that the demand for these services may have increased during this period of time as well as the number of individuals seeking care due to rehabilitation after COVID-19 infection [46,47].

The context described above impacts groups at multiple levels and creates opportunities to enhance equity in the delivery of web-based physical therapy. As such, to consider further adaptations to and appropriate sustainment of web-based physical therapy care, we describe the bridging factors necessary to promote equitable implementation. First, *establishing community partnerships* is needed to engage patients, families, providers, and communities to better identify (needs assessment) who receives physical therapy (or not) when telehealth is offered and *why*. Methods to build community partnerships in implementation research offer opportunities for reflexivity and iteration, which informs strategies to ensure whether telehealth is delivered in a manner that is fair and just. For example, Miller et al [37] showed that patients reached by telehealth delivery of physical therapy during the pandemic were largely younger than 65 years, non-Hispanic White, English-speaking, commercially insured, and with few to no comorbidities. This contrasted with the distribution of patient characteristics seen for in-person physical therapy the year prior to the pandemic, many of whom were older than 65 years, Asian, non–English-speaking, noncommercially insured, and had at least 1 comorbidity. Community partnerships may enhance equitable implementation through the adaptation of physical therapy telehealth to the sociocultural context, thereby increasing the relevance of telehealth to marginalized patient populations and enhancing individual functional outcomes. Community partnerships between patients, providers, and operations are also needed to evaluate organizational capacity to provide both in-person and telehealth options based on patient needs and preferences.

Second, *communicating shared mental models* within and across systems and sectors allows groups of people delivering, receiving, or being impacted by physical therapy telehealth to be on the same page regarding equitable implementation. Telehealth delivery of physical therapy will likely persist post pandemic and adjusting how patients, providers, and systems perceive this new reality is essential to promoting equitable implementation. Discordant mental models may unintentionally hinder access to *any* modality of care such as in the case example where system or clinic policies drive the (1) proportion of telehealth versus in-person appointments available and (2) type of appointment necessary for entry into the care pathway. For example, individuals without stable internet access in a secure, private setting may be unable to engage in physical therapy if telehealth is initially required for entry into that service. Additionally, some patients may feel more comfortable receiving in-person care in a physical therapy clinic during which they can discuss sensitive topics influencing their recovery and feel less vulnerable undressing or exposing certain areas of their body for examination. Alternatively, some patients may feel more comfortable discussing sensitive topics in their own homes, thus creating an inviting atmosphere for greater sharing and conversation between patient and provider.

Convening a diverse group of community members can help build a shared mental model by asking questions such as the following: what proportion of telehealth visits per provider reaches the most patients? What is considered a successful telehealth episode of care? What criteria indicate other modalities of care be considered? Mapping clinic workflow is also essential for identifying gaps where patients may be unable or do not receive the necessary physical therapy services in a timely manner. Practice facilitation may be one strategy to allow physical therapists in a health care system to internalize approaches to ensure the right patient has access to the right modality of care at the right time [48,49]. Practice facilitation is an intervention where an external or internal facilitator interacts with multilevel stakeholders and can offer tools, resources, expertise, and guidance on strategies that address gaps and optimize workflow. Importantly, practice facilitation in the context of web-based physical therapy care can develop an internal capacity for change that can transcend the delivery of telehealth to be adaptive and receptive to evaluating and promoting equitable implementation.

Third, building *digital infrastructure* at multiple levels is necessary to ensure any patient who would benefit from physical therapy services has the option to participate in telehealth, if clinically appropriate and it aligns with patient preferences. To understand the baseline level of infrastructure, research is needed to measure (1) individual-level factors such as technology skills and proficiency, equipment availability, acceptability, and preferences for care and (2) community-level metrics such as neighborhood availability of Wi-Fi or broadband and transportation to in-person appointments. This information and ongoing evaluation inform policies and oversight of policy implementation. Building a digital infrastructure is interconnected with establishing community partnerships and communicating shared mental models as the infrastructure involves cross-sector collaborations for resources, governance, and continual monitoring. An opportunity exists to co-design or adapt aspects of digital infrastructure to better meet the needs of *all* patients who would benefit from physical therapy care, clinicians providing physical therapy, and health systems offering physical therapy services. To be truly successful, a digital infrastructure must develop a plan for transparency and sharing of data, engage the community throughout the infrastructure planning and implementation, and manage data privacy and security [50]. A strong digital infrastructure can provide the foundation to expand the equitable implementation of telehealth to other health and community services.

Future Directions and Recommendations Post Pandemic

Deployment of telehealth will likely persist in the post pandemic era, given patient preferences for such care, emerging advances in technologies, paradigm shifts in health care professional training, and the potential to serve populations with limited potential for high-quality in-person care (eg, residents of rural areas and patients who are homebound) [21,51]. Reevaluating and adapting telehealth to promote equitable implementation is one way to identify patient groups who may be harmed by the web-based delivery of services or those who may be negatively impacted by a full return to non-telehealth delivery. We recommend evaluating the ongoing and future implementation of telehealth by (1) evaluating hybrid care models, (2) identifying multilevel barriers and facilitators to adapting technology resources that enhance access and use across diverse populations, and (3) exploring the intersectionality of telehealth access and usage with respect to age, race, ethnicity, sexual orientation, disability (including visual and hearing impairments), socioeconomic status, social determinants of health, digital health literacy and numeracy, or residence in rural or urban settings. Bridging issues outlined in Table 1 can be alleviated through bridging solutions (Figure 1):

establishing community partnerships, communicating shared mental models, and building digital infrastructure.

First, establishing *community partnerships* (including those often excluded or marginalized) is the intentional and meaningful involvement of impacted community members to understand key issues and problem solving [12,52]. Feedback loops among implementation actors at multiple levels—patients, providers, clinical/health care system leadership, and policy makers—are needed to capture the barriers, facilitators, and unintended consequences to delivering or receiving telehealth, thus enabling a stronger understanding of who is impacted by telehealth delivery and how. At the organizational and policy level, assessing organizational readiness to support multiple modalities and options for care delivery is necessary to honor individual preferences for care while minimizing disruption to clinic workflow [53,54]. We acknowledge that considerable time, support, and relationship building with impacted community members is needed when conducting equitable implementation research. The time, effort, burden, and compatibility with workflow need to be periodically evaluated and modified as needed to make equitable implementation of telehealth sustainable. Future areas of evaluation highlight the need to describe how patient and multilevel partners and contextual factors can impact the uptake and adoption of telehealth through mediation or moderation. Specifically, future work is necessary to examine the uptake and adoption of telehealth by population and setting (high vs low resourced) to promote equitable use of telehealth in health care. Second, *shared mental models* between those impacted by telehealth adoption can determine the level of telehealth they are willing to accept and what changes—such as adaptations to implementation—they may consider for enhanced equity [34,35]. Adaptations are also an important aspect to maximizing equitable implementation by minimizing unintended consequences. For example, systems or clinics may need to integrate assessments of health/technology literacy/numeracy [55-57] into routine clinical care and then create or adopt interventions that address identified gaps. Additionally, for equitable implementation of telehealth, systems or clinics must identify the characteristics of their catchment area that go beyond individual factors (eg, space to complete movement assessments in the home, privacy for web-based connections with physical therapists) to include care delivery constraints (eg, beginning sessions late and running over time of scheduled appointment). Patient, clinical providers, health care systems, and communities need a shared mental model of such adaptations to collectively understand the role and impact of changes on telehealth access, receipt, and delivery. Holtrop et al [35] provide a useful table describing methods to elicit mental models. Understanding mental models can help to select potential implementation strategies needed to promote the equitable implementation of telehealth across a variety of disciplines [35,58,59].

Finally, as we approach postpandemic implementation of telehealth, building a *digital infrastructure* has the potential to mitigate long-standing issues with the inverse relationship between the need for health care and use or access [55] across different populations. Building a digital infrastructure requires attention to engagement, access, training (including cultural humility), and equity-focused measurement [25,60,61]. Future research needs to evaluate the cost-effectiveness of telehealth that includes policy analysis and program evaluation related to the construction and sustainment of digital infrastructure in communities.

Conclusions

Enhancing equitable implementation of telehealth is timely and critical to advancing the health and well-being of *all* persons. The tension between ongoing innovation in telehealth that is occurring in the context of the evolving pandemic creates opportunities for innovation *and* unanticipated challenges to equitable implementation. Equity frameworks help connect internal and external contexts that create disparities and to consider the implementation strategies that may address them. Bridging factors such as community partnerships, shared mental models, and digital infrastructure can guide implementation, adaptations, and sustainability in the setting of a rapidly changing landscape for telehealth.

Acknowledgments

The authors gratefully acknowledge Shelly Johnston, MPH, RD, and Kirsten K Davison, PhD, MS, for their support and feedback during the development and execution of this project. We also thank Emily Hudson for assistance in editing and formatting for the overall improvement of the manuscript. The Institute for Implementation Science Scholars (IS-2) is supported by funding from the National Institutes of Health (NIH) Office of Behavioral and Social Sciences Research (OBSSR) and Office of Disease Prevention, administered by the National Institute of Diabetes and Digestive and Kidney Disorders (R25DK123008, R25DK123008-S1) and the US Veterans Administration. AMG was partially supported during this work by the Agency for Healthcare Research and Quality (AHRQ) and Patient-Centered Outcomes Research Institute (PCORI; grant number K12HS026379); the National Institutes of Health's National Center for Advancing Translational Sciences (grant number KL2TR002492); and the Minneapolis Center of Innovation, Center for Care Delivery and Outcomes Research (CIN 13-406). EEFC is supported by the National Institutes of Health (K08 MH120341). SEL was supported by the National Institutes of Health (K08 HS26517). AAL is supported by the Durham Center of Innovation to Accelerate Discovery and Practice Transformation (grant number CIN 13-410) at the Durham Veterans Affairs Health Care System, and Veteran Affair's Health Services Research and Development Service (VA HSR&D, grant number 18-234). DMO was supported during this work by National Cancer Institute (K99CA256043) and the Robert E Leet and Clara Guthrie Patterson Trust Mentored Research Award. REG was supported by the IS-2 program and the National Cancer Institute (NCI)-funded University of Colorado Implementation Science Center for

Cancer Control (P50 CA244688). LMK was supported by the IS-2 Program (R25DK123008), and the NCI-funded Washington University Implementation Science Center for Cancer Control (P50 CA244431). The views expressed in this article are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States Government, AHRQ, PCORI, Duke University, or Minnesota Learning Health System Mentored Career Development Program (MN-LHS).

Authors' Contributions

All authors contributed to the conception and design of this work, drafted this work, substantively revised it, and approved the submitted version.

Conflicts of Interest

AAL reports receiving funding from Otsuka and the PhRMA Foundation not related to this work. EEFC serves on the clinical advisory board of Beanbag Health and receives consulting fees from Kooth. SEL began a position at Omada Health during the final preparation stage of this manuscript and this affiliation has no influence on the manuscript.

References

1. Lewinski AA, Walsh C, Rushton S, Soliman D, Carlson SM, Luedke MW, et al. Telehealth for the longitudinal management of chronic conditions: systematic review. J Med Internet Res 2022;24(8):e37100. [doi: 10.2196/37100] [Medline: 36018711]
2. What is telehealth? Health Resources and Services Administration (HRSA). 2022. URL: https://www.hrsa.gov/rural-health/topics/telehealth/what-is-telehealth [accessed 2023-04-28]
3. Nouri S, Khoong EC, Lyles CR, Karliner L. Addressing equity in telemedicine for chronic disease management during the Covid-19 pandemic. NEJM Catalyst. 2020. URL: https://catalyst.nejm.org/doi/full/10.1056/CAT.20.0123 [accessed 2023-04-28]
4. Health equity in telehealth. Health Resources & Service Administration (HRSA). 2022. URL: https://telehealth.hhs.gov/providers/health-equity-in-telehealth [accessed 2023-04-28]
5. Bauer MS, Damschroder L, Hagedorn H, Smith J, Kilbourne AM. An introduction to implementation science for the non-specialist. BMC Psychol 2015;3(1):32 [FREE Full text] [doi: 10.1186/s40359-015-0089-9] [Medline: 26376626]
6. Veinot TC, Mitchell H, Ancker JS. Good intentions are not enough: how informatics interventions can worsen inequality. J Am Med Inform Assoc 2018;25(8):1080-1088 [FREE Full text] [doi: 10.1093/jamia/ocy052] [Medline: 29788380]
7. Woodward EN, Matthieu MM, Uchendu US, Rogal S, Kirchner JE. The health equity implementation framework: proposal and preliminary study of hepatitis C virus treatment. Implement Sci 2019;14(1):26 [FREE Full text] [doi: 10.1186/s13012-019-0861-y] [Medline: 30866982]
8. Baumann AA, Cabassa LJ. Reframing implementation science to address inequities in healthcare delivery. BMC Health Serv Res 2020;20(1):190 [FREE Full text] [doi: 10.1186/s12913-020-4975-3] [Medline: 32164706]
9. Galaviz KI, Breland JY, Sanders M, Breathett K, Cerezo A, Gil O, et al. Implementation science to address health disparities during the coronavirus pandemic. Health Equity 2020;4(1):463-467 [FREE Full text] [doi: 10.1089/heq.2020.0044] [Medline: 33111032]
10. Brownson RC, Kumanyika SK, Kreuter MW, Haire-Joshu D. Implementation science should give higher priority to health equity. Implement Sci 2021;16(1):28 [FREE Full text] [doi: 10.1186/s13012-021-01097-0] [Medline: 33740999]
11. Woodward EN, Singh RS, Ndebele-Ngwenya P, Melgar Castillo A, Dickson KS, Kirchner JE. A more practical guide to incorporating health equity domains in implementation determinant frameworks. Implement Sci Commun 2021;2(1):61 [FREE Full text] [doi: 10.1186/s43058-021-00146-5] [Medline: 34090524]
12. Kerkhoff AD, Farrand E, Marquez C, Cattamanchi A, Handley MA. Addressing health disparities through implementation science-a need to integrate an equity lens from the outset. Implement Sci 2022;17(1):13 [FREE Full text] [doi: 10.1186/s13012-022-01189-5] [Medline: 35101088]
13. Proctor EK, Geng E. A new lane for science. Science 2021;374(6568):659 [FREE Full text] [doi: 10.1126/science.abn0184] [Medline: 34735221]
14. Srinivasan M, Asch S, Vilendrer S, Thomas SC, Bajra R, Barman L, et al. Qualitative assessment of rapid system transformation to primary care video visits at an academic medical center. Ann Intern Med 2020;173(7):527-535 [FREE Full text] [doi: 10.7326/M20-1814] [Medline: 32628536]
15. Eruchalu CN, Pichardo MS, Bharadwaj M, Rodriguez CB, Rodriguez JA, Bergmark RW, et al. The expanding digital divide: digital health access inequities during the COVID-19 pandemic in New York city. J Urban Health 2021;98(2):183-186 [FREE Full text] [doi: 10.1007/s11524-020-00508-9] [Medline: 33471281]
16. Litchfield I, Shukla D, Greenfield S. Impact of COVID-19 on the digital divide: a rapid review. BMJ Open 2021;11(10):e053440 [FREE Full text] [doi: 10.1136/bmjopen-2021-053440] [Medline: 34642200]
17. Mohammed HT, Hyseni L, Bui V, Gerritsen B, Fuller K, Sung J, et al. Exploring the use and challenges of implementing virtual visits during COVID-19 in primary care and lessons for sustained use. PLoS One 2021;16(6):e0253665 [FREE Full text] [doi: 10.1371/journal.pone.0253665] [Medline: 34166441]

18. Clare CA. Telehealth and the digital divide as a social determinant of health during the COVID-19 pandemic. Netw Model Anal Health Inform Bioinform 2021;10(1):26 [FREE Full text] [doi: 10.1007/s13721-021-00300-y] [Medline: 33842187]
19. Rhee K, Dankwa-Mullan I, Brennan V, Clark C. What is TechQuity? J Health Care Poor Underserved 2021;32(2S):xiii-xviii [FREE Full text] [doi: 10.1353/hpu.2021.0045]
20. Damschroder LJ, Reardon CM, Opra Widerquist MA, Lowery J. Conceptualizing outcomes for use with the consolidated framework for implementation research (CFIR): the CFIR outcomes addendum. Implement Sci 2022;17(1):7 [FREE Full text] [doi: 10.1186/s13012-021-01181-5] [Medline: 35065675]
21. Lyles CR, Wachter RM, Sarkar U. Focusing on digital health equity. JAMA 2021;326(18):1795-1796 [FREE Full text] [doi: 10.1001/jama.2021.18459] [Medline: 34677577]
22. Sieck CJ, Rastetter M, Hefner JL, Glover AR, Magaña C, Gray DM, et al. The Five A's of access for TechQuity. J Health Care Poor Underserved 2021;32(2):290-299 [FREE Full text] [doi: 10.1353/hpu.2021.0064]
23. Clark CR, Akdas Y, Wilkins CH, Rhee K, Johnson KB, Bates DW, et al. TechQuity is an imperative for health and technology business: let's work together to achieve it. J Am Med Inform Assoc 2021;28(9):2013-2016 [FREE Full text] [doi: 10.1093/jamia/ocab103] [Medline: 34157112]
24. Richardson S, Lawrence K, Schoenthaler AM, Mann D. A framework for digital health equity. NPJ Digit Med 2022;5(1):119 [FREE Full text] [doi: 10.1038/s41746-022-00663-0] [Medline: 35982146]
25. Crawford A, Serhal E. Digital health equity and COVID-19: the innovation curve cannot reinforce the social gradient of health. J Med Internet Res 2020;22(6):e19361 [FREE Full text] [doi: 10.2196/19361] [Medline: 32452816]
26. Moullin JC, Dickson KS, Stadnick NA, Rabin B, Aarons GA. Systematic review of the exploration, preparation, implementation, sustainment (EPIS) framework. Implement Sci 2019;14(1):1 [FREE Full text] [doi: 10.1186/s13012-018-0842-6] [Medline: 30611302]
27. Moullin JC, Dickson KS, Stadnick NA, Albers B, Nilsen P, Broder-Fingert S, et al. Ten recommendations for using implementation frameworks in research and practice. Implement Sci Commun 2020;1:42 [FREE Full text] [doi: 10.1186/s43058-020-00023-7] [Medline: 32885199]
28. Lengnick-Hall R, Stadnick NA, Dickson KS, Moullin JC, Aarons GA. Forms and functions of bridging factors: specifying the dynamic links between outer and inner contexts during implementation and sustainment. Implement Sci 2021;16(1):34 [FREE Full text] [doi: 10.1186/s13012-021-01099-y] [Medline: 33794956]
29. Dover DC, Belon AP. The health equity measurement framework: a comprehensive model to measure social inequities in health. Int J Equity Health 2019;18(1):36 [FREE Full text] [doi: 10.1186/s12939-019-0935-0] [Medline: 30782161]
30. Aarons GA, Hurlburt M, Horwitz SM. Advancing a conceptual model of evidence-based practice implementation in public service sectors. Adm Policy Ment Health 2011;38(1):4-23 [FREE Full text] [doi: 10.1007/s10488-010-0327-7] [Medline: 21197565]
31. Potapchuk M, Leiderman S, Bivens D, Major B. Flipping the script: white privilege and community building. MP Associates, Inc, and the Center for Assessment and Policy Development (CAPD). 2005. URL: https://www.mpassociates.us/uploads/3/7/1/0/37103967/flippingthescriptmostupdated.pdf [accessed 2023-04-28]
32. Chinook F. The four "I's" of oppression. Association for University and College Counseling Center Outreach. 2010. URL: http://auccco.com/resources/Documents/4isofoppression.pdf [accessed 2023-04-28]
33. Madden P, Vera Cruz AC. Integrating Eisner's concepualization of curriculum with the four "I's" of systematic racism to promote anti-racist praxis in education and beyond. J Am Assoc Adv Curric Stud 2021;14(2):1-19 [FREE Full text]
34. Gentner D, Stevens AL. Mental Models. New York: Psychology Press; 1983.
35. Holtrop JS, Scherer LD, Matlock DD, Glasgow RE, Green LA. The importance of mental models in implementation science. Front Public Health 2021;9:680316 [FREE Full text] [doi: 10.3389/fpubh.2021.680316] [Medline: 34295871]
36. Hall JB, Luechtefeld JT, Woods ML. Adoption of telehealth by pediatric physical therapists during COVID-19: a survey study. Pediatr Phys Ther 2021;33(4):237-244 [FREE Full text] [doi: 10.1097/PEP.0000000000000817] [Medline: 34323864]
37. Miller MJ, Pak SS, Keller DR, Barnes DE. Evaluation of pragmatic telehealth physical therapy implementation during the COVID-19 pandemic. Phys Ther 2021;101(1):pzaa193 [FREE Full text] [doi: 10.1093/ptj/pzaa193] [Medline: 33284318]
38. Grundstein MJ, Fisher C, Titmuss M, Cioppa-Mosca J. The role of virtual physical therapy in a post-pandemic world: pearls, pitfalls, challenges, and adaptations. Phys Ther 2021;101(9):pzab145 [FREE Full text] [doi: 10.1093/ptj/pzab145] [Medline: 34106273]
39. Lieneck C, Weaver E, Maryon T. Outpatient telehealth implementation in the United States during the COVID-19 global pandemic: a systematic review. Medicina (Kaunas) 2021;57(5):462 [FREE Full text] [doi: 10.3390/medicina57050462] [Medline: 34065050]
40. LeFevre ML. Behavioral counseling to promote a healthful diet and physical activity for cardiovascular disease prevention in adults with cardiovascular risk factors: U.S. preventive services task force recommendation statement. Ann Intern Med 2014;161(8):587-593 [FREE Full text] [doi: 10.7326/M14-1796] [Medline: 25155419]
41. McCallum CA, DiAngelis T. Direct access: factors that affect physical therapist practice in the state of Ohio. Phys Ther 2012;92(5):688-706 [FREE Full text] [doi: 10.2522/ptj.20100358] [Medline: 22247405]

42. Laliberté M, Williams-Jones B, Feldman DE, Hunt M. Ethical challenges for patient access to physical therapy: views of staff members from three publicly-funded outpatient physical therapy departments. Narrat Inq Bioeth 2017;7(2):157-169 [FREE Full text] [doi: 10.1353/nib.2017.0046] [Medline: 29056647]
43. Sandstrom R, Bruns A. Disparities in access to outpatient rehabilitation therapy for African Americans with arthritis. J Racial Ethn Health Disparities 2017;4(4):599-606 [FREE Full text] [doi: 10.1007/s40615-016-0263-7] [Medline: 27400913]
44. Curry EJ, Penvose IR, Knapp B, Parisien RL, Li X. National disparities in access to physical therapy after rotator cuff repair between patients with medicaid vs. private health insurance. JSES Int 2021;5(3):507-511 [FREE Full text] [doi: 10.1016/j.jseint.2020.11.006] [Medline: 34136862]
45. Lee AC. COVID-19 and the advancement of digital physical therapist practice and telehealth. Phys Ther 2020;100(7):1054-1057 [FREE Full text] [doi: 10.1093/ptj/pzaa079] [Medline: 32343836]
46. Kim SY, Kumble S, Patel B, Pruski AD, Azola A, Tatini AL, et al. Managing the rehabilitation wave: rehabilitation services for COVID-19 survivors. Arch Phys Med Rehabil 2020;101(12):2243-2249 [FREE Full text] [doi: 10.1016/j.apmr.2020.09.372] [Medline: 32971100]
47. Albu S, Zozaya NR, Murillo N, García-Molina A, Chacón CAF, Kumru H. What's going on following acute COVID-19? Clinical characteristics of patients in an out-patient rehabilitation program. NeuroRehabilitation 2021;48(4):469-480 [FREE Full text] [doi: 10.3233/NRE-210025] [Medline: 33998551]
48. Ritchie MJ, Dollar KM, Miller CJ, Oliver KA, Smith JL, Lindsay JA, et al. Implementation facilitation training manual: using implementation faciliation to improve care in the Veterans Health Administration, version 2. Veterans Health Administration. 2017. URL: http /Users/A9075/Downloads/implementation-facilitation-training-manual.pdf [accessed 2023-04-28]
49. Frost MC, Ioannou GN, Tsui JI, Edelman EJ, Weiner BJ, Fletcher OV, et al. Practice facilitation to implement alcohol-related care in veterans health administration liver clinics: a study protocol. Implement Sci Commun 2020;1(1):68 [FREE Full text] [doi: 10.1186/s43058-020-00062-0] [Medline: 32835226]
50. Goldsmith S, Gardner B. Implementing digital infrastructure: responses to equity, sustainability, and safety. Data-Smart City Solutions. 2022. URL: https://datasmart.hks.harvard.edu/implementing-digital-infrastructure [accessed 2023-04-28]
51. Patel SY, Huskamp HA, Busch AB, Mehrotra A. Telemental health and US rural-urban differences in specialty mental health use, 2010-2017. Am J Public Health 2020;110(9):1308-1314 [FREE Full text] [doi: 10.2105/AJPH.2020.305657] [Medline: 32673109]
52. Fields J, Gutierrez JR, Marquez C, Rhoads K, Kushel M, Fernández A, et al. Community-academic partnerships to address covid-19 inequities: lessons from the San Francisco bay area. NEJM Catal Innov Care Deliv 2021;2(3):1-20 [FREE Full text] [doi: 10.1056/CAT.21.0135]
53. Weiner BJ. A theory of organizational readiness for change. Implement Sci 2009;4:67 [FREE Full text] [doi: 10.1186/1748-5908-4-67] [Medline: 19840381]
54. Kononowech J, Hagedorn H, Hall C, Helfrich CD, Lambert-Kerzner AC, Miller SC, et al. Mapping the organizational readiness to change assessment to the consolidated framework for implementation research. Implement Sci Commun 2021;2(1):19 [FREE Full text] [doi: 10.1186/s43058-021-00121-0] [Medline: 33581728]
55. Hart JT. The inverse care law. Lancet 1971;297(7696):405-412 [FREE Full text] [doi: 10.1016/s0140-6736(71)92410-x] [Medline: 4100731]
56. Karnoe A, Furstrand D, Christensen KB, Norgaard O, Kayser L. Assessing competencies needed to engage with digital health services: development of the ehealth literacy assessment toolkit. J Med Internet Res 2018;20(5):e178 [FREE Full text] [doi: 10.2196/jmir.8347] [Medline: 29748163]
57. Oh SS, Kim KA, Kim M, Oh J, Chu SH, Choi J. Measurement of digital literacy among older adults: systematic review. J Med Internet Res 2021;23(2):e26145 [FREE Full text] [doi: 10.2196/26145] [Medline: 33533727]
58. Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, Matthieu MM, et al. A refined compilation of implementation strategies: results from the expert recommendations for implementing change (ERIC) project. Implement Sci 2015;10:21 [FREE Full text] [doi: 10.1186/s13012-015-0209-1] [Medline: 25889199]
59. Waltz TJ, Powell BJ, Fernández ME, Abadie B, Damschroder LJ. Choosing implementation strategies to address contextual barriers: diversity in recommendations and future directions. Implement Sci 2019;14(1):42 [FREE Full text] [doi: 10.1186/s13012-019-0892-4] [Medline: 31036028]
60. Sivashanker K, Duong T, Resnick A, Eappen S. Health care equity: from fragmentation to transformation. NEJM Catal Innov Care Deliv 2020;1(5):1-8 [FREE Full text] [doi: 10.1056/CAT.20.0414]
61. Harrington R, Washington D, Paliani S, Thompson K, Rouse L, Anderson AC. A new effort to address racial and ethnic disparities in care through quality measurement. Health Affairs Forefront. 2021. URL: https://www.healthaffairs.org/do/10.1377/forefront.20210907.568444/full/ [accessed 2023-04-28]

Abbreviations

DHEF: Digital Health Equity Framework
EPIS: Exploration, Preparation, Implementation, and Sustainment

Edited by A Mavragani; submitted 16.06.22; peer-reviewed by J Satterfield, L Nelson, A Schoenthaler; comments to author 08.09.22; revised version received 08.12.22; accepted 30.03.23; published 15.05.23.

Please cite as:
Gustavson AM, Lewinski AA, Fitzsimmons-Craft EE, Coronado GD, Linke SE, O'Malley DM, Adams AS, Glasgow RE, Klesges LM
Strategies to Bridge Equitable Implementation of Telehealth
Interact J Med Res 2023;12:e40358
URL: https://www.i-jmr.org/2023/1/e40358
doi:10.2196/40358
PMID:37184909

Viewpoint

The Utility of Predictive Modeling and a Systems Process Approach to Reduce Emergency Department Crowding: A Position Paper

Ann Corneille Monahan[1], MSHI, PhD; Sue S Feldman[2], MEd, RN, PhD

[1]University of Alabama at Birmingham, Birmingham, AL, United States

[2]Department of Health Services Administration, University of Alabama at Birmingham, Birmingham, AL, United States

Corresponding Author:
Ann Corneille Monahan, MSHI, PhD
University of Alabama at Birmingham
1720 University Blvd
Birmingham, AL, 35294
United States
Phone: 1 205 934 4011
Email: monahanannc@gmail.com

Abstract

Emergency department (ED) crowding and its main causes, exit block and boarding, continue to threaten the quality and safety of ED care. Most interventions to reduce crowding have not been comprehensive or system solutions, only focusing on part of the care procession and not directly affecting boarding reduction. This position paper proposes that the ED crowding problem can be optimally addressed by applying a systems approach using predictive modeling to identify patients at risk of being admitted to the hospital and uses that information to initiate the time-consuming bed management process earlier in the care continuum, shortening the time during which patients wait in the ED for an inpatient bed assignment, thus removing the exit block that causes boarding and subsequently reducing crowding.

(Interact J Med Res 2023;12:e42016) doi:10.2196/42016

KEYWORDS

emergency care, prehospital; information systems; crowding; healthcare service; healthcare system; emergency department; boarding; exit block; medical informatics, application; health services research; personalized medicine; predictive medicine; model, probabilistic; polynomial model; decision support technique; systems approach; predict; evidence based health care; hospital bed management; management information systems; position paper

Introduction

The Emergency Department Crowding Problem

The ED crowding problem occurs when the ED demand exceeds the staff's ability to provide quality care in a reasonable period of time [1,2]. The literature suggests that hospital exit block [3,4] (ie, when patients cannot transition into the hospital from the ED because a hospital bed has not been assigned [3]) and ED boarding [5-8] (ie, when a patient due to be admitted to the hospital remains in the ED, occupying a bed [3]) are the main causes of ED crowding and posits that an impactful solution lies in changes in the bed management strategy, the processes involved in the transition of patients from the ED to the hospital, and when securing a hospital bed [9,10].

This position paper proposes that the complexities of the ED crowding problem can be optimally addressed by applying a systems approach to the hospital bed management strategy. The systems approach views an environment as a whole, which is made up of many parts or subsystems for the purpose of understanding the relationships between the system and its parts and to aid in problem-solving [11]. A systems approach that uses predictive modeling to identify patients who are at risk of being admitted to the hospital and uses that information to initiate the complex and time-consuming bed management process earlier in the care continuum could potentially shorten the time during which patients wait in the ED to be transitioned into an inpatient bed, thus removing the exit block that causes boarding and subsequently reducing crowding [12]. The challenges facing the health care industry are greater than ever before, with increasing complexities of care, regulations, and higher quality care expectations. At the same time, the industry is challenged to address preventable medical errors, poor amenable mortality rates, nursing and physician burnout and shortages, and general inefficiencies. These industry challenges are magnified in the ED where the nature and environment of emergency care cannot tolerate threats to quality care delivery or to patient and provider safety. A systems approach views the ED as a complex microcosm of a larger health ecosystem where optimal functionality requires that it be resilient to the unpredictable demands [13] characteristic of the urgent care environment. To manage new and unpredictable challenges, a systems approach can be used to address known threats. An efficient manner by which to accomplish this is to identify predictable and repeatable processes to which information technology can be applied.

Crowding and its main causes, exit block and boarding, have threatened the quality and safety of ED care for over 20 years despite the efforts of many to resolve it [14-16]. Boarding compromises care quality [3,17], stresses hospital operations [18], and strains resources because boarded patients occupy beds and divert staff resources from new and existing patients [17] and reduce revenue generation through the reduction in available beds for treating new patients [19].

The industry would benefit from improved approaches to resolve the crowding problem. The existing research focused on this problem suggests that predictive modeling holds promise to make a significant contribution toward addressing ED crowding; for example, models have predicted imminent hospital admissions for older ED patients [20], identified patients who are likely to require hospital care in the future [21], predicted ED crowding using calendar and weather variables [22], and forecasted ED flow digitally [23].

Applying Predictive Modeling to Resolve Crowding and Bed Management

Predictive modeling is a form of data mining technology that functions by analyzing historic and current data, and generating a model to help predict a future outcome [24]. It is an explicit, empirical approach for estimating the probabilities that an event will or will not occur in the future [25]—such as death, contracting a disease, surgical complications, or hospital admission—by using statistical techniques to predict future events. Models use data about patients, diseases, or treatment characteristics to estimate the probability that a condition or disease is present or the probability that an outcome will occur [26,27]. However, models are only just starting to be used to produce actionable information to impact operations and patient care. We posit that predictive models to identify patients who are at risk of being admitted could be applied in the bed manager environment to remove the exit block that causes boarding by initiating the bed management process earlier in the care continuum, thereby shortening the time during which patients wait in the ED for an inpatient bed assignment. This, in turn, reduces exit block, boarding, and subsequently crowding. As no clinical decisions are made on the basis of patients' risk of admission, this process could be automated to streamline part of the complicated bed management process and take advantage of predictable and repeatable processes using standardized data.

Many interventions to reduce crowding have not been comprehensive or system solutions but rather focus on part of the care procession and do not directly affect boarding reduction [28]. However, existing interventions that addressed crowding as a systemic problem have reduced the time during which a patient is boarded in the ED [9,10,29], which, in turn, reduces the backlog of boarded patients who contribute to ED crowding. Interdepartmental collaboration with hospital management support was a feature in these interventions. Two of these interventions also used real-time ED data on congestion, flow, and patient admissions to prepare for and manage inpatient admissions and bed demand [9,29]. Individual interventions are parts of the system, rather than being considered a collective, and are automated to contribute valuable data to augment bed management.

The use of predictive models in health care have quadrupled over the last 2 decades and their accuracy has increased [12,28]. While these have traditionally been applied to identify risk, the time is right for integrating predictive models with existing technologies such as electronic health records, clinical decision support systems, and clinical data warehouses, to result in action and efficiencies. n 2019, the use of predictive modeling was reported among 60% of health care executives within their organizations, and another 20% of them expressed intent to begin using them the following year [30], again primarily for risk prediction and not necessarily action. This existing technological infrastructure holds promise to reduce prior barriers to integration. Moreover, as clinical care becomes increasingly personalized, aid from predictive analytics may become a best-practice procedure.

Concerns for Risk Models

Developing quality predictive models is challenging [28,31]. Deciding what variables to measure to answer a particular question can even be problematic. For example, if the goal is to predict "health," which data are measured as indicators of health? The answer to this question varies. If rigorous study methods are not used throughout study development, data gathering, and analysis, there are numerous avenues for the model to make errors and lead to unintentional bias. Recent studies, including a systematic review of admission prediction models [28], have questioned the quality and rigor of existing predictive models [12,28,31-33] due to an overall lack of external validation studies [31], multiple prediction models for the same outcome or target population [26], and risks of bias [34-36]. A bias unique to predictive models, algorithmic bias, occurs when technology reflects the attitudes and values of the humans who coded, collected, selected, or used the data to train the algorithm [37]. Thus, machine-generated algorithms are human products executed by a machine. Algorithms should not be blindly trusted or considered neutral and unbiased [37]. Reliance on an algorithm to predict health-related outcomes or to make decisions about care would increase the pace of decision-making, but the point at which the decision should be transferred from machine to human is necessary, unclear, and currently unregulated [37].

Conclusions

Large amounts of data are available for analysis, and the demands on the health care industry are increasing, making the use of predictive modeling to aid hospital operations sensible and increasingly necessary. The old adage "garbage in garbage out" remains true when applied to predictive models—models developed without quality methodologies risk producing predictions deficient of quality. A model that produces biased predictions may not resolve the problem at hand. Evidence that a model is effective and safe is necessary before its use in a clinical setting. Best practices promoting standards for development and operation will have a role to play in model improvement and their use in clinical settings.

Application of models that predict hospital admission could aid hospital bed managers to secure an appropriate bed for a patient in a timely manner while boosting hospital efficiency and with no harm to patients. The result of this timely and streamlined systems process is better patient care delivered sooner.

We posit that applying a systems approach using prediction models to the hospital bed management strategy for ED patients would reveal the many parts and subsystems involved before and after bed assignment and would ensure that they are part of the solution. This unique application of a prediction model provides bed managers information to support initiation of bed management processes earlier in the care continuum. This strategic use of information has significant potential to reduce hospital exit block and ED boarding, and subsequently ED crowding [12].

Conflicts of Interest

None declared.

References

1. Sinclair D. Emergency department overcrowding - implications for paediatric emergency medicine. Paediatr Child Health 2007 Jul;12(6):491-494 [FREE Full text] [doi: 10.1093/pch/12.6.491] [Medline: 19030415]
2. American College of Emergency Physicians. Crowding. Ann Emerg Med 2006 Jun;47(6):585. [doi: 10.1016/j.annemergmed.2006.02.025] [Medline: 16713796]
3. Richards JR, van der Linden MC, Derlet RW. Providing care in emergency department hallways: demands, dangers, and deaths. Adv J Emerg Med 2014 [FREE Full text] [doi: 10.1155/2014/495219]
4. Henderson K, Boyle A. Exit block in the emergency department: recognition and consequences. Br J Hosp Med (Lond) 2014 Nov 02;75(11):623-626 [FREE Full text] [doi: 10.12968/hmed.2014.75.11.623] [Medline: 25383431]
5. Young C. Hospital Emergency Departments: Crowding Continues to Occur, and Some Patients Wait Longer than Recommended Time Frames. U.S. Government Accountability Office. 2009. URL: https://www.gao.gov/products/gao-09-347 [accessed 2023-06-16]
6. Institute of Medicine. Hospital-Based Emergency Care: At the Breaking Point. Washington, DC: The National Academies Press; 2007.
7. Higginson I. Emergency department crowding. Emerg Med J 2012 Jun 04;29(6):437-443. [doi: 10.1136/emermed-2011-200532] [Medline: 22223713]
8. Mason S, Knowles E, Boyle A. Exit block in emergency departments: a rapid evidence review. Emerg Med J 2017 Jan 27;34(1):46-51 [FREE Full text] [doi: 10.1136/emermed-2015-205201] [Medline: 27789568]
9. Barrett L, Ford S, Ward-Smith P. A bed management strategy for overcrowding in the emergency department. Nurs Econ 2012;30(2):82-5, 116. [Medline: 22558725]
10. Amarasingham R, Swanson TS, Treichler DB, Amarasingham SN, Reed WG. A rapid admission protocol to reduce emergency department boarding times. Qual Saf Health Care 2010 Jun;19(3):200-204. [doi: 10.1136/qshc.2008.031641] [Medline: 20142408]
11. Process approach. Professional Evaluation and Certification Board. 2016 Mar 21. URL: https://pecb.com/article/process-approach [accessed 2022-03-21]
12. Monahan AC, Feldman SS, Fitzgerald TP. Reducing crowding in emergency departments with early prediction of hospital admission of adult patients using biomarkers collected at triage: retrospective cohort study. JMIR Bioinform Biotech 2022 Sep 13;3(1):e38845. [doi: 10.2196/38845]
13. Austin E, Blakely B, Salmon P, Braithwaite J, Clay-Williams R. Technology in the emergency department: using cognitive work analysis to model and design sustainable systems. Safety Science 2022 Mar;147:105613 [FREE Full text] [doi: 10.1016/j.ssci.2021.105613]
14. Stead LG, Jain A, Decker WW. Emergency department over-crowding: a global perspective. Int J Emerg Med 2009 Sep 30;2(3):133-134 [FREE Full text] [doi: 10.1007/s12245-009-0131-x] [Medline: 20157461]
15. Kauppila T, Seppänen K, Mattila J, Kaartinen J. The effect on the patient flow in a local health care after implementing reverse triage in a primary care emergency department: a longitudinal follow-up study. Scand J Prim Health Care 2017 Jun 08;35(2):214-220 [FREE Full text] [doi: 10.1080/02813432.2017.1333320] [Medline: 28593802]
16. Hospital Compare - User Guide. URL: https://health.mo.gov/data/hai/pdf/hospital-compare-user-guide.pdf [accessed 2023-06-16]
17. Improving Patient Flow and Reducing Emergency Department Crowding: A Guide for Hospitals. Agency for Healthcare Research and Quality. URL: https://www.ahrq.gov/research/findings/final-reports/ptflow/index.html [accessed 2023-06-16]

18. Institute of Medicine (US) Committee on Quality of Health Care in America. In: Kohn LT, Corrigan JM, Donaldson MS, editors. To Err is Human: Building a Safer Health System. Washington, DC: National Academies Press; 2000.
19. Falvo T, Grove L, Stachura R, Vega D, Stike R, Schlenker M, et al. The opportunity loss of boarding admitted patients in the emergency department. Acad Emerg Med 2007 Apr;14(4):332-337 [FREE Full text] [doi: 10.1197/j.aem.2006.11.011] [Medline: 17331916]
20. Brink A, Alsma J, Brink HS, de Gelder J, Lucke JA, Mooijaart SP, et al. Prediction admission in the older population in the Emergency Department: the CLEARED tool. Neth J Med 2020 Dec;78(6):357-367 [FREE Full text] [Medline: 33380533]
21. Marcusson J, Nord M, Dong H, Lyth J. Clinically useful prediction of hospital admissions in an older population. BMC Geriatr 2020 Mar 06;20(1):95 [FREE Full text] [doi: 10.1186/s12877-020-1475-6] [Medline: 32143637]
22. Erkamp NS, van Dalen DH, de Vries E. Predicting emergency department visits in a large teaching hospital. Int J Emerg Med 2021 Jun 12;14(1):34 [FREE Full text] [doi: 10.1186/s12245-021-00357-6] [Medline: 34118866]
23. Hoot NR, LeBlanc LJ, Jones I, Levin SR, Zhou C, Gadd CS, et al. Forecasting emergency department crowding: a prospective, real-time evaluation. J Am Med Inform Assoc 2009 May 01;16(3):338-345. [doi: 10.1197/jamia.m2772]
24. Predictive Modeling. Gartner. URL: https://www.gartner.com/en/information-technology/glossary/predictive-modeling [accessed 2022-04-02]
25. Steyerberg EW. Clinical Prediction Models A Practical Approach to Development, Validation, and Updating. New York, NY: Springer International Publishing; 2019.
26. Moons KG, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. Ann Intern Med 2019 Jan 01;170(1):W1. [doi: 10.7326/m18-1377]
27. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 2010 Jan;21(1):128-138 [FREE Full text] [doi: 10.1097/EDE.0b013e3181c30fb2] [Medline: 20010215]
28. Monahan AC, Feldman SS. Models predicting hospital admission of adult patients utilizing prehospital data: systematic review using PROBAST and CHARMS. JMIR Med Inform 2021 Sep 16;9(9):e30022 [FREE Full text] [doi: 10.2196/30022] [Medline: 34528893]
29. Howell E, Bessman E, Kravet S, Kolodner K, Marshall R, Wright S. Active bed management by hospitalists and emergency department throughput. Ann Intern Med 2008 Dec 02;149(11):804-811. [doi: 10.7326/0003-4819-149-11-200812020-00006] [Medline: 19047027]
30. 2019 Predictive Analytics in Health Care Trend Forecast. Society for Actuaries. URL: https://www.soa.org/globalassets/assets/Files/programs/predictive-analytics/2019-health-care-trend.pdf [accessed 2023-06-16]
31. Steyerberg EW, Harrell FE. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol 2016 Jan;69:245-247 [FREE Full text] [doi: 10.1016/j.jclinepi.2015.04.005] [Medline: 25981519]
32. McIntosh GS, Steenstra I, Hogg-Johnson S, Carter T, Hall H. Lack of prognostic model validation in low back pain prediction studies: a systematic review. Clin J Pain 2018 Aug;34(8):748-754. [doi: 10.1097/AJP.0000000000000591] [Medline: 29406366]
33. Shamsoddin E. Can medical practitioners rely on prediction models for COVID-19? A systematic review. Evid Based Dent 2020 Sep 25;21(3):84-86 [FREE Full text] [doi: 10.1038/s41432-020-0115-5] [Medline: 32978532]
34. Miles J, Turner J, Jacques R, Williams J, Mason S. Using machine-learning risk prediction models to triage the acuity of undifferentiated patients entering the emergency care system: a systematic review. Diagn Progn Res 2020 Oct 02;4(1):16 [FREE Full text] [doi: 10.1186/s41512-020-00084-1] [Medline: 33024830]
35. Kareemi H, Vaillancourt C, Rosenberg H, Fournier K, Yadav K. Machine learning versus usual care for diagnostic and prognostic prediction in the emergency department: a systematic review. Acad Emerg Med 2021 Feb 02;28(2):184-196 [FREE Full text] [doi: 10.1111/acem.14190] [Medline: 33277724]
36. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ 2020 Apr 07;369:m1328 [FREE Full text] [doi: 10.1136/bmj.m1328] [Medline: 32265220]
37. Watson K. Predictive analytics in health caremerging value and risks. Deloitte Insights 2021 March 12, 2022. 2019. URL: https://www2.deloitte.com/us/en/insights/topics/analytics/predictive-analytics-health-care-value-risks.html [accessed 2022-03-12]

Abbreviations

ED: emergency department

Edited by T Leung; submitted 29.08.22; peer-reviewed by J Jewer, S Nagavally, G Kolostoumpis; comments to author 25.11.22; revised version received 12.01.23; accepted 10.05.23; published 10.07.23.

Please cite as:
Monahan AC, Feldman SS
The Utility of Predictive Modeling and a Systems Process Approach to Reduce Emergency Department Crowding: A Position Paper
Interact J Med Res 2023;12:e42016
URL: https://www.i-jmr.org/2023/1/e42016
doi:10.2196/42016
PMID:37428536

Viewpoint

Information and Communication Technology Medicine: Integrative Specialty for the Future of Medicine

Peter Joseph Jongen[1,2], MD, PhD

[1]MS4 Research Institute, Nijmegen, Netherlands

[2]Department of Community and Occupational Medicine, University Medical Center Groningen, University of Groningen, Groningen, Netherlands

Corresponding Author:
Peter Joseph Jongen, MD, PhD
MS4 Research Institute
Ubbergseweg 34
Nijmegen, 6522 KJ
Netherlands
Phone: 31 24 3239146
Email: ms4ri@kpnmail.nl

Abstract

The impact of information and communication technology (ICT) on medicine is unprecedented and ever-increasing. This has made it more and more difficult for doctors to keep pace with ICT developments and to adequately match the input of ICT experts. As a result, medical disciplines may not be able to take full advantage of growing possibilities. In this personal viewpoint paper, I argue for the establishment of a novel medical specialty, ICT medicine. ICT medicine is needed to optimally face the challenges of ICT-based developments, including artificial intelligence (AI), and to ensure their efficient and beneficial use. ICT medicine is rooted in both medicine and ICT, and in contrast to existing medical specialties it is integrative in nature, as long-standing structural collaborations with ICT and other stakeholders cross the boundaries between disciplines. Thus, new concepts and theories may evolve that are better suited to addressing ICT-related issues in medicine. ICT doctors will be instrumental in the conception, development, implementation, and evaluation of digital tools, systems, and services. They provide a bridge between ICT professionals and clinical users and educate doctors in digital applications and services. Notably, ICT doctors may have a pivotal role in the validation, verification, and evaluation of AI models. ICT medicine institutes offer a home to these new professionals, enhancing their independence within health care organizations and in relation to ICT companies. Importantly, in an era of growing technicalization and use of AI algorithms, ICT doctors may safeguard the human factor in medicine. And, from a societal perspective, they may promote digital inclusion and the continuing high quality of digital services and provide leadership in the future digitalization of medicine.

(Interact J Med Res 2023;12:e42831) doi:10.2196/42831

KEYWORDS

information and communication technology; ICT; integrative; transdisciplinary; eHealth; internet; medical informatics; application; artificial intelligence; digital medicine; technologies

Introduction

Information is crucial to the success of medicine. In recent decades, the transmission and exchange of information have been revolutionized by the advent of the internet and ongoing developments in information and communication technology (ICT). At this point, the ICT sector provides a broad range of applications and services for medical use [1]. ICT-based fields, including medical informatics, have emerged in the slipstream of this evolution. Increasingly, collaborations between medicine and the ICT specialties are leading to new diagnostic, prognostic, and therapeutic tools and procedures. Among other consequences, these collaborations should ensure that ICT solutions are designed with all current empirical evidence in mind [2].

However, physicians' opportunities and capabilities for interacting optimally with ICT specialists tend to lag behind technical developments. For example, a recent review of smart home technologies for health care identified a lack of collaboration across disciplines and noted that technological developments dominate over the human-centric part of the equation [3]. And the implementation of continuous connected care augmented by remote monitoring, which is technically feasible, was found not to be matched by a corresponding shift in clinicians' approach to the delivery of care [4]. In the Netherlands, general practitioners were early adopters of electronic medical records (EMRs), and they saw the EMR as an instrument for providing evidence-based primary care [5]. However, in terms of ICT-medicine cooperation, the EMR is a rather straightforward use of ICT. At this point, more complex issues are at hand, such as the development and validation of digital biomarkers, the interpretation of automated measurement outcomes and their integrated use in clinical practice, the advancement of connected care models, and the development and evaluation of artificial intelligence (AI)–based algorithms. These current trends place substantially higher demands on the collaboration between ICT and medical experts; therefore, initiatives have a higher risk of not succeeding. Thus, the effectiveness and efficiency of ICT-based solutions may be less than optimal, potentially leading to unexpected challenges and extra expenditures.

Moreover, actual collaborations between medical and ICT professionals are often project-based and ad hoc, whereas implicit differences in perspectives may lead to miscommunication, disagreements, or even the failure of projects. Notably, ICT input is frequently provided by commercial companies whose motives and expectations are potentially at odds with those of medical professionals.

Against the background of both the high potential and the possible pitfalls of medicine-ICT collaborations, appropriate structures are needed to foster optimum cooperation between medical and ICT specialists. As the extensive and radical impact that ICT is having on medical practice and research is unprecedented in history, existing structures cannot be expected to optimally turn possibilities into reality.

In this personal viewpoint paper, I argue for the need for a novel field of medicine: ICT medicine. ICT medicine is needed to optimally face the challenges of new ICT developments, including AI, and to ensure their efficient and beneficial use. It does so by realizing long-standing collaborations between different stakeholders. I argue that ICT medicine—in contrast to existing disciplines—should be integrative in its nature, and I outline the broad range of its activities and the advantages of an institutional organization. Also, in an era of increasing technicalization, ICT doctors may safeguard the human factor in medicine. And, from a societal perspective, they may contribute to digital inclusion and the continuing high quality of digital services and provide leadership in the future digitalization of medicine as a whole.

Integrative Approach

In recent years, the field of digital medicine has emerged. Digital medicine is concerned with the use of high-quality hardware and software technologies as evidence-based tools for measurement and intervention in the service of human health [6]; given the wide range of stakeholders, its activities are dispersed over a great many disciplines [6,7]. Digital medicine is practiced by the same clinicians and health professionals who practice traditional medicine [7], and, in fact, digital medicine is on its way to becoming just plain medicine [7,8].

Importantly, digital medicine represents multidisciplinary or interdisciplinary collaborations [9]. In the multidisciplinary approach, perspectives, notions, and methods are used that prevail in the respective disciplines, and the end result is essentially a combination of the outcomes of the contributing disciplines. In interdisciplinary cooperation, experts from various fields are enabled to exchange ideas and insights, as a result of which the initial research question may be reframed and new questions may emerge. Yet, as in multidisciplinary research, the questions are formulated in discipline-specific wordings and the problems underlying the questions are perceived within conventional conceptual frameworks.

In contrast to multidisciplinary and interdisciplinary approaches, integrative or transdisciplinary collaborations aim toward insights that emerge by crossing the boundaries between disciplines; eventually these collaborations lead to concepts that may be better suited to addressing the problems at hand [10-14]. Integrative approaches transcend traditional boundaries to integrate various sciences [13-15]. Typically, new hypotheses and theories thrive in the context of integrative collaborations. Notably, "integrationality" may be seen as a mental and intellectual disposition, a habit of mind and behavior toward intentional connection seeking and connection making [13]. Thus, the integrative approach is preferably practiced in a framework of longstanding and structural collaborations with frequent and intensive interactions between researchers from various disciplines. Structural collaborations occur in a coherent organization, such as an institute, where the parts are dominated by the integrative character of the whole [16].

ICT Medicine

New Specialty

It has been acknowledged that success in digital medicine requires a fully integrated approach [7]. The existing cooperation between medical and ICT specialists is multidisciplinary and interdisciplinary, often temporary, and focused on selected topics. In order to effectively advance the development and implementation of digital tools and processes it is desirable to facilitate integrative interactions that are enduring and cover all conceivable aspects of ICT. This is best achieved through explicitly and formally integrating the various stakeholders' input and performance in a new organizational structure. Conceivably, integrative collaborations between medical professionals and ICT experts may materialize as a new specialty, ICT medicine. ICT medicine is both overarching and an integral part of the various existing disciplines. It is rooted in the science and practices of medicine and ICT but goes beyond these fields as its activities surpass barriers between disciplines.

Activities

ICT doctors cooperate with a wide range of stakeholders, such as physicians, scientists, medical informaticians, medical engineers, data scientists, cyber security experts, ethicists, sociologists, and legal experts, as well as patients and caregivers. ICT doctors initiate, promote, and integrate collaborations in practice and research and are trained to facilitate the development of ICT-based diagnostic, prognostic, therapeutic, and monitoring tools and processes and to identify problems in real-life situations that might benefit from novel ICT-based solutions. As to the latter, it appears that users find it difficult to conceive of or suggest new e-health services that might be useful to them in terms of demand for new services that do not currently exist [17]. ICT physicians will play a key role in the entire trajectory, from idea to innovation. During the design and development phases, they interact with designers, scientists, and practitioners of various medical disciplines, as well as with eventual commercial partners. They will guide scientific assessments and the evaluation of the evidence regarding the intended and unintended effects of ICT-based solutions. In the implementation phase, they will contribute to the education and coaching of the end users and to embedding the tools and services in daily practice. On a continuous basis, they will scientifically evaluate in real-life settings the acceptance, use, effectiveness, and cost-effectiveness of ICT-based changes. And they will educate doctors in digital technology and connectivity. In this way, ICT medicine substantially increases the likelihood that original ideas will transform into widely used cost-effective improvements or, if indicated by the evidence, see to it that provisory "innovations" are altered or discarded.

The integrative approach of ICT doctors may be particularly important in view of the tremendously rapid developments in the field of AI, such as highly flexible, reusable models (foundation models) [18]. The recently proposed generalist medical AI (GMAI) models are expected to be widely applied across medical applications for, among other uses, bedside decision support, augmented procedures, and chatbots for patients [18]. However, like other AI models, GMAI faces critical challenges regarding validation, verification, social bias, and scale [18]. The structural input of ICT doctors on a continuous basis may be invaluable in addressing these issues in terms of supervising the collection and sharing of the vast amounts of medical data that are required, guiding multidisciplinary verification of the input and output, and auditing for inaccuracies, misstatements, and social biases [18].

As adaptive AI algorithms change continuously in response to various types of use [19], it may be necessary with respect to medical AI applications to continuously study the patterns of interaction of doctors and patients with algorithms [19]. ICT doctors should be capable of cocreating and developing, in close collaboration with technologists and other stakeholders, the appropriate languages and methods to evaluate these interactions [19].

And, just as importantly, ICT physicians could be instrumental in facilitating the use of truly open and transparent AI systems in health care, as well as in researching the reproducibility of clinical AI tools, namely those for diagnostic and prognostic purposes.

Institute

Whereas in multidisciplinary or interdisciplinary collaborations, the various researchers remain employed at their respective organizations and located in their departments, integrative research should be carried out in an institute (an ICT medicine institute) that accommodates all researchers involved and provides them a tenure that complements their activities in other departments. The establishment of ICT medicine institutes also helps to secure the autonomy of ICT physicians, which may be particularly important as ICT-induced changes may at times be revolutionary or disruptive. As an institution, ICT medicine can provide a safe environment for researchers, given that conflicts of interest may arise between the results of an integrative approach and the interests of individual disciplines or commercial partners [20]. In fact, given the unforeseeable consequences of promising technologies like AI and quantum computing, the impact of the activities of ICT medicine is highly unpredictable.

In addition to enabling integrative collaborations with ICT and other experts, ICT medicine institutes may have a coordinating role regarding the often dispersed digital activities in the various departments of hospitals and health care organizations. By so doing they could help not only to prevent an uncontrolled growth of e-health projects and unnecessary duplication, but also promote standardization of and alignment between ICT-based practices inside and outside hospitals. The integration of digital activities between departments would significantly increase their effectiveness and efficiency.

With respect to large-scale ICT systems for public services, including health care, it may be very risky to completely rely on ICT companies. As these companies are shareholder value–maximizing firms, it is highly unlikely that they are objective sources of expertise and competence [21]. Conceivably, given the unique power of big ICT companies through contracts as advisors and vendors, they might ultimately even complicate or hinder ICT-dependent medical innovations [21]. And the updating of systems might be endangered when a company changes policies or ceases to exist. In the end, a dependency on ICT companies may weaken the health care system, as it induces a lack of in-house expertise that is needed for the quality and continuity of core activities [21]. For this reason alone, we should invest in the development of ICT medicine institutes. These institutes can become an independent force that can counterbalance the power of ICT companies. Where appropriate, ICT doctors should be able to cooperate with ICT specialists from public or nonprofit organizations without input from commercial companies, for example, in developing open-source generative AI models [22]. And distinct from commercial companies, ICT medicine institutes may more easily engage medically trained programmers; use programming languages that match the diversity of future users, such as domain-specific languages; and provide user-programmable software [23]. In this way, they may also promote collective digital trust among doctors and patients.

Human Factor

The impact of ICT on medical research and practice may seriously compromise the human factor in more ways than one. To paraphrase Reiser [24], why seek to inquire into the lives of patients to gain insights into their illness, which not only takes time but is fraught with undependability, if ICT-based techniques and procedures exist that give doctors the ability to identify and quantify clinically relevant signs of disease or changes in these signs by themselves? Thus, the widespread use of ICT, including AI, is set to create a new paradigm of examination and evaluation for the medicine of tomorrow [24,25]. However, technology is not a substitute for engaging with the life of the patient [24,25], and ICT doctors could be the primary defenders of patients' rights and perspectives.

Given the rapid developments in the field of AI, it is foreseeable that there will be an increasing tendency to eliminate human interference in the design of medical technology and programs and the building of medical and health care systems [26,27]. In the context of the AI-driven technicalization of medicine and health care, ICT doctors could contribute to the protection of the integrity and dignity of the human person and aspects of human values and humaneness [26-28].

Two scenarios are of particular interest in this respect. First, the medicine of the future might be almost completely determined by AI-based automated assessments, diagnostic decisions, and treatment procedures and practiced in AI-designed digitalized health care systems. Second, according to the "One Health" concept, the daily life events of humans are to be comprehensively monitored and analyzed in ICT-integrated environments, such as smart homes, smart cities, and smart hospitals and health care ecosystems [29,30]. In the end, the consequences of these developments could be that doctors and patients find themselves in fully controlled health care systems and that the doctor has transformed from a medical professional into a medical executive and employee.

Societal Dimension

ICT doctors can also help to ensure that the growing power of ICT-based health care innovations is used appropriately and to facilitate the fair allocation of their benefits [31]. ICT medicine institutes can provide a counterweight to the commercial dimensions of medicine and health care [31]. And in changing political, social, and economic circumstances, ICT physicians may contribute to the quality and continuity of digital medicine and health services [32]. Notably, in the era of ICT, digital inclusion is critical to health care equity; digital inclusion encompasses all activities that ensure that all individuals and communities, including the most disadvantaged, have access to and use of digital services [33]. With respect to medicine and health care, ICT physicians could play a key role in overcoming and preventing structural barriers to digital inclusion relating to age, race, socioeconomic status, language skills, and other factors [34]. They might also see to it that AI-based algorithms are adjusted to local patient populations and health care facilities.

Through vision and by thinking strategically, ICT doctors should be able to lead the way in the field of digital medicine [6,35]. As they have the necessary ICT skills and competence and understand care systems and their complexity, they are very well able to provide leadership in groundbreaking integrative collaborations [36]. Using long-term strategies, ICT doctors may increase awareness among all stakeholders of the potential added value and trust of ICT-based solutions [2]. Importantly, to ensure the implementation and continuity of ICT-based solutions from a financial and administrative perspective, ICT doctors should be able to perform management tasks in health organizations, health insurance companies, and authorities [2].

Historical Perspective

History shows that technical inventions can lay the foundation for innovations in medicine that have a fundamental and lasting impact on practice routines and patients' perspectives. Thus, the discovery of a specific type of electromagnetic radiation led to a revolutionary change in diagnostics and the emergence of the discipline of radiology. The increasingly rapid development of ICT-based technologies will be comparable in its dramatic impact. However, the effects of former technologies were limited to certain aspects of medicine, such as diagnosis (radiology) or the treatment of specific patient groups (radiotherapy); basically, the new disciplines complemented the existing ones. In contrast, the current ICT-based revolution pervades *medicine as a whole*, radically changing virtually every aspect of it. These drastic transformations are inadequately structured and formalized in the dispersed activities of digital medicine; they require an overarching specialty that comprehensively and systemically integrates all ICT-based developments and practices throughout all fields of medicine: ICT medicine. With respect to the all-pervasiveness of their impact, ICT developments are comparable to the expansion of microscopic and histological technologies in the 19th century. The latter made cellular pathology the foundation (ie, the infrastructure) of modern medicine, whereas ICT medicine, by operating on a meta level (as a suprastructure) will define the medicine of the 21st century.

Conclusions

Medicine has entered a period of epochal change. Within a lifetime, age-old practices based on doctors' individual expertise and collective wisdom are being superseded by knowledge- and evidence-based medicine characterized by, among other factors, automated assessments and AI-driven algorithms. ICT-based tools and processes will be indispensable parts of medical practice, in their role comparable to the preeminent position once held by detailed history taking and physical examination. It can be expected that in the near future, a wide array of ICT-based devices and procedures will be broadly applied by virtually all practitioners in most disciplines. To make this historical transformation a success we must create a new, integrative specialty—ICT medicine.

ICT doctors will not only contribute substantially to the development, implementation, and evaluation of digital tools, systems, and services they will provide a bridge between ICT professionals and clinical users, educate and train doctors in digital medicine, and safeguard the human factor; they may also be leaders in digitized health care organizations [37]. ICT medicine is the specialty that will provide a home for these new professionals [37,38]. Otherwise, we are at risk of having to practice a digital medicine that is both ineffective and costly and therefore poorly accepted by professionals and patients alike.

Acknowledgments

I gratefully acknowledge the reviewers for providing feedback that improved the manuscript.

Conflicts of Interest

PJJ has received honoraria from Bayer Netherlands and Orikami Personalized Health Care for consultancy activities and is chairman of the MSmonitor Foundation.

References

1. Ganapathy K. E-Medicine: transforming healthcare with information and communication technology. Med J Armed Forces India 2011 Apr;67(2):106-107 [FREE Full text] [doi: 10.1016/S0377-1237(11)60003-p8] [Medline: 27365780]
2. Hassan AYI. Challenges and recommendations for the deployment of information and communication technology solutions for informal caregivers: Scoping review. JMIR Aging 2020 Jul 29;3(2):e20310 [FREE Full text] [doi: 10.2196/20310] [Medline: 32723720]
3. Morita PP, Sahu KS, Oetomo A. Health monitoring using smart home technologies: Scoping review. JMIR Mhealth Uhealth 2023 Apr 13;11:e37347 [FREE Full text] [doi: 10.2196/37347] [Medline: 37052984]
4. Mann DM, Lawrence K. Reimagining connected care in the era of digital medicine. JMIR Mhealth Uhealth 2022 Apr 15;10(4):e34483 [FREE Full text] [doi: 10.2196/34483] [Medline: 35436238]
5. van Wijk MAM, van der Lei J, Mosseveld M, Bohnen A, van Bemmel JH. Compliance of general practitioners with a guideline-based decision support system for ordering blood tests. Clin Chem 2002 Jan;48(1):55-60. [Medline: 11751538]
6. Digital Medicine Society. URL: https://dimesociety.org/ [accessed 2023-07-06]
7. Goldsack J, Zanetti C. Defining and developing the workforce needed for success in the digital era of medicine. Digit Biomark 2020;4(Suppl 1):136-142 [FREE Full text] [doi: 10.1159/000512382] [Medline: 33442586]
8. Steinhubl SR, Topol EJ. Digital medicine, on its way to being just plain medicine. NPJ Digit Med 2018;1:20175 [FREE Full text] [doi: 10.1038/s41746-017-0005-1] [Medline: 31304349]
9. Bruce C, Harrison P, Giammattei C, Desai S, Sol JR, Jones S, et al. Evaluating patient-centered mobile health technologies: Definitions, methodologies, and outcomes. JMIR Mhealth Uhealth 2020 Nov 11;8(11):e17577 [FREE Full text] [doi: 10.2196/17577] [Medline: 33174846]
10. Coravos A, Goldsack J, Karlin D, Nebeker C, Perakslis E, Zimmerman N, et al. Digital medicine: A primer on measurement. Digit Biomark 2019;3(2):31-71 [FREE Full text] [doi: 10.1159/000500413] [Medline: 32095767]
11. Guimarães M, Pohl C, Bina O, Varanda M. Who is doing inter- and transdisciplinary research, and why? An empirical study of motivations, attitudes, skills, and behaviours. Futures 2019 Sep;112:102441 [FREE Full text] [doi: 10.1016/J.Futures.2019.102441]
12. Drake S, Reid J. 21st century competencies in light of the history of integrated curriculum. Front Educ 2020 Jul 14;5:1-10 [FREE Full text] [doi: 10.3389/feduc.2020.00122]
13. Drake S, Reid J. Thinking now: Transdisciplinary thinking as a disposition. Academia Letters 2021:387 [FREE Full text] [doi: 10.20935/AL387]
14. Bauer HH. Integrated science: Ambitions, practicalities, misunderstandings. In: Rezaei N, editor. Integrated Science. Cham, Switzerland: Springer; 2021.
15. Choi BCK, Pak AWP. Multidisciplinarity, interdisciplinarity, and transdisciplinarity in health research, services, education and policy: 2. Promotors, barriers, and strategies of enhancement. Clin Invest Med 2007;30(6):E224-E232. [doi: 10.25011/cim.v30i6.2950] [Medline: 18053389]
16. "Structure". Merriam-Webster Dictionary. URL: https://www..com/dictionary/structure [accessed 2023-07-06]
17. Cernadas Ramos A, Fernández Da Silva Á, Barral Buceta B, Bouzas-Lorenzo R. Offerings and user demands of eHealth services in Spain: National survey. J Med Internet Res 2023 May 18;25:e42304 [FREE Full text] [doi: 10.2196/42304] [Medline: 37200072]
18. Moor M, Banerjee O, Abad ZSH, Krumholz HM, Leskovec J, Topol EJ, et al. Foundation models for generalist medical artificial intelligence. Nature 2023 Apr;616(7956):259-265. [doi: 10.1038/s41586-023-05881-4] [Medline: 37045921]
19. Matias JN. Humans and algorithms work together - so study them together. Nature 2023 May;617(7960):248-251. [doi: 10.1038/d41586-023-01521-z] [Medline: 37165234]
20. Radtchenko-Draillard S. The fates of the human factor in science and contemporary society at the era of digitization. Academia Letters 2021:873. [doi: 10.20935/al873]
21. Mazzucato M, Collington R. The Big Con: How the Consulting Industry Weakens our Businesses, Infantilizes our Governments and Warps our Economies. London, UK: Penguin Press; 2023.
22. Spirling A. Why open-source generative AI models are an ethical way forward for science. Nature 2023 Apr;616(7957):413. [doi: 10.1038/d41586-023-01295-4] [Medline: 37072520]
23. Hermans F, Pinzger M, van Deursen A. Domain-specific languages in practice: A user study on the success factors. In: Proceedings in Lecture Notes in Computer Science. 2009 Presented at: Model Driven Engineering Languages and Systems: 12th International Conference, MODELS 2009; October 4-9, 2009; Denver, CO p. 4-9. [doi: 10.1007/978-3-642-04425-0_33]

24. Reiser S. Technology and the patient-physician relationship: a defining historic moment. Virtual Mentor 2007 Feb 01;9(2):143-146 [FREE Full text] [doi: 10.1001/virtualmentor.2007.9.2.mhum1-0702] [Medline: 23217764]
25. Rosenberg C. Back to the future. Lancet 2013 Sep 07;382(9895):851-852. [doi: 10.1016/s0140-6736(13)61852-8] [Medline: 24020069]
26. Lenk H. "Intelligent" systems technology: robots, autonomy and SET systems. Academia Letters 2021:2788 [FREE Full text] [doi: 10.20935/al2788]
27. Nabavi E. Why the huge growth in AI spells a big opportunity for transdisciplinary researchers. Nature 2019 Apr 23:429. [doi: 10.1038/d41586-019-01251-1] [Medline: 32317782]
28. Cipriano LE. Evaluating the impact and potential impact of machine learning on medical decision making. Med Decis Making 2023 Feb;43(2):147-149 [FREE Full text] [doi: 10.1177/0272989X221146506] [Medline: 36575951]
29. Ho CWL. Operationalizing "One health" as "One digital health" through a global framework that emphasizes fair and equitable sharing of benefits from the use of artificial intelligence and related digital technologies. Front Public Health 2022;10:768977 [FREE Full text] [doi: 10.3389/fpubh.2022.768977] [Medline: 35592084]
30. Benis A, Haghi M, Deserno TM, Tamburis O. One digital health intervention for monitoring human and animal welfare in smart cities: Viewpoint and use case. JMIR Med Inform 2023 May 19;11:e43871 [FREE Full text] [doi: 10.2196/43871] [Medline: 36305540]
31. Reiser SJ. The concept of service in medicine. J Pain Symptom Manage 2012 Jul;44(1):150-153 [FREE Full text] [doi: 10.1016/j.jpainsymman.2012.02.005] [Medline: 22704057]
32. Laur CV, Agarwal P, Mukerji G, Goulbourne E, Baranek H, Pus L, et al. Building health services in a rapidly changing landscape: Lessons in adaptive leadership and pivots in a COVID-19 remote monitoring program. J Med Internet Res 2021 Jan 13;23(1):e25507 [FREE Full text] [doi: 10.2196/25507] [Medline: 33417588]
33. National Digital Inclusion Alliance. URL: https://www.digitalinclusion.org/ [accessed 2023-07-06]
34. Rodriguez JA, Shachar C, Bates DW. Digital inclusion as health care - supporting health care equity with digital-infrastructure initiatives. N Engl J Med 2022 Mar 24;386(12):1101-1103. [doi: 10.1056/NEJMp2115646] [Medline: 35302722]
35. Laukka E, Pölkki T, Kanste O. Leadership in the context of digital health services: A concept analysis. J Nurs Manag 2022 Oct;30(7):2763-2780 [FREE Full text] [doi: 10.1111/jonm.13763] [Medline: 35942802]
36. Laukka E, Pölkki T, Heponiemi T, Kaihlanen A, Kanste O. Leadership in digital health services: Protocol for a concept analysis. JMIR Res Protoc 2021 Feb 04;10(2):e25495 [FREE Full text] [doi: 10.2196/25495] [Medline: 33538702]
37. Hersh W. Who are the informaticians? What we know and should know. J Am Med Inform Assoc 2006;13(2):166-170 [FREE Full text] [doi: 10.1197/jamia.M1912] [Medline: 16357354]
38. Ertl G. [Do we need a specialist physician for digital medicine?]. Dtsch Med Wochenschr 2018 Oct;143(20):1421. [doi: 10.1055/a-0669-1618] [Medline: 30286487]

Abbreviations

AI: artificial intelligence
EMR: electronic medical record
GMAI: generalist medical artificial intelligence
ICT: information and communication technology

Edited by T de Azevedo Cardoso; submitted 21.09.22; peer-reviewed by J Sedlakova, J Aarts; comments to author 26.01.23; revised version received 15.05.23; accepted 29.06.23; published 13.07.23.

Please cite as:
Jongen PJ
Information and Communication Technology Medicine: Integrative Specialty for the Future of Medicine
Interact J Med Res 2023;12:e42831
URL: https://www.i-jmr.org/2023/1/e42831
doi:10.2196/42831
PMID:37440294

Viewpoint

Supporting Public Health Research Capacity, Quality, and Productivity in a Diverse Region

Rana AlHamawi[1], MPH; Randa K Saad[1], MD; Hanan F Abdul Rahim[2], PhD; Khwaja Mir Islam Saeed[3], MSc; Abdullatif Husseini[4], PhD; Yousef Khader[5], SCD; Mohannad Al Nsour[1], PhD

[1]Global Health Development/Eastern Mediterranean Public Health Network, Amman, Jordan

[2]Qatar University Health, College of Health Sciences, Qatar University, Doha, Qatar

[3]Afghanistan Field Epidemiology Training Program, Afghanistan National Public Health Institute, kabul, Afghanistan

[4]Institute of Community and Public Health, Birzeit University, Ramallah, Occupied Palestinian Territory

[5]Public Health, Faculty of Medicine, Jordan University of Science and Technology, Irbid, Jordan

Corresponding Author:
Randa K Saad, MD
Global Health Development/Eastern Mediterranean Public Health Network
Abdallah Ben Abbas St, Building No. 42
Amman, 11196
Jordan
Phone: 962 6 5519962
Email: randaksaad@gmail.com

Abstract

Public health research plays a critical role in strengthening health systems and improving their performance and impact. However, scholarly production in public health coming from the Eastern Mediterranean Region (EMR) remains well below the world average and lacks a tangible growth trend over time. During the seventh Eastern Mediterranean Public Health Network Regional Conference, a roundtable session brought together a panel of public health experts representing Global Health Development/Eastern Mediterranean Public Health Network affiliates, universities or academia, and research institutions from the region, where they shared insights on the current situation of public health research; challenges and barriers to research facing the different countries in the EMR and the region in general; and how research agendas, productivity, and quality can be supported through strengthening research capacity in the region. Although the region is diverse in terms of health system capacity and socioeconomic development, several common challenges were identified, including a lack of strategic prioritization to guide health research, insufficient funding, ineffective transfer of knowledge to policy and practice, limited availability of research facilities, and limited national and international research collaboration. Occupied countries and countries in a state of conflict, such as Palestine, face additional barriers, such as personal and social security, lack of control of borders and natural resources, travel and movement restrictions, and confidentiality challenges because of the continuing war conditions and occupation. However, there have been success stories in the EMR regarding research publications and their positive and effective impact on policy and decision-makers. To improve research resilience and public health care in the region, a collaborative approach involving institutions, policymakers, and relevant stakeholders is critical.

(Interact J Med Res 2023;12:e39154) doi:10.2196/39154

KEYWORDS
health research; Public health; Eastern Mediterranean region; Research capacity; Research Quality

Introduction

Public health research plays a critical role in strengthening health systems, improving their performance and public health impact, and adding value to society, as it provides critical information regarding disease trends, risk factors, outcomes of treatments or health interventions, patterns of care, and health care use and costs [1]. The importance of building research capacity in low-and middle-income countries (LMICs) has been recognized for well over 2 decades. The 1990 Commission on Health Research for Development reported that strengthening research capacity in LMICs is "one of the most powerful, cost-effective and sustainable means of advancing health and development" [2]. Indeed, locally led health research is important in overcoming global health barriers and challenges in LMICs [3].

Researchers usually initiate research studies with the best of intentions. However, the research process may face several challenges that hinder attempts for achieving successful results, especially in a region as diverse and facing as many political and financial challenges as the Eastern Mediterranean Region (EMR), which includes 22 countries as defined by the World Health Organization [4]. Therefore, research efforts and capacities in the EMR must be comprehensively reviewed within a national, regional, and global context.

This viewpoint aims to share insights from a roundtable discussion on the challenges and barriers facing public health research in the EMR, in particular the LMICs, and share views on what can be done to strengthen research capacity and improve research quality and productivity in the EMR.

Roundtable Description

Overview

A roundtable session was held on November 18, 2021, as part of the Seventh Eastern Mediterranean Public Health Network (EMPHNET) Regional Conference. The conference was attended by many experts from ministries of health, academic institutions, and public health institutes who represent most of the EMR. The roundtable brought together a panel of public health experts representing Global Health Development (GHD)|EMPHNET affiliates, universities or academia, and research institutions at regional and global levels. It provided a space for sharing insights on the challenges and barriers to research in the EMR and how research agenda can be supported by strengthening research capacity and increasing research productivity and quality in the region. The roundtable included oral presentations and an interactive discussion of questions and comments from participants. The following topics were presented to address the roundtable objectives: barriers and challenges to public health research in the EMR and strategies for improvement; investing in public health research and the COVID-19 response; a perspective from Qatar; public health research under occupation; an example from Palestine; and public health research in the EMR and the contribution of GHD|EMPHNET.

Public Health Research in the EMR

Public health research and scientific production coming from the Middle East remains well below the world average and has not increased significantly over time [5]. The global share of EMR health research publications is much smaller than its global share of population or wealth. Between the years 2004 and 2013, the EMR contributed to only 0.85%-2.36% of total PubMed publications in health, and only 5 countries in the region accounted for 80% of the total publications: the Islamic Republic of Iran (39%), Egypt (14%), Saudi Arabia (11%), Tunisia (8%), and Pakistan (8%) [6]. Additionally, even though the population of the EMR has reached nearly 745 million people [4], only 2.35 articles were published per 100,000 population per year during 2003 to 2013 [6]. The low number of publications and lack of a comprehensive contribution from all countries in the EMR can be attributed to the several challenges faced in the region.

Barriers and Challenges to Public Health Research in the EMR and Strategies for Improvement

Table 1 illustrates some of the challenges and barriers affecting the quality, capacity, and productivity of research in the EMR, including strategies for mitigating these barriers and improving overall research in the region.

Table 1. Barriers facing research in the Eastern Mediterranean Region and strategies for improvement.

Barriers	Possible strategies for improvement
Lack of an identified research agenda based on emerging priorities [7-9]	• Set national research priorities and raise awareness about the importance of research for evidence-based decisions and policies [7,10]
Limited national and regional strategies for health research, and a lack of national policies and regulations that govern the conduct of health research [7,9,11]	• Adopt national plans, strategies, and policies for effective health research [10] • Adopt a regional strategy between supranational organizations and groupings to steer improvements in health research and encourage effective research stewardship in the Eastern Mediterranean Region [11].
Ineffective knowledge transfer to policy and practice [7,9,11]	• Build capacities of policymakers to demand information for decision-making and carry out informed decision-making [12] • Develop knowledge translation platforms to support the dissemination of evidence-based data and facilitate the dialogue between researchers, policymakers, and relevant stakeholders [13]
Limited availability of research facilities, equipment, and resources [9]	• Foster an appropriate and effective research environment by advocating for national investment in public health infrastructure, such as research infrastructure and resources (libraries, scientific literature, communication facilities) [11,14], and maintain it by encouraging regional collaboration of countries with different resource levels.
Limited collaboration between research institutions and other national, regional, and international research institutions [9]	• Develop and share a database of researchers and their expertise [15] • Build capacities in grant writing and ensure that researchers from the region engage in international platforms, including conferences and meetings, to advocate for the Eastern Mediterranean Region's issues [16]
Insufficient human resource capacity regarding research skills and competencies; lack of overall human resources [9,17,18]	• Building capacities in research should be tied to the educational system, especially higher education. Academic research centers or institutes should focus on capacity building, improving scientific writing or publishing, health research proposal writing or funding, and quantitative and qualitative research methods. Continuous learning should be adopted as a strategy to build research capacities [19] • Organize training courses for researchers and editors, such as "train the trainers" courses [20]
Limited national and regional research funding and ongoing sustainability of funding [8,11]	• Establish a research finance system using innovative revenue generation [21] • Advocate for funding through shared causes and engaging with the media [22]
Inadequate mentorship [9]	• Provide recognition and long-term funded positions to mentors [23] • When mentoring is locally unavailable, peer mentorship or institutional exchange or partnerships can be implemented [24]
Difficulties in retaining qualified researchers, research assistants, and associates (brain drain). As well as a lack of motivation and incentives to encourage researchers to conduct and engage in public health research [25]	• To discourage brain drain, countries should ensure an entire ecosystem exists that supports and strengthens research, including making available research support (funding), human resource support (qualified research assistants), opportunities for dissemination and networking, physical infrastructure, and research compliance mechanisms [26] • Provide higher salaries and link research production and publication to academic promotion [27]

Fostering and promoting an efficient environment appropriate for health research is required for planning, designing, and implementing research, disseminating evidence-based data, and translating the findings into evidence-informed policies and decisions [7]. One of the challenges facing the advancement of research in most countries in the EMR is that health and health research are not seen as a priority in terms of funding by most governments, the private sector, or the third sector (organizations). Investments in research systems are low in comparison to health in general [11]. Evidence has shown that research and development in the EMR is among the lowest in the world with an average of 0.3% of the region's gross domestic product, which is well below world leaders such as Japan (2.8%) and the United Kingdom (1.8%) [8]. Research institutions conducting quantitative or qualitative research are receiving limited national, regional, and international funds [11]. The limitations in national funding can be due to national governments and policymakers allocating limited budgets for health research [2,8,11], policymakers being unaware of the value and importance of health research in improving public health and health systems, lack of engagement of policymakers in health research, and lack of a national health strategy and policies that define research priorities and guide health research in the region [7,9,11]. Indeed, previous research has shown that only 3 out of 10 EMR countries had reported setting national research priorities and only 2 countries had a national health research policy in place [5,8]. Additionally, it has been shown that only 29.3% of research institutions in the EMR reported having policymakers as part of their advisory board [11].

The low rate of investment of governments in health research is becoming apparent in the productivity of health research in the EMR. Previous research has shown that the research productivity of the EMR is very low next to international comparators (Europe, the Americas, and the Western Pacific). The EMR is only comparable to Africa and Southeast Asia. However, it is still behind the latter [28,29]. Additionally, research productivity in the EMR is not evenly distributed among countries, with Egypt and Pakistan contributing around 62% of the articles in the Index Medicus, while Afghanistan, Djibouti, and Somalia have not contributed any articles to the Index Medicus [30]. Evidence has shown that the publishing of biomedical and health research is directly linked to the state of socioeconomic development and political situation of the country [31,32]. The 3 countries that have contributed the least are among the least developed in the world. Somalia specifically has been experiencing a protracted emergency (a civil war) since the 1990s. The civil war, accompanied by famine and population displacement, has severely affected Somalia’s health system. In 2010, the transition to recovery of the health system organization, regulation, and workforce development began, with the institution of a transitional federal government. This transition created opportunities to initiate the pursuit of universal health coverage. During this phase, around 25 academic institutions with undergraduate medical or health courses and a few master’s courses were operationalized, in addition to reviving the collaboration between Swedish universities that was offset by the civil war [33,34]. There have been some additional efforts to accelerate and prioritize research in Somalia; for example, in 2022, the Somali National Institute of Health (NIH) and Federal Ministry of Health organized Somalia’s first NIH Health Research Conference in collaboration with multiple actors such as the World Health Organization, the Alliance for Health Policy and Systems Research, the Public Health Agency of Sweden, the African Field Epidemiology Network, Somali universities, the Somali-Swedish Research Cooperation Initiative, and the Somali Swedish Researchers’ Association [35]. The conference placed research at the forefront to accelerate progress toward universal health coverage [35]. The *Somali Health Action Journal*, established in collaboration between Somalia and Swedish universities, was showcased during the conference. The *Somali Health Action Journal* offers an open-access dissemination platform for addressing Somalia’s health challenges [35,36]. Despite all these efforts, Somalia’s national health system remains fragile, and national public health research is an essential component toward health system resilience.

Another barrier to a resilient public health system in the EMR is that research agendas in LMICs are set by international funders rather than local research institutions. Furthermore, the research implemented in LMICs is mainly led by researchers from high-income countries (HICs), with little contribution from LMIC researchers [9]. The small contribution from LMICs in conducting the research can create what is known as “parachute research,” where research is collected in LMICs but further work is carried out in HIC research institutions. This can severely diminish research capacity development [37] within the EMR and implement priorities that match the funder’s priorities than the actual health needs of the region [17].

The research in the EMR experience from lack of knowledge dissemination and transfer to policy [9], lack of national policies and regulations that govern the conduct of research [7,9,11], brain drain of researchers to HICs [25], lack of mentorship [9], and insufficient human resource competencies and skills [9,17]. For research to be effective in the EMR, a systems approach to capacity building should be adopted [38]. This approach addresses the national research system on 3 levels: macro, institutional, and individual. For health research to contribute to the reduction of health inequalities, research must be conducted on the basis of a research system that has well-defined goals and priorities [39]. To develop such a resilient system, countries must invest in capacity building at the macro level, where leadership and management skills are fostered [40]. At the institutional level, countries must focus on fostering an appropriate and effective research environment for their researchers. For example, having adequate and appropriate research infrastructure and resources (scientific literature, communication facilities) and maintaining the continuity of funding [14]. The individual level focuses on building technical competence (in data analysis and protocol development) but also builds the individual’s capacities in other aspects of research such as priority setting, networking and leadership, disseminating and translating knowledge, and partnership development. This level is not only targeted at researchers but must include other stakeholders, such as decision-makers, health workers, research managers, and community members [14].

The unstable political leadership in the EMR is another significant barrier for public health research. In countries with unstable political leadership, research funding and priorities are inconsistent and unpredictable, making it difficult for researchers to plan and conduct their work effectively. Additionally, political instability can create an environment of insecurity and uncertainty, which can discourage international collaboration and investment in research. Political instability can also lead to brain drain, as researchers may leave the country in search of more stable and supportive research environments.

Public Health Research Under Occupation: An Example From Palestine

It is evident that colonial structures where indigenous people live foster material and social inequalities, which lead to health disparities that persist over several generations. Diminished life expectancy, a disproportionate burden of communicable and noncommunicable diseases, social violence, and addiction have been linked to colonial structures [11]. “Research has increasingly established that poor health outcomes in Indigenous peoples, and the health disparities realized by Indigenous peoples in almost all sectors of life as compared with their nonindigenous counterparts, stem from or are related to colonial disruptions and ongoing erosion of human rights” [41]. Therefore, encouraging research in such countries is critical for the advancement of the health of indigenous people. However, public health research is severely affected in occupied countries such as Palestine, where fragmentation of communities, land,

and the health system and dependence on international aid are prevalent.

Despite the benefits of research in conflict zones, researchers are faced with challenges at every step of the research process, from conducting fieldwork and disseminating research findings to the repatriation of researchers [40]. Researchers conducting research in occupied countries such as Palestine are faced with the abovementioned barriers and challenges. However, due to the continuing war conditions and the Israeli military occupation in Palestine [42], they are also faced with personal and social security and confidentiality challenges. For example, safety is a significant issue facing both researchers and communities in conflict zones. Researchers should sometimes avoid the use of participatory methods such as the gathering of many people in one place, as they represent high-risk strategies in areas subject to military aggression [14]. Researchers must also be reflective on where and how they conduct research, what they talk about, and who they talk to to avoid jeopardizing communities' safety. Researchers must be thoroughly trained on how to conduct research in conflict zones and should develop adequate skills and competencies that allow them to accurately assess the political situation. They should be able to identify which topics are too sensitive to talk about and can therefore put lives in the communities at risk [14,43]. Additional obstacles related to conducting research in occupied countries include a lack of control of borders and natural resources and travel and movement restrictions. This may affect sample transportation to other countries and purchasing equipment needed for research, causing delays and affecting the research timelines and outcomes. Finally, the dependence on international aid to fund health research, which may have its own research agenda, may not be compatible with occupied countries' research needs and agendas.

Communities living in conflict zones might revert to a strategy of silence, where they keep a low profile and mind their own business to protect themselves from militarized violence, including ethnic cleansing and demonstrative killings. Reverting to silence might also be a coping mechanism used by traumatized individuals. Therefore, researchers must be aware of the sensitivity of the collected information and the fears of the communities and respect the boundaries of the individuals. Researchers might be faced with challenges when it comes to disseminating such sensitive information for the greater good of these communities without risking the welfare of research subjects [14].

Researchers must follow the "do no harm" strategy in conflict zones to reduce the negative impact of research. This strategy requires the selection of well-trained and mature researchers that are aware of ethical dilemmas. It also involves a balance between insider and outsider researchers with relevant ethnic backgrounds and language, cultural sensitivity, and religious skills. Researchers are required to blend in with their surroundings, keep a low profile, prepare methodological contingency plans, frequently monitor the political situation, analyze risk, obtain informed consent, and maintain confidentiality [14,43].

Benefits of Investing in Public Health Research: A Perspective From Qatar

Despite the challenges that face the EMR, there have been some success stories. Qatar's response to the COVID-19 pandemic has been effective due to a number of factors, including the use of high-quality epidemiological and clinical data to support the national response efforts. In the first wave of COVID-19, the number of COVID-19 cases in Qatar was especially high among craft and manual workers living in dormitory-like conditions. However, case fatality has been among the lowest globally [7]. The success of the national COVID-19 response can be attributed to the interplay of a number of factors, including (1) linking science to policy through a multi-stakeholder platform as part of an effective governance mechanism, where scientific evidence was used to direct appropriate public health and health care measures against the pandemic; (2) providing rapid and flexible research funding (the Qatar National Research Fund established a rapid response call, offering researchers grants of up to US $27,460 each for impactful 3-month projects related to COVID-19, and Qatar University offered an emergency response grant of up to US $39,000 for each COVID-19–related research project); (3) readying infrastructure for infectious disease research ahead of the pandemic, including the biosafety level 3 laboratory at Qatar University; and (4) ensuring the availability of centralized and complete data, where the national SARS-CoV-2 databases are integrated using a digital health information platform [44]. Because of these factors, scientists from Qatar were able to contribute to international scholarship on the unfolding pandemic through a number of high-profile publications. It was seen that gains made in COVID-19 research in terms of data access were linked to policy, and collaboration should be sustained and implemented in other research areas.

Contribution of GHD|EMPHNET to Public Health Research

Despite the many challenges and barriers that face the EMR, GHD|EMPHNET was able to contribute to the improvement of the region's public health resilience and strengthen public health researchers' capacities. GHD|EMPHNET supports researchers through a hybrid model, where face-to-face and preconference workshops, Zoom, and Microsoft Team meetings are carried out whenever possible. For example, GHD|EMPHNET's knowledge exchange network creates a space for field epidemiology training program residents and graduates to share and exchange their ideas, discuss their work, learn from one another, and improve their skills through continuous collaboration. It also provides technical support to researchers. It has formed a research group targeting priority areas that involves a core team of lead researchers and more than 40 coresearchers. It has also published its own e-book and its fair share of case studies and papers. GHD|EMPHNET launched its own webinar initiative, the EMPHNET WEBi series, that serves as a web-based dialogue opportunity to bring together a wide range of audiences. Its objectives are to enhance relationship and coordination between countries, stakeholders, and partners; maximize the use of available measures to enhance and develop public health expertise, capacity, and community awareness; and to disseminate information and knowledge applicable to the public health and EMR priorities and needs.

Conclusions

There are many challenges and barriers facing public health research capacity, quality, and productivity in the EMR. The identified challenges in the EMR are a lack of national health strategy and policies that define research priorities and guide health research, lack of funding, lack of effective knowledge transfer to policy and practice, limited availability of research facilities and resources, limited national and international research collaboration, inadequate competencies of human resources, brain drain, and inadequate mentorship. Researchers conducting research in conflict contexts are faced with more challenges compared to many other settings, such as social security and confidentiality barriers. However, there have been success stories in the EMR regarding research publication and its positive and effective impact on policy- and decision-makers. To improve research resilience and public health care in the region, a collaborative approach involving institutions, policymakers, and relevant stakeholders is critical.

Recommendations and Key Areas for Improvement

- Public health research capacity and competencies can be improved through collaborative work with local research teams. Collaboration with research team members and coauthors from credible international universities leads to improvements in research capacities and strengthens research in the region.
- In a conflict country, efforts to contribute to public health research in contexts of violence and political and economic oppression are of priority to the region, are highly valued by the global research community, and may merit the granting of research funds.
- Research in conflict areas needs to investigate the needs of people living in such areas and consequently the appropriate responses to their needs. Therefore, a framework should be developed to assist researchers committed to ethical decision-making.
- Official guidelines on data sharing should be developed, should be clear and consistent, and should balance between making data available and accessible to researchers and safeguarding privacy. A centralized mechanism such as a secretariat or a commission should monitor and facilitate data sharing among various stakeholders for efficient and fair use of data for the health of the community.
- Universities must make research skills and competencies part of the curriculum for undergraduate students. Efforts should be made to mentor undergraduate students and develop their research competencies. Mentors should be given incentives to encourage mentorship. Furthermore, universities should develop better archival databases and consequently provide better access to up-to-date data for their students.
- Research institutions must focus on developing a research capacity educational program such as “train the trainer,” where institutions adopting such a program must be flexible and willing to revise the plan if faced with barriers and challenges.
- Knowledge transfer frameworks and programs should be developed and implemented for collaborative knowledge transfer between researchers, policymakers, and other relevant stakeholders to facilitate the linkage of science to policy.

Conflicts of Interest

None declared.

References

1. Institute of Medicine (US) Committee on Health Research and the Privacy of Health Information: The HIPAA Privacy Rule. In: Nass SJ, Levit LA, Gostin LO, editors. Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research. Washington, DC: The National Academies Press; 2009.
2. The Commission on Health Research for Development. Health Research: Essential Link to Equity in Development. New York, NY: Oxford University Press; 1990.
3. Dye C, Evans D, Harries A, Lienhardt C, McManus J, Pang T, et al. The World Health Report 2013: Research for Universal Health Coverage. World Health Organisation. 2013. URL: https://www.who.int/publications/i/item/9789240690837 [accessed 2023-07-20]
4. Regional Office for the Eastern Mediterranean. World Health Organization. 2023. URL: https://www.emro.who.int/countries.html [accessed 2023-03-18]
5. Cavacini A. Recent trends in Middle Eastern scientific production. Scientometrics 2016;109(1):423-432. [doi: 10.1007/s11192-016-1932-3]
6. Tadmouri GO, Mandil A, Rashidian A. Biomedical and health research geography in the Eastern Mediterranean region. East Mediterr Health J 2019;25(10):728-743 [FREE Full text] [doi: 10.26719/emhj.19.082] [Medline: 31774139]
7. Mandil A, El-Jardali F, El-Feky S, Nour M, Al-Abbar M, Bou-Karroum L. Health research institutional mapping: an Eastern Mediterranean regional perspective. East Mediterr Health J 2018;24(2):189-197 [FREE Full text] [Medline: 29748948]
8. Kennedy A, Khoja TAM, Abou-Zeid AH, Ghannem H, IJsselmuiden C, WHO-EMRO/COHRED/GCC NHRS Collaborative Group. National health research system mapping in 10 Eastern Mediterranean countries. East Mediterr Health J 2008;14(3):502-517 [FREE Full text] [Medline: 18720615]

9. Franzen SRP, Chandler C, Lang T. Health research capacity development in low and middle income countries: reality or rhetoric? A systematic meta-narrative review of the qualitative literature. BMJ Open 2017;7(1):e012332 [FREE Full text] [doi: 10.1136/bmjopen-2016-012332] [Medline: 28131997]
10. McKee M, Stuckler D, Basu S. Where there is no health research: what can be done to fill the global gaps in health research? PLoS Med 2012;9(4):e1001209 [FREE Full text] [doi: 10.1371/journal.pmed.1001209] [Medline: 22545025]
11. Ismail SA, McDonald A, Dubois E, Aljohani FG, Coutts AP, Majeed A, et al. Assessing the state of health research in the Eastern Mediterranean region. J R Soc Med 2013;106(6):224-233 [FREE Full text] [doi: 10.1258/jrsm.2012.120240] [Medline: 23761582]
12. Sadana R, D'Souza C, Hyder AA, Chowdhury AMR. Importance of health research in South Asia. BMJ 2004;328(7443):826-830 [FREE Full text] [doi: 10.1136/bmj.328.7443.826] [Medline: 15070643]
13. Kasonde JM, Campbell S. Creating a knowledge translation platform: nine lessons from the Zambia Forum for health research. Health Res Policy Syst 2012;10:31 [FREE Full text] [doi: 10.1186/1478-4505-10-31] [Medline: 23034056]
14. Minja H, Nsanzabana C, Maure C, Hoffmann A, Rumisha S, Ogundahunsi O, et al. Impact of health research capacity strengthening in low- and middle-income countries: the case of WHO/TDR programmes. PLoS Negl Trop Dis 2011;5(10):e1351 [FREE Full text] [doi: 10.1371/journal.pntd.0001351] [Medline: 22022630]
15. Chandiwana S, Ornbjerg N. Review of North-South and South-South cooperation and conditions necessary to sustain research capability in developing countries. J Health Popul Nutr 2003;21(3):288-297. [Medline: 14717574]
16. Thornicroft G, Cooper S, Bortel TV, Kakuma R, Lund C. Capacity building in global mental health research. Harv Rev Psychiatry 2012;20(1):13-24 [FREE Full text] [doi: 10.3109/10673229.2012.649117] [Medline: 22335179]
17. Kok MO, Rodrigues A, Silva AP, de Haan S. The emergence and current performance of a health research system: lessons from Guinea Bissau. Health Res Policy Syst 2012;10:5 [FREE Full text] [doi: 10.1186/1478-4505-10-5] [Medline: 22321566]
18. Ogundahunsi OAT, Vahedi M, Kamau EM, Aslanyan G, Terry RF, Zicker F, et al. Strengthening research capacity-TDR's evolving experience in low- and middle-income countries? PLoS Negl Trop Dis 2015;9(1):e3380 [FREE Full text] [doi: 10.1371/journal.pntd.0003380] [Medline: 25569232]
19. Planning, monitoring and evaluation framework for capacity strengthening in health research. Essence on Health Research. 2011. URL: https://tdr.who.int/docs/librariesprovider10/essence/essence-frwk-2016-web-pdf.pdf?sfvrsn=7282f353_6 [accessed 2023-07-14]
20. Gómez L, Jaramillo A, Halpaap B, Launois P, Cuervo LG, Saravia NG. Building research capacity through "Planning for Success". PLoS Negl Trop Dis 2019;13(8):e0007426 [FREE Full text] [doi: 10.1371/journal.pntd.0007426] [Medline: 31369556]
21. Zachariah R, Reid T, Srinath S, Chakaya J, Legins K, Karunakara U, et al. Building leadership capacity and future leaders in operational research in low-income countries: why and how? Int J Tuberc Lung Dis 2011;15(11):1426-1435, i. [doi: 10.5588/ijtld.11.0316] [Medline: 22008755]
22. Ntoumi F. Networking and capacity building for health research in Central Africa. Wien Klin Wochenschr 2010;122(Suppl 1):23-26. [doi: 10.1007/s00508-010-1331-3] [Medline: 20376720]
23. Whitworth JAG, Kokwaro G, Kinyanjui S, Snewin VA, Tanner M, Walport M, et al. Strengthening capacity for health research in Africa. Lancet 2008;372(9649):1590-1593 [FREE Full text] [doi: 10.1016/S0140-6736(08)61660-8] [Medline: 18984193]
24. Bates I, Akoto AYO, Ansong D, Karikari P, Bedu-Addo G, Critchley J, et al. Evaluating health research capacity building: an evidence-based tool. PLoS Med 2006;3(8):e299 [FREE Full text] [doi: 10.1371/journal.pmed.0030299] [Medline: 16942394]
25. Lansang MA, Dennis R. Building capacity in health research in the developing world. Bull World Health Organ 2004;82(10):764-770 [FREE Full text] [Medline: 15643798]
26. Kupfer L, Hofman K, Jarawan R, McDermott J, Bridbord K. Roundtable. Strategies to discourage brain drain. Bull World Health Organ 2004;82(8):616-619, discussion 619-623 [FREE Full text] [Medline: 15375452]
27. Dakik HA, Kaidbey H, Sabra R. Research productivity of the medical faculty at the American University of Beirut. Postgrad Med J 2006;82(969):462-464 [FREE Full text] [doi: 10.1136/pgmj.2005.042713] [Medline: 16822923]
28. Adams J, King C, Pendlebury D, Hook D, Wilsdon J, Zewail A. Global Research Report Middle East: Exploring the Changing Landscape of Arabian, Persian and Turkish Research. Leeds, UK: Evidence/Thomson Reuters; 2011.
29. Adams J, King C, Hook D. Global research report: Africa. Evidence/Thomson Reuters. 2010. URL: https://hedbib.iau-aiu.net/pdf/ThompsonReuters_globalresearchreport-africa.pdf [accessed 2023-07-20]
30. Al-Shorbaji N. Index Medicus for the eastern Mediterranean region. Emerg Themes Epidemiol 2008;5(1):14 [FREE Full text] [doi: 10.1186/1742-7622-5-14] [Medline: 18826568]
31. Bliziotis IA, Paraschakis K, Vergidis PI, Karavasiou AI, Falagas ME. Worldwide trends in quantity and quality of published articles in the field of infectious diseases. BMC Infect Dis 2005;5:16 [FREE Full text] [doi: 10.1186/1471-2334-5-16] [Medline: 15780136]
32. Rahman M, Fukui T. Biomedical research productivity: factors across the countries. Int J Technol Assess Health Care 2003;19(1):249-252. [doi: 10.1017/s0266462303000229] [Medline: 12701955]

33. Somali-Swedish Action Group for Health Research and Development. Healing the health system after civil unrest. Glob Health Action 2015;8:27381 [FREE Full text] [doi: 10.3402/gha.v8.27381] [Medline: 25828070]
34. Dalmar AA, Hussein AS, Walhad SA, Ibrahim AO, Abdi AA, Ali MK, et al. Rebuilding research capacity in fragile states: the case of a Somali-Swedish global health initiative. Glob Health Action 2017;10(1):1348693 [FREE Full text] [doi: 10.1080/16549716.2017.1348693] [Medline: 28799463]
35. Bile K, Warsame M, Ahmed AD. Fragile states need essential national health research: the case of Somalia. Lancet Glob Health 2022;10(5):e617-e618 [FREE Full text] [doi: 10.1016/S2214-109X(22)00122-X] [Medline: 35427511]
36. Mutapi F. Africa should set its own health-research agenda. Nature 2019;575(7784):567. [doi: 10.1038/d41586-019-03627-9] [Medline: 31772371]
37. Mony PK, Kurpad A, Vaz M. Capacity building in collaborative research is essential. BMJ 2005;331(7520):843-844 [FREE Full text] [doi: 10.1136/bmj.331.7520.843-b] [Medline: 16210297]
38. Annerstedt J, Liyanage S. Challenges when Shaping Capabilities for Research: Swedish Support to Bilateral Research Cooperation with Sri Lanka and Vietnam, 1976-2006, and a Look Ahead. Sweden: Swedish International Development Cooperation Agency; 2008.
39. International Organizing Committee. Health research for development: the continuing challenge: a discussion paper prepared for the International Conference on Health Research for Development. 2000 Presented at: International Conference on Health Research for Development; October 10-13, 2000; Bangkok, Thailand URL: https://www.cohred.org/downloads/727.pdf
40. Nuyens Y, McKee NA. No Development without Research: A Challenge for Capacity Strengthening. Geneva, Switzerland: Global Forum for Health Research; 2005.
41. Czyzewski K. Colonialism as a broader social determinant of health. Int Indig Policy J 2011;2(1):5. [doi: 10.18584/iipj.2011.2.1.5]
42. Giacaman R, Khatib R, Shabaneh L, Ramlawi A, Sabri B, Sabatinelli G, et al. Health status and health services in the occupied Palestinian territory. Lancet 2009;373(9666):837-849. [doi: 10.1016/s0140-6736(09)60107-0]
43. Drake G. The ethical and methodological challenges of social work research with participants who fear retribution: To 'do no harm'. Qual Soc Work 2013;13(2):304-319. [doi: 10.1177/1473325012473499]
44. Al Khal A, Al-Kaabi S, Checketts RJ. Qatar's response to COVID-19 pandemic. Heart Views 2020;21(3):129-132 [FREE Full text] [doi: 10.4103/HEARTVIEWS.HEARTVIEWS_161_20] [Medline: 33688398]

Abbreviations

EMR: Eastern Mediterranean Region
GHD|EMPHNET: Global Health Development|Eastern Mediterranean Public Health Network
HIC: high-income country
LMIC: low-and middle-income country
NIH: National Institutes of Health

Edited by A Mavragani; submitted 30.04.22; peer-reviewed by D Verran, C Abbafati, A Ahmed, Y Zhang; comments to author 14.03.23; revised version received 18.03.23; accepted 08.05.23; published 31.07.23.

Please cite as:
AlHamawi R, Saad RK, Abdul Rahim HF, Mir Islam Saeed K, Husseini A, Khader Y, Al Nsour M
Supporting Public Health Research Capacity, Quality, and Productivity in a Diverse Region
Interact J Med Res 2023;12:e39154
URL: https://www.i-jmr.org/2023/1/e39154
doi:10.2196/39154
PMID:37523227

Viewpoint

Normal Workflow and Key Strategies for Data Cleaning Toward Real-World Data: Viewpoint

Manping Guo[1,2,3], PhD; Yiming Wang[3], MD; Qiaoning Yang[3], PhD; Rui Li[4], PhD; Yang Zhao[3], MD; Chenfei Li[3], MD; Mingbo Zhu[5], BSc; Yao Cui[5], BSc; Xin Jiang[5], BSc; Song Sheng[3], PhD; Qingna Li[3*], MD; Rui Gao[3*], PhD

[1]Postdoctoral Research Station, China Academy of Chinese Medical Sciences, Beijing, China
[2]Postdoctoral Works Station, Yabao Pharmaceutical Group Co, Ltd, Yuncheng, China
[3]Xiyuan Hospital, China Academy of Chinese Medical Sciences, Beijing, China
[4]First Affiliated Hospital, Henan University of Traditional Chinese Medicine, Zhenzhou, China
[5]Beijing Xinshi Yuanzhi Technology Consulting Co, Ltd, Beijing, China
[*]these authors contributed equally

Corresponding Author:
Rui Gao, PhD
Xiyuan Hospital
China Academy of Chinese Medical Sciences
GCP, Xiyuan Hospital
1 Xiyuan Playground, Haidian District
Beijing, 100091
China
Phone: 86 010 62835653
Email: ruigao@126.com

Abstract

With the rapid development of science, technology, and engineering, large amounts of data have been generated in many fields in the past 20 years. In the process of medical research, data are constantly generated, and large amounts of real-world data form a “data disaster.” Effective data analysis and mining are based on data availability and high data quality. The premise of high data quality is the need to clean the data. Data cleaning is the process of detecting and correcting “dirty data,” which is the basis of data analysis and management. Moreover, data cleaning is a common technology for improving data quality. However, the current literature on real-world research provides little guidance on how to efficiently and ethically set up and perform data cleaning. To address this issue, we proposed a data cleaning framework for real-world research, focusing on the 3 most common types of dirty data (duplicate, missing, and outlier data), and a normal workflow for data cleaning to serve as a reference for the application of such technologies in future studies. We also provided relevant suggestions for common problems in data cleaning.

(Interact J Med Res 2023;12:e44310) doi:10.2196/44310

KEYWORDS

data cleaning; data quality; key technologies; real-world data; viewpoint

Introduction

Randomized controlled trials (RCTs) are considered to yield the highest-level evidence in the practice of evidence-based medicine, representing the “gold standard” for evaluating the safety and efficacy of drugs [1]. However, the extrapolation of RCT results to real-world situations is limited because of strict screening conditions, single-intervention measures, and limited sample sizes [2]. To compensate for the shortcomings of RCTs, Kaplan et al [3] first proposed the concept of real-world research (RWS) in 1993. RWS focuses on using high-quality, real-world data to generate reliable evidence regarding the effects of medical interventions in a real-world environment to complement the evidence generated from traditional RCTs.

The role of massive real-world data in academic and business environments has become increasingly significant [4]. To better understand the value of these data, data mining and analyses are required [5]. However, real-world data from medical practice are usually generated without the strict process controls applied in clinical trials. For example, when data are collected from multiple sources, such as different hospitals or hospital systems, the rules for integration may not be identical, leading to quality issues such as duplicate, missing, and outlier data. These “dirty data” are ubiquitous in RWS. Among them, the large amount of storage space required for duplicate data affects the efficiency of the database, and the inappropriate processing of missing data can result in the loss of a considerable amount of potentially useful information. Furthermore, inconsistent or incorrect outlier data can seriously affect the results of data analyses and key calculations, which may even provide incorrect directions for subsequent academic research, resulting in a loss of time, effort, and funding [6].

The China Center for Drug Evaluation issued a document titled "Guidance on Real-World Data for Generating Real-World Evidence (Trial)" [7] (referred to as "the guidance" in this study) in 2021, which emphasizes that "not all real-world data can produce real-world evidence when analyzed." The role of data cleaning is to process dirty data to regenerate real-world data that can be used to form real-world evidence. A standardized data cleaning process is critical to improving data quality. The guidance proposes essential requirements for real-world data governance; however, it does not outline the specific processes and approaches for data cleaning in detail. In this study, we outlined the current data-cleaning approaches for RWS and proposed a normal workflow for data cleaning to serve as a reference for applying such technologies in future studies.

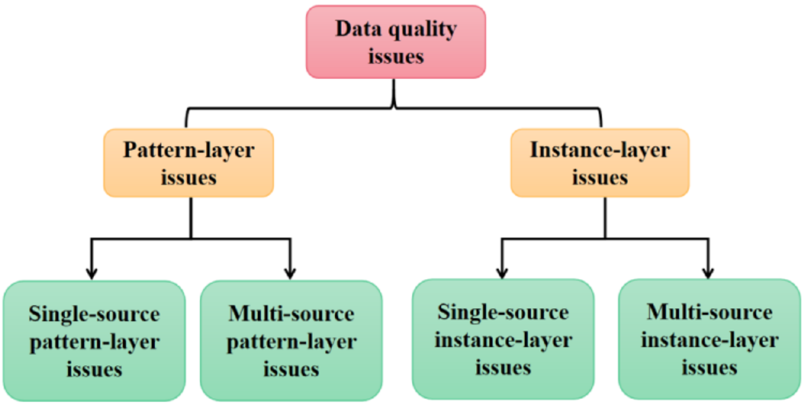
Impact of Data Cleaning on Data Quality

Data cleaning is the process of identifying and solving problems, which is crucial for the management of data quality [8]. The lack of an effective data cleaning process may result in a "garbage in and garbage out" scenario [8], adversely affecting the subsequent data analysis. In contrast, an effective data-cleaning process can transform dirty data into clean, reliable data that reflect real-world situations, providing researchers with more valuable information [9]. Therefore, data cleaning plays a decisive role in improving data quality.

Categorizing Issues With Data Quality

Data quality is the degree to which the accuracy, completeness, consistency, and timeliness of the data satisfy the expected needs of specific users. Issues with data quality can be categorized as either pattern-layer or instance-layer, depending on the level at which the issues are observed. Similarly, issues can be categorized as single-source or multi-source, depending on the data source. Therefore, issues with data quality are typically divided into 4 categories: single-source pattern-layer issues, multi-source pattern-layer issues, single-source instance-layer issues, and multi-source instance-layer issues (Figure 1) [10].

Figure 1. Classification of data quality issues.



Causes of Data Quality Issues

Pattern-layer issues originate from deficiencies in the system design. For single data sources, pattern-layer issues include a lack of integrity constraints and low-end architectural design. For multiple data sources, pattern-layer issues can also include structural and naming conflicts among the sources. Pattern-layer issues are not the main focus of data governance for RWS; however, many issues in the instance layer are caused by unresolved errors in the pattern layer.

At the instance layer, data issues mainly arise from human errors, which is a key focus of RWS on data governance. Common causes of data record exceptions at the single-source instance layer include data input errors, similar or duplicate records, and missing values. Input errors occur mostly during the process of case recording and are common in data sources that rely heavily on manual input, such as hospital information system data and individual health monitoring data from mobile devices. Similar or duplicate records may arise from operational errors during manual data entry. However, they may also arise when 2 cases with different levels of completeness are stored for the same patient during the same time period. This latter scenario is common when exporting data for different time periods, such as from January to June and June to December successively. Missing values may arise from technical errors in recording or deliberate concealment on the part of the patient (eg, refusal to provide relevant information). Alternatively, missing values can be caused by failures in data storage or error clearance resulting from equipment issues. In some cases, highly sensitive data may also be difficult to obtain (eg, medical insurance payment data).

In addition to all the problems that can arise at the instance layer for single sources, unique multisource issues in the instance layer include inconsistent data time and aggregation. Among them, similar or duplicate records resulting from identifying the same content as different objects (ie, use of different expressions) are the main problems.

Data cleaning can effectively address issues at the instance layer. To improve data quality, this step should be integrated into the processing pattern layer.

Data Cleaning

Definition of Data Cleaning

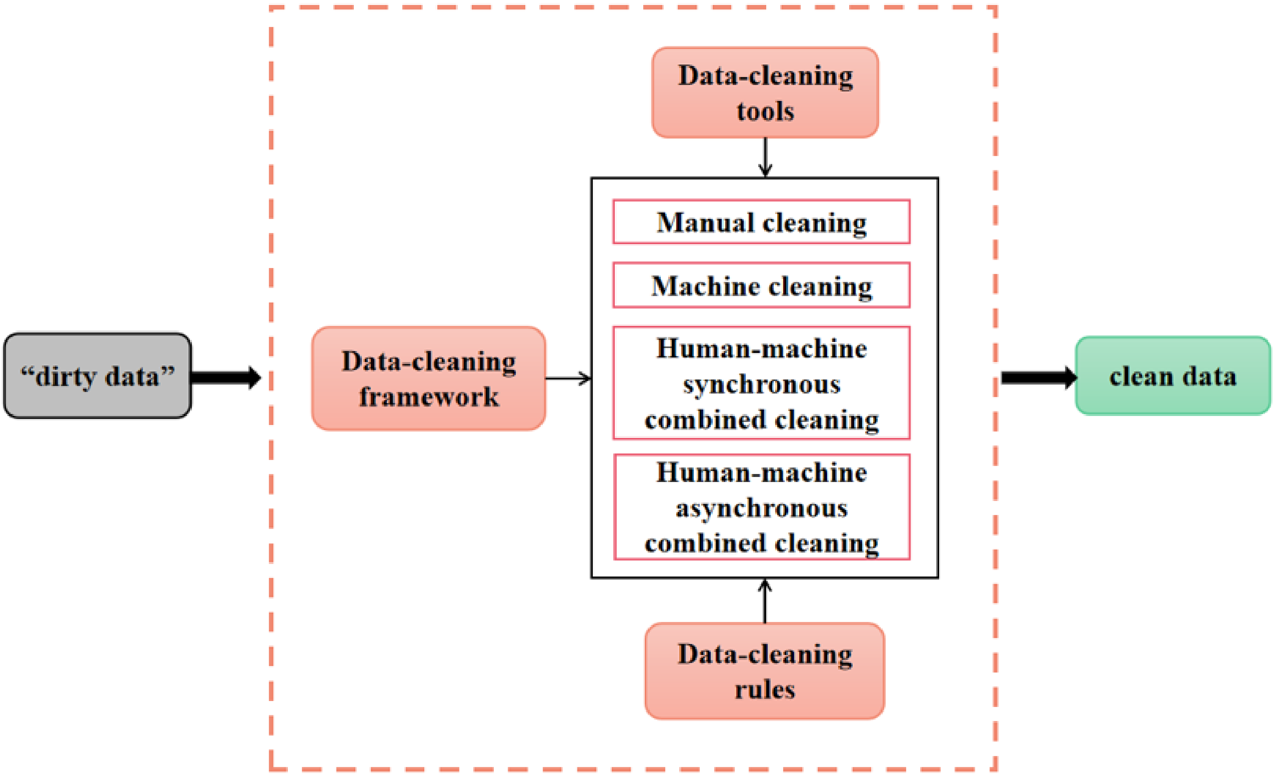
Data cleaning is a series of processes that streamline the database to remove duplicate records and convert the remaining information into standard-compliant data. More specifically, data cleaning involves the preprocessing of the extracted raw data, which includes elements such as the removal of duplicate or redundant data, logical verification of variable values, treatment of outliers, and processing of missing data. Thus, any operation performed to improve data quality can be classified as data cleaning. Data cleaning also encompasses all processes used to detect and repair irregularities in data collection and improve data quality. In the process of data cleaning, the corresponding cleaning rules can be formulated, and the data-cleaning framework and cleaning algorithms can be used to make the data-cleaning process easier and more efficient.

The guidance suggests that real-world data can be obtained prospectively and retrospectively, requiring data management and governance, respectively. Data cleaning is an element of data governance and is not required in the data management process. Therefore, data cleaning is generally suitable for real-world data collected retrospectively. The guidance divides data cleaning in RWS into the processing of duplicate, outlier, and missing data.

Basic Process for Data Cleaning

Relevant technical means, such as data mining, mathematical statistics, or predefined cleaning rules, are used to convert dirty data into data that meet quality requirements (Figure 2). Data cleaning is generally divided into 4 types: manual cleaning, machine cleaning, synchronous human-machine combined cleaning, and asynchronous human-machine combined cleaning [11]. The unprocessed source data are first collected directly from the database (ie, dirty data), following which the corresponding data cleaning rules are applied. The process can be streamlined using an appropriate data-cleaning framework and cleaning algorithms. Fully manual, fully automated, or combined strategies can be used until quality requirements are met.

Figure 2. Basic process of data cleaning.

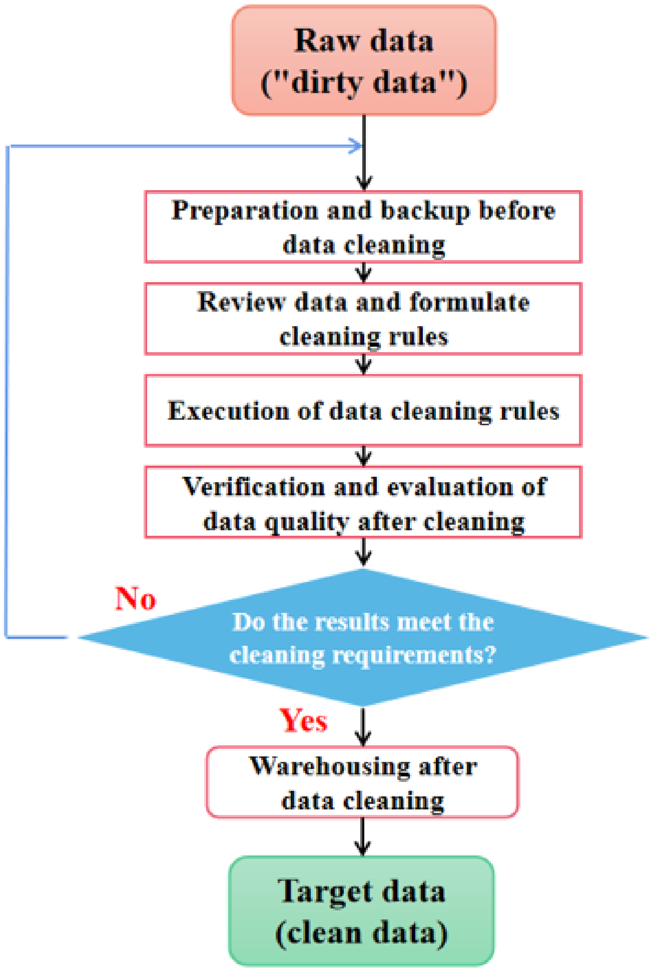


Despite its high accuracy, manual cleaning is only suitable for smaller data sets, given its time-consuming nature. In contrast, machine cleaning is more suitable for processing larger data sets since the process is completely automated. However, the cleaning plan and program must still be developed in advance, making later-stage maintenance difficult. In the synchronous human-machine strategy, problems that cannot be handled by the machine are manually addressed through an appropriate interface. This method is advantageous because it reduces the workload of manual cleaning while reducing the difficulty of designing the machine cleaning strategy. In principle, the asynchronous human–machine strategy is similar to its synchronous counterpart; however, when problems that cannot be handled by the machine are encountered, the issues are not addressed in real time. Instead, a problem report is generated, and cleaning proceeds to the next step. Thus, manual processing occurs after cleaning and is the method currently used by most cleaning software.

Normal Workflow for Data Cleaning

Depending on the task requirements, the data-cleaning workflow can be performed differently. The general data-cleaning process can be divided into 5 components (Figure 3).

Figure 3. Normal workflow for data cleaning.



Step 1: Back Up and Prepare the Raw Data for Cleaning

Data collected from different sources must be combined before further data governance and analysis can be performed. Therefore, it is necessary to unify the data types, formats, and key variable names in different databases before data cleaning. In addition, the original data must be backed up and archived before cleaning to prevent damage or loss of data during the cleaning process. This step is also crucial in cases requiring cleaning policy changes.

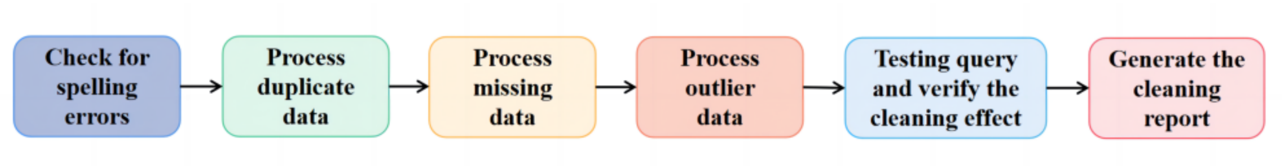
Step 2: Review the Data to Formulate Cleaning Rules

Appropriate cleaning methods (manual, machine, or combined) should be selected according to the size of the data set. After analyzing and summarizing the characteristics of the data source, the proposed cleaning algorithm and corresponding cleaning rules are formulated. Cleaning rules are divided into 3 categories: processing of missing, duplicate, and outlier data.

Step 3: Implement the Cleaning Rules

The execution of cleaning rules is the core step in the data cleaning process, and data processing can be performed in the following order: duplicate, missing, and outlier data (Figure 4). However, given the differences in professional fields and situational factors, adopting a common, unified standard for data cleaning is difficult. In addition, there are many types of data quality problems and complex situations, making a generalization based on categories difficult. Therefore, the corresponding cleaning rules must be formulated on a situational basis.

Figure 4. Execution of data cleaning rules: an example sequence.



Step 4: Verify and Evaluate the Quality of the Cleaned Data

Following data cleaning, the quality of the data should be assessed according to the cleaning report generated, and problems that could not be addressed by the machine must be handled manually. Evaluating the data quality will also enable the optimization of the program and algorithm to ensure that future processes yield data of sufficient quality. After redesigning the program based on these observations, the cleaning step should be repeated as needed until the requirements for analysis have been met.

Step 5: Warehouse After Data Cleaning

Following data cleaning, a new target database should be established for the cleaned data. While this aids in archiving and preservation, appropriate warehousing of the data can prevent the need for repeated cleaning work in the future.

Summary of Data Cleaning Methods for the Instance Layer

This section describes the methodology of the data cleaning methods, including the data sets, the 3 types of dirty data, and the corresponding data cleaning methods.

Data Set

In this study, we used a data set from a retrospective heart failure cohort in the Research Resource for Complex Physiologic Signals (PhysioNet) database [12]. This heart failure cohort retrospectively collected electronic medical records of 2008 hospitalized patients with heart failure from the Fourth People's Hospital of Zigong City, Sichuan Province, China, from December 2016 to June 2019. The identification of hospitalized patients with heart failure was based on the International Classification of Diseases-9 code. Furthermore, the diagnostic criteria followed the 2016 European Society of Cardiology Heart Failure Diagnosis and Treatment Guidelines. The partial information contained in the data set is presented in Figure S1 in Multimedia Appendix 1, with 167 variables (n=167), including 2008 records (N=2008), and saved as a CSV file.

To provide a more intuitive demonstration of the results in the following examples, we added 30 records as "duplicate data" in the heart failure data set and manually adjusted the "systolic blood pressure" values in 11 records as "abnormal data." According to the admission way (column E), patients with heart failure were divided into 2 groups: the emergency and nonemergency groups, and the urea values (column BO) were analyzed and compared between these 2 groups. There are "missing data" in the urea values of the 2 groups (Figure S2 in Multimedia Appendix 1):

Processing of Duplicate Data

Methods for detecting duplicate data can be divided into record-based and field-based methods. Record-based duplicate detection algorithms include the N-grams, sorted-neighborhood method (SNM), clustering, and most probable number (MPN) algorithms [13-15]. Field-based repeat detection algorithms include the cosine similarity function [16] and Levenshtein distance algorithms [17]. The main processes involved in duplicate data cleaning are as follows:

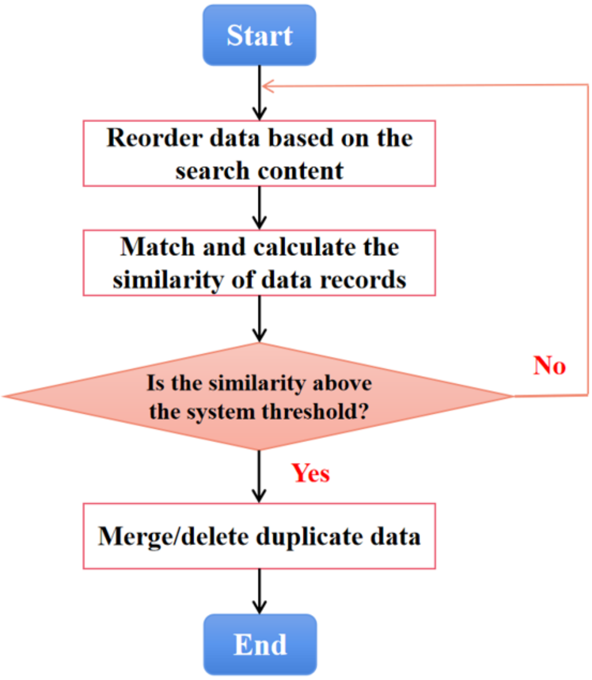
Step 1: Analyze the attribute segments of the source database, limit the key search values of the attributes (eg, name, patient ID, and date of treatment), and sort the data in the source database from the bottom to the top or top to bottom according to the key search values.

Step 2: Scan all data records according to the order of arrangement, compare the adjacent data, and calculate the similarity of the matching data records. The duplicate data retrieval code for the sample data set is presented in Figure S3 in Multimedia Appendix 1.

Step 3: Deduplicate or merge the duplicate data. When the similarity value of adjacent data is higher than the threshold defined by the system, the continuous data records are identified as similar or duplicate data. The duplicate data retrieval results of the sample data set are presented in Figure S4 in Multimedia Appendix 1. These data should be deduplicated or merged. Similarly, when the similarity value is below the threshold defined by the system, scanning should be continued, and steps 2 and 3 should be repeated as necessary.

Step 4: After testing all data records, generate a report and archive the data before and after it. The workflow for cleaning duplicate data is shown in Figure 5.

Figure 5. Cleaning workflow for duplicate data.



Processing of Missing Data

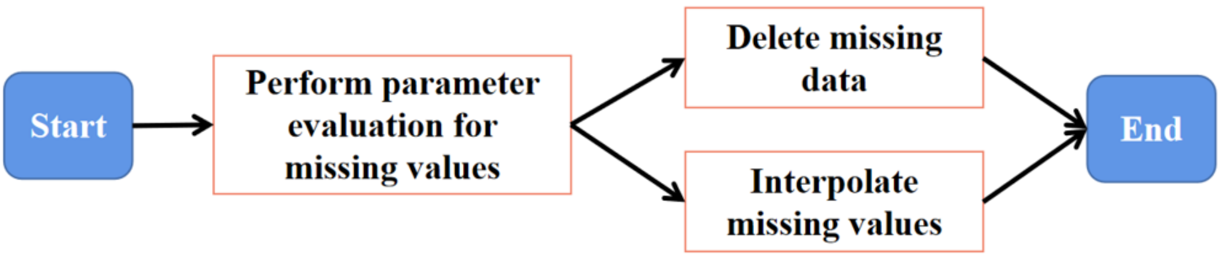
While deletion is initially considered for missing values, this only applies if the data set is large, and the proportion of missing data is not large. If the amount of data is not sufficiently large, directly deleting missing values will lead to data loss, resulting in the deletion of many useful statistics. Feasible methods for addressing missing values include using missing-value imputation technology for repair. Commonly used methods for imputation include mean imputation, mode imputation, minimum imputation, regression imputation, and maximum likelihood estimation [18-21], and using a Bayesian or decision tree classifier [22] to model the missing value imputation as a classification problem. The main processes involved in cleaning missing data are as follows:

Step 1: Perform parameter estimation for missing values in the source data and select the deletion method or imputation method according to the proportion of missing values. The missing data retrieval code for the sample data set is presented in Figure S5 in Multimedia Appendix 1.

Step 2: Fill in the missing data according to the data-filling algorithm. The missing data retrieval and interpolation results of the sample data set are presented in Figure S6 in Multimedia Appendix 1. For the convenience of demonstration, the mean interpolation method was chosen, and specific problems should be analyzed in practical application.

Step 3: Output and archive the complete data. The workflow for cleaning missing data is shown in Figure 6.

Figure 6. Cleaning workflow for missing data.



Processing of Outlier Data

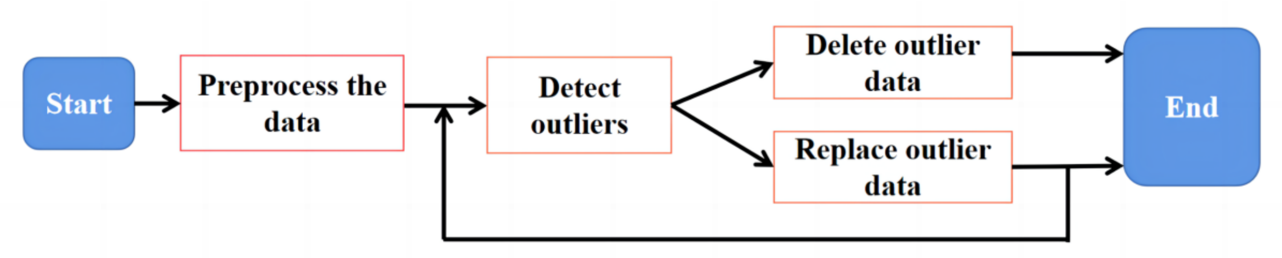
An outlier is a value that does not conform to attribute semantics. There are 2 methods for handling outlier data: deletion and replacement. However, the appropriate methods should be selected based on the nature of the data. If the nature of the outlier data is unsuitable for replacement, such as age data, outlier analysis can be used to detect and delete outliers. Outlier detection algorithms mainly include cluster-based, statistical model-based, density-based, and proximity-based algorithms [23]. If the nature of the abnormal data is suitable for replacement, the regression method or mean smoothing method can be used to replace the abnormal values. The regression method is applicable to data conforming to a linear trend, while the mean smoothing method is more effective in cleaning data with sinusoidal time-series characteristics [24]. The main processes involved in cleaning outlier data are as follows:

Step 1: Convert the source data into the data format required for detection and conduct data preprocessing. The exception data retrieval code for the sample data set is presented in Figure S7 in Multimedia Appendix 1.

Step 2: Perform outlier detection on the data after preprocessing. The abnormal data retrieval results of the sample data set are presented in Figures S8 and S9 in Multimedia Appendix 1. If the nature of the outlier data is not suitable for replacement, delete the outliers. If the nature of the outlier data is suitable for replacement, use the regression or mean smoothing method to replace the outliers. Often, the repaired data can lead to new data exceptions, making it necessary to repeat steps 1 and 2 until the requirements are met.

Step 3: Restore the repaired data to its original format and perform archiving. The workflow for cleaning outlier data is shown in Figure 7.

Figure 7. Cleaning workflow for outlier data.



Data Cleaning Tools

Oni et al [25] reported 4 tools commonly used in the data cleaning industry: Data Wrangler, OpenRefine, Python, and R. They explained that these tools are the most popular tools for data cleaning in RWS. OpenRefine, R, and Python are all open-source tools, making them easy to access and use. Data Wrangler is a commercial tool, but there is a community version that efficiently cleans up data. The characteristics of these tools are described below and presented in Table 1.

Table 1. Comparison of Data Wrangler, Python, R, and OpenRefine.

Criteria	Data Wrangler	Python	R	OpenRefine
Import format	Excel, CSV, and text	All	All	Excel, CSV, TSV[a], XML, JSON, and RDF[b]
Factors affecting the performance time	Data size and user choice	User programming skill level	User programming skill level	Data size and data format
Output format	CSV, JSON, and TDE[c]	Any format	Any format	Excel, TSV, CSV, and HTML table
Skill level	Basic level	Advanced level	Advanced level	Basic or intermediate level
Running platform	Windows and Mac	All	All	All
Accuracy	Depends on the specific data quality issues (eg, missing values)	Depends on the user's programming skill level	Depends on the user's programming skill level	Depends on the specific data quality issues (eg, missing values)
Possibility to embedded	No	Yes	Yes	No, but code is available
Data set processing scale	Big data	Big data	Big data	Up to 5000 records
Graphic user interface	Yes	No	No	Yes

[a]TSV: tab separated value.

[b]RDF: resource description framework.

[c]TDE: tableau data extract.

Data Wrangler

Data Wrangler is a web-based data cleaning and reorganization project developed by Stanford University [26]. It is a web-based data cleaning tool, mainly used to remove invalid data and organize data into user-required formats. Several data collations can be done in Data Wrangler with a simple click. It also lists the history of data modifications, making it extremely convenient for users to view past modifications and undo a modification operation. Data Wrangler can process data in 2 ways: users can either paste the data into its web interface or use the web interface to export any data operations to Python code and process them.

Advantages of Data Wrangler are that it has column and grid views, uses natural language to describe transformations, supports data visualization and every step of data cleaning, and supports large-scale editing. Disadvantages are that the free version of Data Wrangler provides only limited functionality and consumes a large amount of memory.

Python

Python is a concise, easy-to-read, and extensible data-cleaning tool [27]. Currently, Numpy and PANDAS (Python Data Analysis Library) are the most used mainstream modules in Python for data cleaning. The PANDAS module is mainly used for data analysis, of which data cleaning is a part. The Numpy module has a powerful N-dimensional array object, and vectorization operations make data processing efficient and helpful in cleaning large data sets.

Advantages of Python are that it is easy to embed into other tools and applications, and users can customize solutions based on their needs. Disadvantages are that it requires users to have advanced programming skills, learn how to use many modules in Python, and understand the required steps during the cleaning process in advance, making it difficult to implement.

R

R is the language and operating environment used for statistical calculations, data analysis, and graphics [28]. R is a free, open-source software belonging to the GNU's Not Unix (GNU) system. It can provide some integrated statistical tools. More importantly, it can provide various mathematical and statistical calculation functions, allowing users to flexibly analyze data and create new statistical calculation methods that meet their needs. R has a set of tools that can effectively and comprehensively clean data. The R environment can read data in multiple formats and process these files. R provides sufficient visualization tools. During the cleaning process, visualization of data at each stage is useful.

Advantages of R are that it supports the visualization of data and each step of data cleaning, making it more suitable for analyzing statistical data. Disadvantages of R are that it is not a good choice for projects outside data science. Users must understand the required steps during the cleaning process in advance, making it difficult to implement.

OpenRefine

OpenRefine is a web-based ,independent, open-source application with various functions such as data portrait, cleaning, and conversion [29]. It can perform visual manipulations on data. It is similar to traditional Microsoft Excel software. However, it works like a database, as it does not deal with individual cells but rather with columns and fields. OpenRefine, formerly known as Google Refine, is a tool for cleaning, reshaping, and editing bulk, unstructured, and cluttered data. OpenRefine is a desktop application that opens as a local web server in a browser. Since it is an open-source project, its code can be reused in other projects. OpenRefine performs cleanup tasks by filtering and faceting, and then converts the data into a more structured format.

Advantages of OpenRefine are that it is a desktop application that does not require networking, making data sets more difficult to tamper with and relatively secure. It can be easily operated and has powerful functions for converting data. Users can use its facet function to filter data into subsets. Disadvantages include a limit of 5000 records, making OpenRefine not suitable for processing large data sets. It assumes that data is organized in a tabular format with limited operations and an unfriendly user interface. In addition, Google has removed support for the tool.

Documentation and Reporting

The Guidelines for Real-World Evidence to Support Drug Research and Development and Review (Trial) of the National Medical Products Administration of China (No 1 of 2020) stipulated that "transparency and reproducibility of evidence" should be achieved in the process of translating real-world data into real-world evidence, noting that proper documentation retention is the basis for ensuring transparency and reproducibility. We recommend that the data cleaning plan be stipulated in the RWS data governance plan, which should include personnel requirements, previous expectations for screening suspicious data, diagnostic procedures for identifying errors in the source data, cleaning tools, and decision rules to be applied in the cleaning phase.

Additionally, appropriate documentation should be provided at each of the following points: (1) precleaning (the raw data stage); (2) cleaning operation (during this stage, documentation should include differential markers of suspicious feature types, diagnostic information related to the type of dirty data, application algorithms and operational steps for data editing, and corresponding cleaning reports generated after cleaning is complete; simultaneously, the modification date must be marked for each operation, and the information of the relevant personnel involved in the modification must be saved); (3) the retention stage (after cleaning the data).

Recommendations for Data Cleaning

Most research projects do not formulate data-cleaning plans in advance. Analyses performed without complete cleaning of the dirty data will lead to biased results, and identifying the causes of any deviations from scratch will further delay the progress of the work. As the diversity of data sources increases the difficulty and workload of data cleaning, we recommend the following strategy.

First, formulate the cleaning plan in advance. As mentioned above, the results of statistical analyses are closely related to the cleanliness of the data. Data cleaning plans should be formulated in advance to ensure sufficient time and technical guidance for data cleaning.

Second, cultivate medical and computer talent. While analyzing real-world data, many medical researchers find that they do not understand computer programming. Conversely, many computer programmers do not have much medical expertise, resulting in poor communication between the two sides and affecting the development of data-cleaning strategies. Therefore, it is necessary to cultivate a group with compound talents who understand both medical statistics and computer applications.

Third, strengthen the computer skills training required for data cleaning. Hospitals and data companies should work together to organize and implement skills training for data cleaning in a timely manner. In addition, medical researchers and computer programmers should participate simultaneously to acquire professional knowledge from each other. Machine-based and manual methods can be selected to improve work efficiency when adopting combined human-machine cleaning strategies.

Fourth, establish a unified data governance and management platform. Researchers should fully use modern technical means to realize the collection, review, governance, and management of RWS data. Moreover, project researchers should perform unified management and maintenance of platform data.

Conclusions

Real-world data are large-scale with low value density. The data source yields dirty data, plagued by issues such as duplication, missing values, and outliers owing to various reasons. Analyses based on such data can severely reduce the efficiency of data use and negatively affect the quality of decision-making. Data cleaning technology can improve data quality and provides more accurate and realistic target data than the source data, which can then be used to support data consumers in making appropriate decisions. The data cleaning principles and workflows discussed in this study may aid in developing standardized methods for data cleaning in RWS.

Acknowledgments

The authors would like to acknowledge support from the foundation of the Science and Technology Innovation Project of the Chinese Academy of Traditional Chinese Medicine (Research on Key Technologies for Clinical Evaluation of New Chinese Medicines under the Three Combination Evaluation Evidence System), with the subject number CI2021A04701.

Authors' Contributions

RG and QL put forward the writing direction of this paper, MG was responsible for the idea and writing the first draft of this paper, and other authors participated in the discussion during the preparation of the paper. MZ, YC, and XJ were responsible for demonstrating the practical part of data cleaning. All authors contributed to the paper and approved the submitted version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary figures 1-9.
[DOCX File , 1148 KB - ijmr_v12i1e44310_app1.docx]

References

1. Bhide A, Shah PS, Acharya G. A simplified guide to randomized controlled trials. Acta Obstet Gynecol Scand 2018;97(4):380-387 [FREE Full text] [doi: 10.1111/aogs.13309] [Medline: 29377058]
2. Bothwell LE, Podolsky SH. The emergence of the randomized, controlled trial. N Engl J Med 2016;375(6):501-504. [doi: 10.1056/NEJMp1604635] [Medline: 27509097]
3. Kaplan NM, Sproul LE, Mulcahy WS. Large prospective study of ramipril in patients with hypertension. CARE investigators. Clin Ther 1993;15(5):810-818. [Medline: 8269447]
4. Schalk E, Hentrich M. Real-world data. Dtsch Arztebl Int 2022;119(8):134 [FREE Full text] [doi: 10.3238/arztebl.m2022.0035] [Medline: 35506295]
5. Corrigan-Curay J, Sacks L, Woodcock J. Real-world evidence and real-world data for evaluating drug safety and effectiveness. JAMA 2018;320(9):867-868. [doi: 10.1001/jama.2018.10136] [Medline: 30105359]
6. Page L. The high cost of dirty data. Mater Manag Health Care 2005 Nov;14(11):22-25. [Medline: 16396144]
7. The guidance on real-world data for the generation of real-world evidence (Trial). The China Center for Drug Evaluation. 2021. URL: https://www.mediecogroup.com/zhuanlan/lessons/863/
8. Benevento E, Aloini D, van der Aalst WMP. How can interactive process discovery address data quality issues in real business settings? Evidence from a case study in healthcare. J Biomed Inform 2022;130:104083 [FREE Full text] [doi: 10.1016/j.jbi.2022.104083] [Medline: 35504544]
9. Bogani R, Theodorou A, Arnaboldi L, Wortham RH. Garbage in, toxic data out: a proposal for ethical artificial intelligence sustainability impact statements. AI Ethics 2022:1-8 [FREE Full text] [doi: 10.1007/s43681-022-00221-0] [Medline: 36281314]
10. Rahm E, Do HH. Data cleaning: problems and current approaches. IEEE Data Eng 2000;23(4):3-13 [FREE Full text]
11. Gesicho MB, Were MC, Babic A. Data cleaning process for HIV-indicator data extracted from DHIS2 national reporting system: a case study of Kenya. BMC Med Inform Decis Mak 2020;20(1):293 [FREE Full text] [doi: 10.1186/s12911-020-01315-7] [Medline: 33187520]

12. Zhang Z, Cao L, Zhao Y. Hospitalized patients with heart failure: integrating electronic healthcare records and external outcome data. PhysioNet 2000;101(23):e215-e220 [FREE Full text] [doi: 10.13026/8A9E-W734]
13. Nawab RMA, Stevenson M, Clough P. Comparing medline citations using modified N-grams. J Am Med Inform Assoc 2014;21(1):105-110 [FREE Full text] [doi: 10.1136/amiajnl-2012-001552] [Medline: 23715801]
14. van Wunnik BPW, Visschers RGJ, van Asselt ADI, Baeten CGMI. Cost-effectiveness analysis of sacral neuromodulation for faecal incontinence in the Netherlands. Colorectal Dis 2012;14(12):e807-e814 [FREE Full text] [doi: 10.1111/codi.12002] [Medline: 22943485]
15. Jahan M, Hasan M. A robust fuzzy approach for gene expression data clustering. Soft Comput 2021;25(23):14583-14596. [doi: 10.1007/s00500-021-06397-7]
16. Akbaş CE, Günay O, Taşdemir K, Çetin AE. Energy efficient cosine similarity measures according to a convex cost function. Signal Image Video Process 2017;11(2):349-356. [doi: 10.1007/s11760-016-0949-7]
17. Berger B, Waterman MS, Yu YW. Levenshtein distance, sequence comparison and biological database search. IEEE Trans Inf Theory 2021;67(6):3287-3294 [FREE Full text] [doi: 10.1109/tit.2020.2996543] [Medline: 34257466]
18. Raja PS, Thangavel KJSC. Missing value imputation using unsupervised machine learning techniques. Soft Comput 2020;24(6):4361-4392. [doi: 10.1007/s00500-019-04199-6]
19. Silva-Ramírez EL, Pino-Mejías R, López-Coello M. Single imputation with multilayer perceptron and multiple imputation combining multilayer perceptron and k-nearest neighbours for monotone patterns. Applied Soft Computing 2015;29:65-74 [FREE Full text] [doi: 10.1016/j.asoc.2014.09.052]
20. Beesley LJ, Bondarenko I, Elliot MR, Kurian AW, Katz SJ, Taylor JM. Multiple imputation with missing data indicators. Stat Methods Med Res 2021;30(12):2685-2700 [FREE Full text] [doi: 10.1177/09622802211047346] [Medline: 34643465]
21. Li M, Liu X. Maximum likelihood least squares based iterative estimation for a class of bilinear systems using the data filtering technique. Int J Control Autom Syst 2020;18(6):1581-1592. [doi: 10.1007/s12555-019-0191-5]
22. Duan Z, Wang L, Sun M. Efficient heuristics for learning Bayesian network from labeled and unlabeled data. Intell Data Anal 2020;24(2):385-408. [doi: 10.3233/ida-194509]
23. Shao M, Qi D, Xue H. Big data outlier detection model based on improved density peak algorithm. J Intell Fuzzy Syst 2021;40(4):6185-6194. [doi: 10.3233/jifs-189456]
24. Yang J, Zhu H, Choi T, Cox DD. Smoothing and mean-covariance estimation of functional data with a bayesian hierarchical model. Bayesian Anal 2016;11(3):649-670 [FREE Full text] [doi: 10.1214/15-ba967] [Medline: 34457106]
25. Oni S, Chen Z, Hoban S, Jademi O. A comparative study of data cleaning tools. Int J Data Warehous 2019;15(4):48-65. [doi: 10.4018/ijdwm.2019100103]
26. Kandel S, Paepcke A, Hellerstein J, Heer J. Wrangler: interactive visual specification of data transformation scripts. 2011 Presented at: CHI '11: CHI Conference on Human Factors in Computing Systems; May 7-12, 2011; Vancouver, BC. [doi: 10.1145/1978942.1979444]
27. Raschka S, Patterson J, Nolet C. Machine learning in python: main developments and technology trends in data science, machine learning, and artificial intelligence. Information 2020;11(4):193 [FREE Full text] [doi: 10.3390/info11040193]
28. Fox J, Leanage A. R and the journal of statistical software. J Stat Soft 2016;73(2):1-13 [FREE Full text] [doi: 10.18637/jss.v073.i02]
29. Carlson S, Seely A. Using OpenRefine's reconciliation to validate local authority headings. Cat Classif Q 2017;55(1):1-11. [doi: 10.1080/01639374.2016.1245693]

Abbreviations

GNU: GNU's Not Unix
MPN: most probable number
PANDAS: Python Data Analysis Library
RCT: randomized controlled trial
RDF: resource description framework
RWS: real-world research
SNM: sorted-neighborhood method
TDE: tableau data extract
TSV: tab separated value

Edited by A Mavragani; submitted 15.11.22; peer-reviewed by R Alexander, HJ Kim; comments to author 29.12.22; revised version received 30.01.23; accepted 09.05.23; published 21.09.23.

Please cite as:
Guo M, Wang Y, Yang Q, Li R, Zhao Y, Li C, Zhu M, Cui Y, Jiang X, Sheng S, Li Q, Gao R
Normal Workflow and Key Strategies for Data Cleaning Toward Real-World Data: Viewpoint
Interact J Med Res 2023;12:e44310
URL: https://www.i-jmr.org/2023/1/e44310
doi:10.2196/44310
PMID:37733421

Viewpoint

Understanding Loneliness in Younger People: Review of the Opportunities and Challenges for Loneliness Interventions

Hurmat Ali Shah[1], PhD; Mowafa Househ[1], PhD

Hamad bin Khalifa University, Doha, Qatar

Corresponding Author:
Mowafa Househ, PhD
Hamad bin Khalifa University
Education City - Gate 8
Ar Rayyan
Doha
Qatar
Phone: 974 44547278
Email: mhouseh@hbku.edu.qa

Abstract

Loneliness affects the quality of life of people all around the world. Loneliness is also shown to be directly associated with mental health issues and is often the cause of mental health problems. It is also shown to increase the risk of heart diseases and other physical illnesses. Loneliness is studied both from the social and medical sciences perspectives. There are also interventions on the basis of health informatics, information and communication technologies (ICTs), social media, and other technological solutions. In the literature, loneliness is studied from various angles and perspectives ranging from biological to socioeconomical and through anthropological understandings of technology. From the ICT and technological sides, there are multiple reviews studying the effectiveness of intervention strategies and solutions. However, there is a lack of a comprehensive review on loneliness that engulfs the psychological, social, and technological studies of loneliness. From the perspective of loneliness informatics (ie, the application of health informatics practices and tools), it is important to understand the psychological and biological basis of loneliness. When it comes to technological interventions to fight off loneliness, the majority of interventions focus on older people. While loneliness is highest among older people, theoretical and demographical studies of loneliness give a U-shaped distribution age-wise to loneliness; that is, younger people and older people are the demographics most affected by loneliness. But the strategies and interventions designed for older people cannot be directly applied to younger people. We present the dynamics of loneliness in younger people and also provide an overview of the technological interventions for loneliness in younger people. This paper presents an approach wherein the studies carried out from the perspectives of digital health and informatics are discussed in detail. A comprehensive overview of the understanding of loneliness and the study of the overall field of tools and strategies of loneliness informatics was carried out. The need to study loneliness in younger people is addressed and particular digital solutions and interventions developed for younger people are presented. This paper can be used to overcome the challenges of technological gaps in the studies and strategies developed for loneliness. The findings of this study show that the majority of interventions and reviews are focused on older people, with ICT-based and social media–based interventions showing promise for countering the effects of loneliness. There are new technologies, such as conversational agents and robots, which are tailored to the particular needs of younger people. This literature review suggests that the digital solutions developed to overcome loneliness can benefit people, and younger people in particular, more if they are made interactive in order to retain users.

(Interact J Med Res 2023;12:e45197) doi:10.2196/45197

KEYWORDS

health informatics; loneliness informatics; loneliness theory; health effects; loneliness interventions; information and communication technology; ICT-based interventions; social-media–based interventions; social media; ICT; lonely; loneliness; social isolation; mental health; psychological

Introduction

Loneliness is a global health epidemic that affects a significant number of global populations. In the United States, an estimated 17% of adults aged 18 to 70 years report loneliness. Monetary loss as a result of loneliness is estimated to be between US$8074.80 and US$12,0777.70 per person per year in the United Kingdom [1]. The monetary cost of lost days and loss in productivity is estimated to be US$3.14 billion per year for employees in the United Kingdom. Loneliness is also linked to a 30% increase in heart disease, stroke, dementia, depression, and anxiety [2-4]. Loneliness and isolation are interlinked yet separate concepts. Loneliness is defined as the subjective difference between a person's desired and actual social contact and relations [5]. Loneliness can be subjective, while isolation, on the other hand, is a social phenomenon in which there is actual absence of social engagement and contact both with the immediate family and larger community [6]. This can be acutely the case for older adults, where one-third and one-quarter of them will experience feelings of loneliness and isolation, respectively [7,8].

Social isolation results in the loss of self-esteem and self-confidence and hence the ability to form meaningful social relationships. Stigma may be particularly significant for certain age groups, such as young people [9]. Studies have linked loneliness and social isolation to numerous determinants, such as possible psychological, social, neurocognitive, and genetic mechanisms, as well as relationships with community and social factors [9]. What makes loneliness hard to study as a category is that it can be viewed as a transdiagnostic construct that can occur alongside, as well as cause and predict, a variety of mental health conditions [10]. Transient loneliness results in emotional distress, which can be caused by social disconnection but can be coped with. The problem arises if loneliness persists in the long-term, at which point it can result in the altering of neurobiological and behavioral patterns. Cacioppo and Hawkley [11] theorize about a self-enforcing loop of chronic loneliness, leading to an increase in hypervigilance and cognitive bias toward social threat, hence eliciting hostile behavior toward social interaction.

There have been several systematic and scoping reviews done to understand loneliness. Similarly, from a health informatics perspective, there have been multiple systematic reviews of the connection between loneliness and mental health [6,8,12-14]. From the health informatics side, there also have been studies that deal with the application of technology-based interventions to cope with loneliness. There are also scoping reviews to find the effectiveness of technology-based intervention strategies for loneliness [14]. However, there are limitations to the scoping reviews done on loneliness, both theoretical (ie, psychological aspect) and technical (ie, from a health informatics perspective). The first limitation is that almost all of them focus on intervention strategies for older people. Second, from a health informatics perspective, a comprehensive study of technology-based interventions has not been carried such that a general overview of different technologies used to cope with loneliness can be drawn.

Young people are particularly prone to being lonely. The reasons for this are multifaceted. The first context is a developmental one, wherein adolescent social interaction is key to identity formation and individuation from family [15,16]. Any negative experience as to social interaction can result in self-imposed isolation, and thus loneliness. The other context is that of increased risky behavior, which can be exacerbated by the feeling of loneliness. Studies have shown that loneliness in young people results in increased risks of smoking, taking drugs, and consuming alcohol [17]. Studies have also shown that young people are particularly prone to mental health problems, with 1 in every 8 young people aged 5 to 19 years having a mental health issue [18]. Because loneliness and mental health problems are related in adults and 75% of all mental health problems emerge before the age of 24 years, it is important to study loneliness in younger people and overcome the condition earlier to avoid consequent mental health problems.

Finally, the COVID-19 pandemic has had a greater impact on the feeling of loneliness among younger people than the general population, with 50.8% of people aged 16 to 24 years reporting feeling lonely as compared to the general population ratio of 30.9% [17]. Other research has shown an increase in feelings of anger and perceived stress during the COVID-19 pandemic as compared to prepandemic levels [19].

These factors make studying loneliness in young people urgent. A review is important to find out what technology-based interventions are available for loneliness among young people and what kind of interventions will be more effective than others in reducing loneliness among young people. This paper was written after consulting a wide range of resources on loneliness, from theoretical understandings of loneliness to technological interventions for overcoming loneliness. The resources and papers were found using Google Scholar. No particular inclusion methodology was used to review or study papers other than to find papers that gave an introduction to both loneliness on a theoretical level and for technological interventions. The available literature was further studied to identify different categories of technological interventions to overcome loneliness. The survey of available literature revealed that a relatively smaller number of technological interventions focus on overcoming loneliness in younger people than in older people. Summarily, there are 3 types of technology-based interventions in the literature on loneliness in young people. The first type deals with social media–based intervention strategies for fighting loneliness in young people. The second type deals with broader applications of information and communication technology (ICT) ranging from videoconferencing, psychotherapy, and remote counselling. The third type of literature available in loneliness informatics is about exploring the effects of the prevalent use of social media in young people and its relationship with loneliness.

The major contributions of this literature review are as follows:

- Situating the literature on loneliness in young people in loneliness informatics through a short overview of the literature on loneliness across all age groups.
- Comprehensively reviewing technology-based intervention strategies for reducing loneliness in younger people.
- Highlighting gaps in the current literature on loneliness in young people.
- Exploring and proposing methods and technology tools that can be used to build more effective tools for reducing loneliness in younger people.

Theoretical Understanding of Loneliness

While we provided some definitions of loneliness in the Introduction section, it is imperative for the field of loneliness informatics to be well-connected with the theoretical grounding of loneliness. For this purpose, this section will provide a detailed theoretical understanding of loneliness through different psychological, biological, and social theories.

While loneliness can have positive connotations in philosophical contexts and in some evolutionary theories, the core elements of the concept of loneliness are unwelcome feelings associated with a lack of companionship, either in quantity or quality. Cognitive processes determine how a person processes the feeling of loneliness. Taking this, Paplau and Perlman [20] formally defined loneliness as "the unpleasant experience that occurs when a person's network of social relationships is deficient in some important way, quantitively or qualitatively." Loneliness is a multidimensional phenomenon with 3 clear, major components [21]. The first component is deprivation, which can be called the core component. The second component is the temporal dimension, that is, whether the feeling of loneliness is perceived as changeable or hopeless. The third component is that of different types of emotional aspects, such as feelings of sorrow, sadness, guilt, and shame. The deprivation and emotional components of this understanding of loneliness point to the personal and social determinants of loneliness. Thus, loneliness can be the result of emotional isolation caused by the absence of attachment and social isolation caused by the absence of community.

Loneliness is understood by the following three major theoretical frameworks [22]:

1. Social needs perspective: This theory claims a direct relationship between the objective social reality of relationship deficit and the subjective experience of loneliness. When a relationship does not satisfy the desired social needs, loneliness arises. Desired social needs change over time and the adherents of this theory claim that the experiences of loneliness are expected to change over the lifespan [20].
2. Cognitive discrepancy model: This model focuses on a person's subjective evaluation of the fulfillment relationships bring rather than the social needs fulfilled by the relationships. People tend to judge the satisfaction their relationships bring on their own internal scale. Through this scale, comparison is made with other relationships in their circle or through a perceived sense of other people's relationships. The cognitive discrepancy arises when the actual relationships do not meet qualitatively with the perceived standard as judged by a subjective internal scale.
3. Evolutionary theory of loneliness (ETL): The previous 2 theories emphasize the social environment and its relationship to the subjective feeling of loneliness. The ETL, on the other hand, focuses both on the social environment and genes in developing feelings of loneliness. According to the ETL, the feeling of being lonely even in the presence of others is a biological warning system, which is found across all species, that signals that the present relationships one finds themselves in are either damaged or threatened. The negative feelings associated with being lonely motivate one to replace the negative bonding or relationships. The ETL states that the feeling of loneliness promotes emphasis on vigilance for social threats and an increased concern for one's self-interests. This behavior is influenced both by environmental factors and genetic inheritance. Furthermore, according to the ETL, the onset of loneliness is not controlled by one gene but by multiple genes. The role of these genes in determining loneliness varies from one person to the other. For the social component, the ETL claims that the expression of loneliness genes depends on the social environment. Therefore, the variability in feelings of loneliness is determined by the interaction of genes and the social environment. The ETL has multiple components, but for the study of loneliness informatics and technology-based intervention, one component is important. Reaffiliation motive, which is part of the ETL, states that aversive feelings of social isolation caused by loneliness motivate individuals to reconnect with other people [23].

Situating Loneliness Informatics for Young People: Context for Technology-Based Interventions

This section will give an in-depth literature review of loneliness in younger people and technology-based interventions for loneliness in general. Loneliness in younger people has different social and generational dynamics. Therefore, loneliness in younger people must be understood from its particular dynamics in order to design effective digital interventions. In the first subsection below, the theoretical understanding of loneliness in younger people is given, while in the second subsection, an overall picture of loneliness informatics is given with the aim of situating loneliness informatics for younger people.

Understanding Loneliness in Younger People

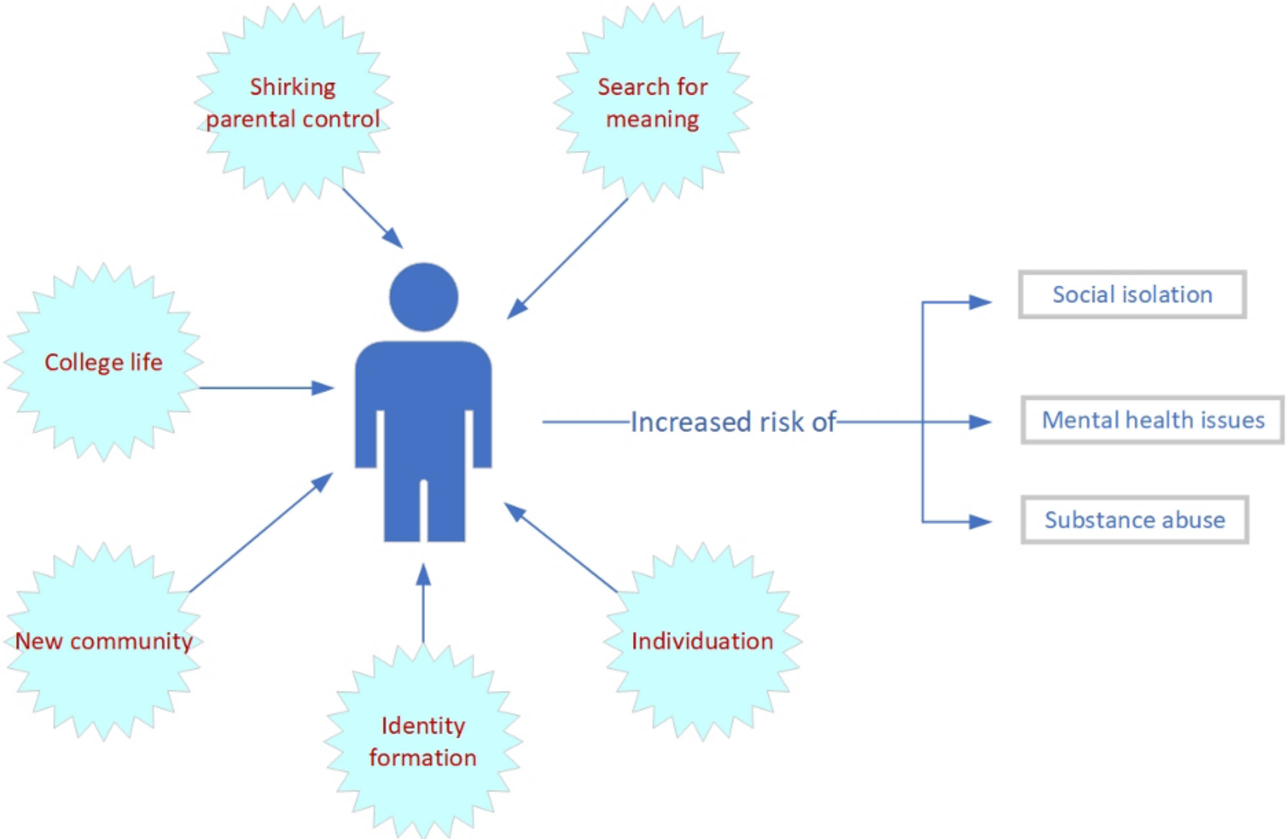
Loneliness in younger people has yet to be explored and tackled proportionately to the prevalence of loneliness in younger people. Population studies of loneliness find a U-shaped age distribution of loneliness; that is, the rate of loneliness is higher in older and younger people. The epidemiological work carried out on loneliness and its health impacts focus more on older people [24-26]. The findings as well as theoretical underpinnings of the works that focus on older people cannot be directly linked and applied to younger people as the dynamics of loneliness vary by age and cultural contexts and groups. Weiss [27] suggested that the primary function of young adulthood is to shirk off parental attachments and to make life and social connections according to one's own subjective perceptions and interests. Young adulthood is also the age when the new life of college begins, which also brings with it the experience of social distance from the established social networks.

The reported evidence about loneliness in younger people suggests an increased occurrence of loneliness in adolescence years. Younger people and children aged 10 to 24 years report loneliness as a feeling of isolation and a sense of exclusion and disconnection from their social context [28]. It was also noted by the UK Office of National Statistics that people aged 16 to 24 years were the loneliest of all age groups [28]. It was shown by Qualter et al [29] that chronic loneliness predicts adolescent depression. Loneliness affects the quality of life particularly because of the stigma associated with seeking formal or informal help [30]. The loneliness reported by younger people is even higher than that reported by older people, as reported by some studies. A web-based survey found that out of 55,000 people, 40% of people aged 16 to 24 years reported loneliness as compared to 27% of people older than 75 years [28]. Behavioral genetic analysis indicates that loneliness and depression are influenced by the same genes, while the effects of these genes can be higher in younger people. In view of the high prevalence of loneliness in younger people and its potential lifelong consequences, it is paramount to gain a better understanding of loneliness and its relation to depression in younger people and to design age-appropriate treatments.

Psychological interventions for overcoming loneliness have proven to work if the target areas are changing maladaptive social perception, increasing social contact, increasing opportunities for social interaction, and improving social interaction and interpersonal skills [31]. Technology-based interventions for loneliness focus on one or more of these target areas. Technology can engulf different modes of interventions, such as individual psychotherapy with cognitive behavioral therapy, mindfulness therapies, and social support groups [32-34]. Because of the COVID-19 pandemic, remote psychotherapies and support groups have been used. Internet-based cognitive behavioral therapy and skills training to access internet and web-based support have been used for remote therapies [35,36].

Figure 1 presents the particular dynamics of loneliness in younger people and shows that the causes of loneliness in younger people are distinct; therefore, the solutions for overcoming loneliness in younger people should address these distinct causes. These distinct causes include processes of identity formation, individuation from family, as well as experiencing new environments, such as that of college and a new city [15,16]. Figure 1 has presented these parameters in graphical form with the consequences of increased loneliness. The increased risk of mental health issues, isolation, and substance abuse are presented [18,37].

Figure 1. Dynamics of loneliness in young people.



Situating Loneliness Informatics: Overview of the Literature

This section aims to provide a brief overview of studies and technology-based interventions for areas and concerns other than loneliness in younger people.

There have been scoping reviews done for ICT-based programs and interventions for older people. A scoping review of reviews was carried out to study the effectiveness of communication technologies to reduce the feeling of loneliness in older people [14]. The review included 28 studies, which combined covered 248 primary studies spanning over 50 years. The conclusion was that communication-based technologies do in fact reduce feelings of loneliness in older people. Although, the study also found that there was a lack of evidence and limited insight into innovative technologies, such as extended reality frameworks. In work by Shah et al [1], a systematic review and meta-analysis was carried out for digital technology–based interventions to reduce loneliness in older adults. The study analyzed 6 articles including a total of 646 participants. The study showed no statistical difference between the effectiveness of digital intervention, but it self-reported the lack of enough studies and small sample size of participants as the cause for the lack of validating effectiveness.

However, there are studies (eg, [38]) that establish the relationship between communication-based technologies and reduction in the feeling of loneliness. A cross-sectional study of 4315 adults older than 50 years reported that rural older adults who used technology less frequently felt loneliness more than urban older adults. The study also explored the relationship between race, urbanity, and age, and recommended that technology-based intervention be consider a social technology to address rural and racial disparities. Nevertheless, the connection or direct correlation between social media technology use and reduction in the feeling of loneliness is not strongly established. Wiwatkunupakarn et al [12] performed a study of the effectiveness of social networking site usage in older people for reducing loneliness. The study analyzed 10 observational studies and 5 experimental studies, of which 5 studies focused on loneliness and social isolation. Among the observational studies, some evidence was found that the use of social networking sites was associated with reductions in the feelings of loneliness and depression; however, the study was lacking on the experimental side. The conclusion here, too, required more studies to establish unambiguously the relationship between technology use and loneliness.

As can be seen from the reviews presented in this section, work in technology-based intervention to reduce loneliness has been focused on older people. There is a protocol design study for carrying out technology-based interventions for older people, which details the parameters on which such a scoping review can be done [6]. There is also a survey study about designing a digital psychoeducation tool for reducing loneliness in older people [39]. The study highlighted how the older adults' concerns about technology could be incorporated through an iterative process in a co-design approach to designing a technological approach. Budak et al [40] studied a technology-based intervention for dementia and loneliness with a particular focus on the role of assistive technologies. It was concluded that assistive technologies meant to fight dementia have a positive effect in reducing loneliness in older people.

Many studies meant to analyze loneliness in young people also consider older people. For example, Loveys et al [31] found that technology-based interventions for both older and younger people were acceptable to both groups. Other studies also focused on surveying and designing feasibility cases for technology-based intervention in older people [41,42]. The common denominator in these technology-based studies is that both older and younger people are considered.

Loneliness Informatics for Younger People

This section presents in detail the literature on loneliness informatics for younger people. This section aims to present focused technological interventions for younger people in order to build upon them. The following subsections discuss each in detail.

ICT-Based Interventions

While ICT-based interventions are used extensively for overcoming loneliness in older people in the literature, it can also be applied to younger people. ICT-based interventions for loneliness in older people suit their needs of communication and connection as their loneliness is often born out of social isolation. Pitman et al [9] presented a scoping review of reviews of ICT-based interventions for older people; however, such reports do not exist for younger people.

Studies on the effectiveness of ICT-based interventions for young people are rare. Internet, videoconferencing, email, telephone, video game consoles, tablets, and smartphones are the mediums for implementing ICT-based loneliness intervention strategies. The use of these mediums and tools is prevalent in younger people, and intervention strategies for loneliness for younger people can be designed around them. In work by Stephens-Reicher et al [43], an evidence base was built to see the efficacy of ICT interventions for youths facing social isolation and ensuing mental health problems. It has to be noted that social isolation is not totally specific to young people, and other groups, such as transgender people, may feel it and may rely upon the internet to find safety and connection. Similarly, for young people experiencing limited career opportunities as well as limited social and educational opportunities, ICT tools may be the gateway to building friendships and connection and seeking ways for growth. It was also reported by Yeo and Sawyer [44] that young people with chronic illnesses may have limited opportunities for social interaction, which can be overcome by the use of ICT for continual social and educational participation [45].

The social isolation born out of geographical location and socioeconomic parameters, such as race, class, and religious identity, can reduce access to face-to-face services for loneliness and other mental health–related services. ICT-based interventions for loneliness and related mental health problems, if delivered strategically, can overcome geographic and socioeconomic discriminations. The developmental challenges faced by youths that lead to social isolation, such as growth and consolidation of identity, maturation of identity, and transition into formal schooling, can well be addressed by ICT-based interventions. The specific importance of social and digital media in web-based spaces for providing support during the COVID-19 pandemic has been reported [46]. The ability to connect with others is also reported to be associated with well-being achieved through using communication technologies [47]. The positive effects of technology intervention were reported to be 44% [29]. A scoping review of interventions specifically for youths suggested that technology may be an effective alternative to face-to-face interventions as 90% of younger adults use the internet at least occasionally [48]. The current literature is mostly focused on exploring technology-based interventions to overcome loneliness for older adults. This calls for the need to explore ICT and other technology-based interventions to overcome loneliness in younger people.

Social Media–Based Interventions

Like ICT-based interventions for loneliness, social networking sites (SNSs) and social media–based interventions for loneliness have been considered for older adults in the literature. Because of the wide appeal and use of social media and SNSs, there are a few studies of these mediums for loneliness interventions in younger people as well. The main distinction between social media and SNSs is that SNSs allow users to share content with limited users to see responses from friends, family members, or followers [12,49]. This makes popular social media sites, such as Twitter, Facebook, and Instagram, SNSs but not the other way around. SNSs with limited circles of interaction were found to be effective in social support and be a source of health information [50]. Loneliness is reported to be reduced through the use of SNS-based interventions for older people [51].

Social media–based interventions for loneliness can address the needs of younger people to connect and create bonds. The displacement of identity and social relations, which result in loneliness in younger people, can be overcome through SNS-based interventions. Studies show that social media technology enables personal relationships inside and outside of one's social circle. The use of social media is also shown to result in a deepened sense of identity and purpose [52,53]. When considering the use of social media–based interventions for youths, the following questions should be addressed: Does social media help in forming social relationships among youths or the opposite? Do some youths, on the basis of their socioeconomic or geographical realities, become more prone to using social media as their primary source of socialization? Finally, is social media use associated with problematic internet use (PIU), and at what point do the benefits of social media use change into PIU?

A study in 2018 explored the relationship of social media use over time in youths and frequent face-to-face communication with close social circles, such as family and friends, and their associated subjective well-being [54]. The research provided little support for the social displacement model theory, which states that social media displaces or replaces real-life social relationships, which have depth with remote social media connections. In fact, the study found a positive association between social media use and well-being changes. Similarly, there is little support for the assertion that increased social media use can result in loneliness and depression [55]. If social media and the internet are better integrated into the lives of the participants, such as when they are used to interact with friends and family, the result is the feeling of loneliness fades over time.

The social compensation model, contrary to the social displacement model, asserts that social media may reduce the feeling of loneliness because it may be seen by users as a safer space for exploring social connections. Ellison et al [56] found a positive correlation between the self-esteem of college students and the use of social media sites. Individuals who are socially anxious, introverted, or less likely to self-disclose may find social media particularly helpful in establishing contacts, thus reducing the feelings of social isolation and loneliness. The self-disclosure hypothesis has also been supported in the literature [57]. It was showed that, because of reduced visual, auditory, and contextual signals, the users may become more confident to share their feelings, vulnerabilities, and emotions as they perceive the medium and the interaction to be less judgmental, hence leading to more self-disclosure and deeper relationships.

In a report by Vincent [58], the use of social media to develop a sense of belonging was studied. Belonging and loneliness are associated—a sense of belonging to a community brings down the feeling of loneliness. The study explored the role of social media as a therapeutic model used by college counsellors. The results showed that such use of social media increased the sense of belonging. This study suggests that social media can be used by college counselors as a potential tool to overcome loneliness in college students. Similar results were reported by Liu et al [59] for minority youths to identify more with web-based friends and to report more support on the internet than in real life. When belonging and overcoming loneliness is the central motivating factor behind the use of social media, it was found that the sense of community is likely to increase, resulting in a reduction in loneliness [60].

Robots, Conversational Agents, and Digital Humans: Application to Loneliness in Younger People

Robots are another area of technology-based interventions for loneliness. Robots are particularly useful for older adults [31,61]. Studies have suggested different robots for companionship and reported a general reduction in loneliness and an increase in the sense of companionship [46,62-64]. Social robots are also reported to improve loneliness in younger adults [65]. Similarly, conversational agents (CAs), where robots or interactive mobile or software applications conversate, have been effective in older people, especially if they are human-like embodied and use proactive communications.

CAs can be employed for all age groups as they respond to the needs of people with loneliness. Digital humans (DHs) can be classified as a type of CA, with the difference that the agent who is interacting with the user has a human-like appearance and uses artificial intelligence to build real-time social and emotional engagement with users [66]. Interacting with such an agent can reduce the feeling of loneliness, as the appearance of DHs is human-like because they are based on real-life characters. Some DHs are designed to include a complex cognitive architecture modelled on human behavior, which is able to show attachment and separation toward users to influence their behavior. DHs can be useful in delivering remote loneliness intervention as they are scalable and require access to the internet and a device, be it a computer, tablet, or mobile device.

The features of DHs that can be useful for countering loneliness are the real-life tasks that can be recommended by the human-like CA to the user. In a study by Loveys et al [31], a DH was designed by the name of "Bella". Bella was modelled to be a woman of Maori and New Zealand European decent, and the program was accessible through the web. The conversation facilitator (ie, Bella), responded to users' input through speech, text, or a prompt to press a button on a screen. Bella engaged in human-like facial gestures, and her face portrayed joy and concern. Bella also had linguistic variation of her own such that the user did not feel mechanical repetition of the same phrases. The relationship-building strategies employed by Bella were taken from psychology and human-computer interaction research [67]. These included tasks such as positive self-affirmation, reaching out to a friend, and complimenting someone you know. Participants who took part in interactions with Bella reported that Bella was useful in improving the feeling of loneliness and that they would like to interact with Bella again in the future. The study also reported that it was easy to train younger people to use the service either in a clinic room at the university or through video calling, and technical support requests were low for younger people during the training. DHs are used for a variety of purposes. DHs play a key role in augmenting reality and in other forms of extended reality. These are not limited to mental health applications, or in the case of this paper, to loneliness informatics.

Conclusion and Future Directions

This paper presented a comprehensive review of loneliness informatics. While there has been significant work done on countering loneliness in older adults, the same is not true for younger people, who have the highest loneliness numbers. This paper discussed in detail the particular dynamics of loneliness in younger people, which range from developmental changes to transitioning into new social contexts to cultural and psychological dynamics that begin to form at adolescence, such as the search for personal identity, meaning, and purpose. This paper presented an overview of loneliness informatics for older people, which included the use of digital media, social media, and robotics for interventions to cope with loneliness.

The focus of loneliness informatics is not on the younger population. However, in terms of social media, there have been some studies carried out. One downside of social media and digital media use is that it is often asserted that they are responsible for increased loneliness in younger people. The reported studies in this paper suggested that this may not be the case. However, there still is the possibility of PIU. There are factors that may affect addiction possibilities, such as low social capital, which can carry over to social media and bullying, which can happen in social media and web-based spaces. Some students and younger people who have social anxiety issues use social media for socialization. Overall, social media was found to be an effective intervention tool and strategy for countering loneliness. CAs and DHs, forms of digital technology with the added features of human-like language processing or human-like appearance, were also found to be helpful as coping mechanisms for loneliness. Because of the requirement of technology competency, the study found that they are more helpful to younger people.

Understanding loneliness from a health informatics perspective is an open field. There are issues, gaps, and challenges that should be addressed in future works. A comprehensive review of different aspects of the application of technology to understand loneliness and to design intervention programs for overcoming loneliness can be undertaken. Examples of such reviews can be the application of virtual reality– and extended reality–based interventions for loneliness, machine learning–based applications in serious games for loneliness, and comparative analysis of these technologies for younger and older people. There are also other aspects of loneliness and its intersection with technology that need to be explored, such as whether the use of technology, especially social media, leads to loneliness and how to use social media effectively for countering loneliness. Other than analysis of the literature, loneliness informatics can use data to understand loneliness in different geographical regions as well as different contexts. One challenge to loneliness informatics is the availability of data, which needs to be approximated from different sources, such as social media and news analyses. Another challenge is that the intervention strategies have to consider the social and economic contexts of loneliness, which need policymaking at multiple levels of government and organizations.

Authors' Contributions

HAS designed the study, selected sources to review, and wrote the manuscript. MH supervised the study and reviewed the draft of the paper.

Conflicts of Interest

None declared.

References

1. Shah SGS, Nogueras D, van Woerden HC, Kiparoglou V. Evaluation of the effectiveness of digital technology interventions to reduce loneliness in older adults: systematic review and meta-analysis. J Med Internet Res 2021 Jun 04;23(6):e24712 [FREE Full text] [doi: 10.2196/24712] [Medline: 34085942]
2. Schrempft S, Jackowska M, Hamer M, Steptoe A. Associations between social isolation, loneliness, and objective physical activity in older men and women. BMC Public Health 2019 Jan 16;19(1):74 [FREE Full text] [doi: 10.1186/s12889-019-6424-y] [Medline: 30651092]

3. Okwaraji FE, Obiechina KI, Onyebueke GC, Udegbunam ON, Nnadum GS. Loneliness, life satisfaction and psychological distress among out-of-school adolescents in a Nigerian urban city. Psychol Health Med 2018 Oct;23(9):1106-1112. [doi: 10.1080/13548506.2018.1476726] [Medline: 29792068]
4. van den Broek T. Gender differences in the correlates of loneliness among Japanese persons aged 50-70. Australas J Ageing 2017 Sep;36(3):234-237. [doi: 10.1111/ajag.12448] [Medline: 28685951]
5. Holt-Lunstad J. The potential public health relevance of social isolation and loneliness: prevalence, epidemiology, and risk factors. Public Policy Aging Rep 2017;27(4):127-130. [doi: 10.1093/ppar/prx030]
6. Wister A, Fyffe I, O'Dea E. Technological interventions for loneliness and social isolation among older adults: a scoping review protocol. Syst Rev 2021 Aug 07;10(1):217 [FREE Full text] [doi: 10.1186/s13643-021-01775-6] [Medline: 34362447]
7. Grenade L, Boldy D. Social isolation and loneliness among older people: issues and future challenges in community and residential settings. Aust Health Rev 2008 Aug;32(3):468-478. [doi: 10.1071/ah080468] [Medline: 18666874]
8. Wister A, Menec V, Mugford G. Loneliness, social isolation, and social engagement. The Canadian Longitudinal Study on Aging (CLSA) Report on Health and Aging in Canada. 2018. URL: https://www.clsa-elcv.ca/doc/2639 [accessed 2023-10-12]
9. Pitman A, Mann F, Johnson S. Advancing our understanding of loneliness and mental health problems in young people. Lancet Psychiatry 2018 Dec;5(12):955-956. [doi: 10.1016/s2215-0366(18)30436-x]
10. Hickin N, Käll A, Shafran R, Sutcliffe S, Manzotti G, Langan D. The effectiveness of psychological interventions for loneliness: a systematic review and meta-analysis. Clin Psychol Rev 2021 Aug;88:102066. [doi: 10.1016/j.cpr.2021.102066] [Medline: 34339939]
11. Cacioppo JT, Hawkley LC. Perceived social isolation and cognition. Trends Cogn Sci 2009 Oct;13(10):447-454 [FREE Full text] [doi: 10.1016/j.tics.2009.06.005] [Medline: 19726219]
12. Wiwatkunupakarn N, Pateekhum C, Aramrat C, Jirapornchaoren W, Pinyopornpanish K, Angkurawaranon C. Social networking site usage: a systematic review of its relationship with social isolation, loneliness, and depression among older adults. Aging Ment Health 2022 Jul;26(7):1318-1326. [doi: 10.1080/13607863.2021.1966745] [Medline: 34427132]
13. Choi HK, Lee SH. Trends and effectiveness of ICT interventions for the elderly to reduce loneliness: a systematic review. Healthcare (Basel) 2021 Mar 07;9(3):293 [FREE Full text] [doi: 10.3390/healthcare9030293] [Medline: 33800099]
14. Döring N, Conde M, Brandenburg K, Broll W, Gross H, Werner S, et al. Can communication technologies reduce loneliness and social isolation in older people? A scoping review of reviews. Int J Environ Res Public Health 2022 Sep 08;19(18):11310 [FREE Full text] [doi: 10.3390/ijerph191811310] [Medline: 36141581]
15. Hards E, Loades ME, Higson-Sweeney N, Shafran R, Serafimova T, Brigden A, et al. Loneliness and mental health in children and adolescents with pre-existing mental health problems: a rapid systematic review. Br J Clin Psychol 2022 Jun;61(2):313-334. [doi: 10.1111/bjc.12331] [Medline: 34529837]
16. Blakemore S, Mills KL. Is adolescence a sensitive period for sociocultural processing? Annu Rev Psychol 2014;65:187-207. [doi: 10.1146/annurev-psych-010213-115202] [Medline: 24016274]
17. Stickley A, Koyanagi A, Koposov R, Schwab-Stone M, Ruchkin V. Loneliness and health risk behaviours among Russian and U.S. adolescents: a cross-sectional study. BMC Public Health 2014 Apr 16;14:366 [FREE Full text] [doi: 10.1186/1471-2458-14-366] [Medline: 24735570]
18. Study Quality Assessment Tools. National Heart, Lung, and Blood Institute. URL: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools [accessed 2023-10-12]
19. Shanahan L, Steinhoff A, Bechtiger L, Murray AL, Nivette A, Hepp U, et al. Emotional distress in young adults during the COVID-19 pandemic: evidence of risk and resilience from a longitudinal cohort study. Psychol Med 2020 Jun 23;52(5):824-833. [doi: 10.1017/s003329172000241x]
20. Peplau LA, Perlman D, editors. Loneliness: A Sourcebook of Current Theory, Research and Therapy. New York: Wiley; 1982.
21. Gierveld JDJ. A review of loneliness: concept and definitions, determinants and consequences. Rev Clin Gerontol 1998 Feb 01;8(1):73-80. [doi: 10.1017/s0959259898008090]
22. Spithoven AWM, Cacioppo S, Goossens L, Cacioppo JT. Genetic contributions to loneliness and their relevance to the evolutionary theory of loneliness. Perspect Psychol Sci 2019 May;14(3):376-396. [doi: 10.1177/1745691618812684] [Medline: 30844327]
23. Maes M, Qualter P, Vanhalst J, Van den Noortgate W, Goossens L. Gender differences in loneliness across the lifespan: a meta–analysis. Eur J Pers 2019 Nov 01;33(6):642-654. [doi: 10.1002/per.2220]
24. Achterbergh L, Pitman A, Birken M, Pearce E, Sno H, Johnson S. The experience of loneliness among young people with depression: a qualitative meta-synthesis of the literature. BMC Psychiatry 2020 Aug 24;20(1):415 [FREE Full text] [doi: 10.1186/s12888-020-02818-3] [Medline: 32831064]
25. Stessman J, Rottenberg Y, Shimshilashvili I, Ein-Mor E, Jacobs JM. Loneliness, health, and longevity. J Gerontol A Biol Sci Med Sci 2014 Jun;69(6):744-750. [doi: 10.1093/gerona/glt147] [Medline: 24077598]
26. Vanhalst J, Goossens L, Luyckx K, Scholte RH, Engels RC. The development of loneliness from mid- to late adolescence: trajectory classes, personality traits, and psychosocial functioning. J Adolesc 2013 Dec;36(6):1305-1312. [doi: 10.1016/j.adolescence.2012.04.002] [Medline: 22560517]

27. Weiss RS, editor. Loneliness: The Experience of Emotional and Social Isolation. Cambridge: The MIT Press; 1975.
28. Children's and young people's experiences of loneliness: 2018. Office for National Statistics. URL: https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/articles/childrensandyoungpeoplesexperiencesofloneliness/2018 [accessed 2023-10-12]
29. Qualter P, Brown SL, Munn P, Rotenberg KJ. Childhood loneliness as a predictor of adolescent depressive symptoms: an 8-year longitudinal study. Eur Child Adolesc Psychiatry 2010 Jun;19(6):493-501. [doi: 10.1007/s00787-009-0059-y] [Medline: 19777287]
30. Salaheddin K, Mason B. Identifying barriers to mental health help-seeking among young adults in the UK: a cross-sectional survey. Br J Gen Pract 2016 Sep 29;66(651):e686-e692. [doi: 10.3399/bjgp16x687313]
31. Loveys K, Sagar M, Pickering I, Broadbent E. A digital human for delivering a remote loneliness and stress intervention to at-risk younger and older adults during the COVID-19 pandemic: randomized pilot trial. JMIR Ment Health 2021 Nov 08;8(11):e31586 [FREE Full text] [doi: 10.2196/31586] [Medline: 34596572]
32. Chiang K, Chu H, Chang H, Chung M, Chen C, Chiou H, et al. The effects of reminiscence therapy on psychological well-being, depression, and loneliness among the institutionalized aged. Int J Geriatr Psychiatry 2010 Apr;25(4):380-388. [doi: 10.1002/gps.2350] [Medline: 19697299]
33. Creswell JD, Irwin MR, Burklund LJ, Lieberman MD, Arevalo JM, Ma J, et al. Mindfulness-based stress reduction training reduces loneliness and pro-inflammatory gene expression in older adults: a small randomized controlled trial. Brain Behav Immun 2012 Oct;26(7):1095-1101 [FREE Full text] [doi: 10.1016/j.bbi.2012.07.006] [Medline: 22820409]
34. Collins CC, Benedict J. Evaluation of a community-based health promotion program for the elderly: lessons from Seniors CAN. Am J Health Promot 2006 Sep;21(1):45-48. [doi: 10.4278/0890-1171-21.1.45]
35. Käll A, Jägholm S, Hesser H, Andersson F, Mathaldi A, Norkvist BT, et al. Internet-based cognitive behavior therapy for loneliness: a pilot randomized controlled trial. Behav Ther 2020 Jan;51(1):54-68. [doi: 10.1016/j.beth.2019.05.001] [Medline: 32005340]
36. Jones RB, Ashurst EJ, Atkey J, Duffy B. Older people going online: its value and before-after evaluation of volunteer support. J Med Internet Res 2015 May 18;17(5):e122 [FREE Full text] [doi: 10.2196/jmir.3943] [Medline: 25986724]
37. Kessler R, Angermeyer M, Anthony JC, DE Graaf R, Demyttenaere K, Gasquet I, et al. Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's World Mental Health Survey Initiative. World Psychiatry 2007 Oct;6(3):168-176 [FREE Full text] [Medline: 18188442]
38. Byrne KA, Anaraky RG, Dye C, Ross LA, Chalil Madathil K, Knijnenburg B, et al. Examining rural and racial disparities in the relationship between loneliness and social technology use among older adults. Front Public Health 2021;9:723925 [FREE Full text] [doi: 10.3389/fpubh.2021.723925] [Medline: 34532308]
39. Al Mahmud A, Long KM, Harrington KD, Casey K, Bhar S, Curran S, et al. Developing a digital psychoeducational tool to reduce loneliness in older adults: a design case study. Int J Hum-Comput Int 2021 Aug 04;38(6):499-528. [doi: 10.1080/10447318.2021.1949854]
40. Budak KB, Atefi G, Hoel V, Laporte Uribe F, Meiland F, Teupen S, et al. Can technology impact loneliness in dementia? A scoping review on the role of assistive technologies in delivering psychosocial interventions in long-term care. Disabil Rehabil Assist Technol 2021 Nov 09:1-13. [doi: 10.1080/17483107.2021.1984594] [Medline: 34752177]
41. Amundsen D. Digital technologies as a panacea for social isolation and loneliness among older adults: an intervention model for flourishing and wellbeing. Video J Educ Pedagog 2021 Jan 11;5(1):1-14. [doi: 10.1163/23644583-00501008]
42. Hoang P, Whaley C, Thompson K, Ho V, Rehman U, Boluk K, et al. Evaluation of an intergenerational and technological intervention for loneliness: protocol for a feasibility randomized controlled trial. JMIR Res Protoc 2021 Feb 17;10(2):e23767 [FREE Full text] [doi: 10.2196/23767] [Medline: 33595443]
43. Stephens-Reicher J, Metcalf A, Blanchard M, Mangan C, Burns J. Reaching the hard-to-reach: how information communication technologies can reach young people at greater risk of mental health difficulties. Australas Psychiatry 2011 Jul:S58-S61. [doi: 10.3109/10398562.2011.583077] [Medline: 21878021]
44. Yeo M, Sawyer S. Chronic illness and disability. BMJ 2005 Mar 26;330(7493):721-723 [FREE Full text] [doi: 10.1136/bmj.330.7493.721] [Medline: 15790645]
45. Third A, Richardson I. Analysing the impacts of social networking for young people living with chronic illness, a serious condition or a disability: an evaluation of the livewire online community. Murdoch Univeristy. 2009. URL: https://researchportal.murdoch.edu.au/esploro/outputs/report/Analysing-the-impacts-of-social-networking/991005543794107891 [accessed 2023-10-12]
46. Robinson H, Macdonald B, Kerse N, Broadbent E. The psychosocial effects of a companion robot: a randomized controlled trial. J Am Med Dir Assoc 2013 Sep;14(9):661-667. [doi: 10.1016/j.jamda.2013.02.007] [Medline: 23545466]
47. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
48. Eccles AM, Qualter P. Alleviating loneliness in young people - a meta-analysis of interventions. Child Adolesc Ment Health 2021 Feb;26(1):17-33. [doi: 10.1111/camh.12389] [Medline: 32406165]

49. Casanova G, Zaccaria D, Rolandi E, Guaita A. The effect of information and communication technology and social networking site use on older people's well-being in relation to loneliness: review of experimental studies. J Med Internet Res 2021 Mar 01;23(3):e23588 [FREE Full text] [doi: 10.2196/23588] [Medline: 33439127]
50. Teng C, Joo TM. Analyzing the usage of social media: A study on elderly in Malaysia. Int J Humanit Soc Sci 2017;11(3):737-743.
51. Goswami S, Köbler F, Leimeister JM, Krcmar H. Using online social networking to enhance social connectedness and social support for the elderly. In: ICIS 2010 PROCEEDINGS. 2010 Presented at: ICIS 2010; December 12-15, 2010; St. Louis.
52. Smith D, Leonis T, Anandavalli S. Belonging and loneliness in cyberspace: impacts of social media on adolescents' well-being. Aust J Psychol 2021 Mar 31;73(1):12-23. [doi: 10.1080/00049530.2021.1898914]
53. Verduyn P, Ybarra O, Résibois M, Jonides J, Kross E. Do social network sites enhance or undermine subjective well - being? A critical review. Soc Issues Policy Rev 2017 Jan 13;11(1):274-302. [doi: 10.1111/sipr.12033]
54. Hall JA, Kearney MW, Xing C. Two tests of social displacement through social media use. Inf Commun Soc 2018 Feb 02;22(10):1396-1413. [doi: 10.1080/1369118x.2018.1430162]
55. Houghton S, Lawrence D, Hunter SC, Rosenberg M, Zadow C, Wood L, et al. Reciprocal relationships between trajectories of depressive symptoms and screen media use during adolescence. J Youth Adolesc 2018 Nov;47(11):2453-2467 [FREE Full text] [doi: 10.1007/s10964-018-0901-y] [Medline: 30046970]
56. Ellison N, Steinfield C, Lampe C. The benefits of Facebook "friends:" social capital and college students' use of online social network sites. J Comput-Mediat Comm 2007;12(4):1143-1168. [doi: 10.1111/j.1083-6101.2007.00367.x]
57. Valkenburg PM, Peter J. Online communication among adolescents: an integrated model of its attraction, opportunities, and risks. J Adolesc Health 2011 Feb;48(2):121-127. [doi: 10.1016/j.jadohealth.2010.08.020] [Medline: 21257109]
58. Vincent EA. Social media as an avenue to achieving sense of belonging among college students.Vistas Online,36. Vistas ONLINE. 2016. URL: https://www.counseling.org/docs/default-source/vistas/social-media-as-an-avenue.pdf [accessed 2023-10-12]
59. Liu D, Baumeister RF, Yang CC, Hu B. Digital communication media use and psychological well-being: a meta-analysis. J Comput-Mediat Comm 2019;24(5):259-273. [doi: 10.1093/jcmc/zmz013]
60. Liu Q, Shao Z, Fan W. The impact of users' sense of belonging on social media habit formation: empirical evidence from social networking and microblogging websites in China. Int J Inf Manage 2018 Dec;43:209-223. [doi: 10.1016/j.ijinfomgt.2018.08.005]
61. Gasteiger N, Loveys K, Law M, Broadbent E. Friends from the future: a scoping review of research into robots and computer agents to combat loneliness in older people. Clin Interv Aging 2021 May;16:941-971. [doi: 10.2147/cia.s282709]
62. Niemelä M, van Aerschot L, Tammela A, Aaltonen I, Lammi H. Towards ethical guidelines of using telepresence robots in residential care. Int J of Soc Robotics 2019 Feb 22;13(3):431-439. [doi: 10.1007/s12369-019-00529-8]
63. Barrett E, Burke M, Whelan S, Santorelli A, Oliveira BL, Cavallo F, et al. Evaluation of a companion robot for individuals with dementia: quantitative findings of the MARIO project in an Irish residential care setting. J Gerontol Nurs 2019 Jul 01;45(7):36-45. [doi: 10.3928/00989134-20190531-01] [Medline: 31237660]
64. Gross HM, Scheidig A, Müller S, Schütz B, Fricke C, Meyer S. Living with a mobile companion robot in your own apartment-final implementation and results of a 20-weeks field study with 20 seniors. 2019 Presented at: 2019 IEEE International Conference on Robotics and Automation; May 20-24, 2019; Montreal p. 2253-2259. [doi: 10.1109/icra.2019.8793693]
65. Jeong K, Sung J, Lee HS, Kim A, Kim H, Park C, et al. Fribo: a social networking robot for increasing social connectedness through sharing daily home activities from living noise data. In: Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction. 2018 Presented at: HRI '18; March 5-8, 2018; Chicago p. 114-122. [doi: 10.1145/3171221.3171254]
66. Sagar M, Bullivant D, Robertson P, Efimov O, Jawed K, Kalarot R, et al. A neurobehavioural framework for autonomous animation of virtual human faces. In: SIGGRAPH Asia 2014 Autonomous Virtual Humans and Social Robot for Telepresence. 2014 Presented at: SA '14; December 3-6, 2014; Shenzhen p. 1-10. [doi: 10.1145/2668956.2668960]
67. Loveys K, Sebaratnam G, Sagar M, Broadbent E. The effect of design features on relationship quality with embodied conversational agents: a systematic review. Int J of Soc Robotics 2020 Sep 24;12(6):1293-1312. [doi: 10.1007/s12369-020-00680-7]

Abbreviations

CA: conversational agent
DH: digital human
ETL: evolutionary theory of loneliness
ICT: information and communication technology
PIU: problematic internet use
SNS: social networking site

Edited by A Mavragani; submitted 20.12.22; peer-reviewed by A Molinard-Chenu, J Taylor; comments to author 02.02.23; revised version received 20.03.23; accepted 30.04.23; published 02.11.23.

Please cite as:
Shah HA, Househ M
Understanding Loneliness in Younger People: Review of the Opportunities and Challenges for Loneliness Interventions
Interact J Med Res 2023;12:e45197
URL: https://www.i-jmr.org/2023/1/e45197
doi:10.2196/45197
PMID:37917125

Review

Electronic Medical Record System Use and Determinants in Ethiopia: Systematic Review and Meta-Analysis

Masresha Derese Tegegne[1], BSc, MPH; Sisay Maru Wubante[1], BSc, MPH; Mulugeta Hayelom Kalayou[2], BSc, MPH; Mequannent Sharew Melaku[1], BSc, MPH; Binyam Tilahun[1], BSc, MPH, PhD; Tesfahun Melese Yilma[1], BSc, MPH, PhD; Hiwote Simane Dessie[3], BSc

[1]Department of Health Informatics, Institute of Public Health, College of Medicine and Health Sciences, University of Gondar, Gondar, Ethiopia
[2]Department of Health Informatics, School of Public Health, College of Medicine and Health Sciences, Wollo University, Dessie, Ethiopia
[3]Department of Health Informatics, School of Public Health, College of Medicine and Health Sciences, Wachemo University, Hosaena, Ethiopia

Corresponding Author:
Masresha Derese Tegegne, BSc, MPH
Department of Health Informatics
Institute of Public Health, College of Medicine and Health Sciences
University of Gondar
PO Box 196
Gondar
Ethiopia
Phone: 251 935474755
Email: masresha1derese@gmail.com

Abstract

Background: The strategic plan of the Ethiopian Ministry of Health recommends an electronic medical record (EMR) system to enhance health care delivery and streamline data systems. However, only a few exhaustive systematic reviews and meta-analyses have been conducted on the degree of EMR use in Ethiopia and the factors influencing success. This will emphasize the factors that make EMR effective and increase awareness of its widespread use among future implementers in Ethiopia.

Objective: This study aims to determine the pooled estimate of EMR use and success determinants among health professionals in Ethiopia.

Methods: We developed a protocol and searched PubMed, Web of Sciences, African Journals OnLine, Embase, MEDLINE, and Scopus to identify relevant studies. To assess the quality of each included study, we used the Joanna Briggs Institute quality assessment tool using 9 criteria. The applicable data were extracted using Microsoft Excel 2019, and the data were then analyzed using Stata software (version 11; StataCorp). The presence of total heterogeneity across included studies was calculated using the index of heterogeneity I^2 statistics. The pooled size of EMR use was estimated using a random effect model with a 95% CI.

Results: After reviewing 11,026 research papers, 5 papers with a combined total of 2439 health workers were included in the evaluation and meta-analysis. The pooled estimate of EMR usage in Ethiopia was 51.85% (95% CI 37.14%-66.55%). The subgroup study found that the northern Ethiopian region had the greatest EMR utilization rate (58.75%) and that higher (54.99%) utilization was also seen in publications published after 2016. Age groups <30 years, access to an EMR manual, EMR-related training, and managerial support were identified factors associated with EMR use among health workers.

Conclusions: The use of EMR systems in Ethiopia is relatively low. Belonging to a young age group, accessing an EMR manual, receiving EMR-related training, and managerial support were identified as factors associated with EMR use among health workers. As a result, to increase the use of EMRs by health care providers, it is essential to provide management support and an EMR training program and make the EMR manual accessible to health professionals.

(Interact J Med Res 2023;12:e40721) doi:10.2196/40721

KEYWORDS

electronic medical record system; health professional; utilization; determinants; Ethiopia; medical record; EMR; EHR; electronic health record; health information technology; systematic review

Introduction

Background

Health information technology has transformed and improved health care delivery worldwide. Health information technology has been used for patient administration and management in health care systems. The electronic medical record (EMR) is widely regarded as a critical health information technology tool for improving the quality of medical care [1]. EMRs are computerized patient record systems introduced in the early 1970s to collect, store, and display patient information [2,3]. EMRs can include a variety of clinical services units, such as test ordering, consultation, e-prescription, decision support system, digital imaging, and telemedicine, while protecting patient privacy and confidentiality [4-6].

Implementing the EMR system is the priority agenda in both high-income and resource-limited countries [7]. The adoption of EMRs is a prerequisite for improving clinical decision-making as well as the privacy and security of patients' information [1]. The perceived benefits that EMRs could provide for the health care system include the following: safety, the organization of patient information, coordination of care, communication, health history, timely access to medical information, and the effectiveness of care [7-9]. Furthermore, evidence shows that EMRs can improve data quality by recording patient information and performing health care functions [8]. This prompted health administrators to develop a program to promote the use of EMRs in the health care system. However, a small proportion of low-income countries have successfully implemented national health information systems.

The Ethiopian Ministry of Health, with the assistance of various nongovernmental organizations, adapted the SmartCare EMR system as a national EMR system for all hospitals and scaled it up to additional hospitals and regions [10,11]. However, individual studies report that this EMR system is underused in Ethiopia, and the system faces sustainability challenges. According to a survey of the comprehensive evaluation of EMR systems in 5 Ethiopian hospitals, only about 31.7% of the participants used EMRs [7]. Similar studies in eastern Ethiopia revealed that EMRs are being used optimally [2]. Another study in the northwestern part of Ethiopia found that only 46.5% of participants used hospital EMR systems [11]. The main reasons for low utilization are implementation challenges and a lack of preimplementation measures, such as EMR readiness, knowledge of EMR, attitude toward EMR, and preimplementation training [2,10,11].

Ethiopia is currently implementing several initiatives to address the abovementioned challenges and strengthen national e-health systems to improve health data availability, accessibility, quality, and use in decision-making processes [12]. The strategic plan calls for an EMR system to streamline data systems and improve the health care delivery [13]. However, only a few comprehensive systematic reviews and meta-analyses are available on the level of EMR use in Ethiopia and the factors that influence its success. As a result, determining the combined level of use and identifying determinants affecting health professionals' EMR use is critical in confirming its optimal integration and ultimately measuring the benefits within the health care system.

Objective

This study is unique as it aims to expand our knowledge of the combined level of EMR usage by a health practitioner and offers important recommendations for the effective, efficient, and desirable integration of EMR systems into the Ethiopian health care system.

In our review, we specifically looked into the following questions:

- What is the pooled level of EMR use in Ethiopia?
- What are the determinant factors for EMR use in Ethiopia?

Methods

Reporting

This study followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines in its design and reporting (Multimedia Appendix 1) [14].

Search Strategy and Study Selection

We developed a protocol and searched PubMed, MEDLINE, Web of Sciences, African Journals OnLine, EMBASE, and Scopus to research EMR use and determinants in Ethiopia. To find publications, the following search strategy is used to do extensive searches in web-based databases: ["electronic medical record' OR 'electronic health record" OR 'electronic patient record' OR 'Decision Support Systems'] AND ['determinant' OR 'associated factors' OR 'barriers'].

Inclusion and Exclusion Criteria

Studies investigating the utilization and determinants of EMR systems in Ethiopia by the end of June 2022 were considered eligible. Studies that were published in English, in peer-reviewed journals, or as freely accessible full-text publications in the grey literature were all included in this analysis. However, studies without full text and with data that are difficult to extract, studies that are not published in English, studies that do not categorize outcome variables, and studies that do not reflect EMR use in Ethiopia were excluded from this analysis.

Measurement of the Outcome Variable

The main objectives of this review are to determine the pooled prevalence of EMR use and its determinants. EMR use was assessed based on published literature, with a category of "utilized" or "not utilized." The review's second outcome variable sought to uncover factors associated with Ethiopian health workers' use of EMR systems, which were measured using the odds ratio. The odds ratio for each identified factor was determined using the binary outcome data provided by each primary study.

Data Extraction and Management

Two authors (MDT and SMW) used Microsoft Excel to extract all the essential parameters independently. The first author's last name, year of publication, region, study area, study design, study population, sample size, percentage of EMR use with standard error, and determinant factors that affect utilization with the standard error were all extracted from each study. The disagreements between the two authors were resolved through discussion.

Quality Appraisal of the Individual Studies

Two authors appraised each study's quality independently (MDT and TMY). To assess the quality of each included study, we used the Joanna Briggs Institute quality assessment tool using 9 criteria [15]. The tool mainly included (1) an appropriate sample frame; (2) an appropriate sampling strategy; (3) an adequate sample size; (4) a description of the study subjects and setting; (5) data analysis conducted with sufficient coverage; (6) valid methods for condition identification; (7) the condition measured in a standard, reliable way for all participants; (8) appropriate statistical analysis; and (9) an adequate response rate. Each item was given a rating of "yes," "not reported," or "not appropriate." Finally, the total quality score was assigned based on the number of "yes" responses per study. Papers with a rating of 5 or above out of 9 were included in the final review (Multimedia Appendix 2).

Data Processing and Analysis

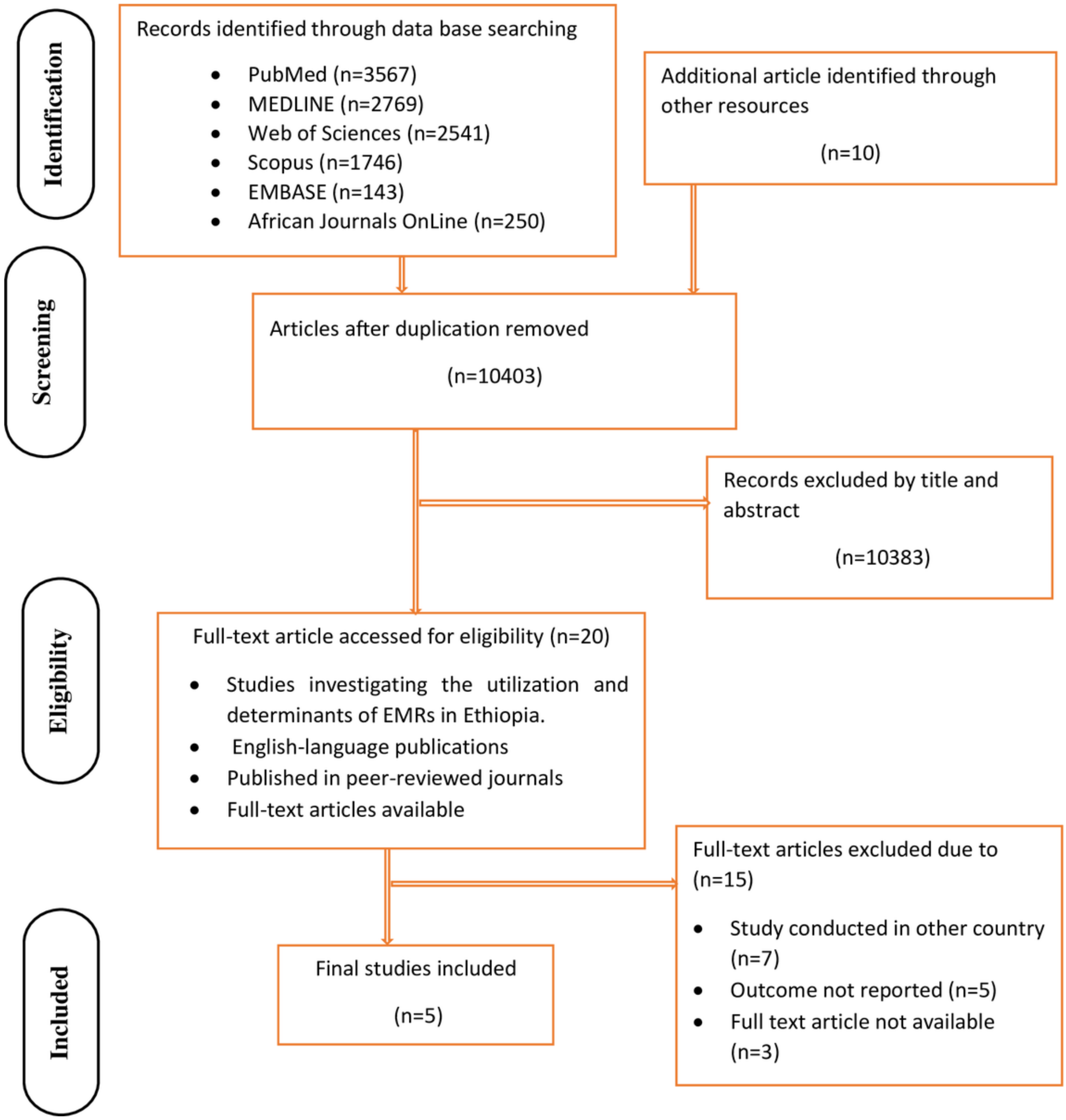
The relevant data were extracted using Microsoft Excel 2019. The data were then analyzed using Stata software (version 11; StataCorp). The pooled size of EMR use was estimated using a random effect model with a 95% CI [16]. The percentage of total variation across studies was calculated using the index of heterogeneity I^2 statistics [17]. Due to the heterogeneity of the included studies (I^2>75% and P<.05), the data were divided into subgroups according to the study region and year of publication. This was due to the highly diverse study regions and publication years of the included research. As a result, the random differences between the point estimations in the primary research are reduced. Researchers employed Egger's regression test and funnel plot analysis to identify publication bias [18,19]. P<.05 was considered a statistically significant publication bias in Egger's test.

Results

Search Results

A total of 11,026 articles on the use and determinants of EMRs in Ethiopia were found in PubMed, MEDLINE, Web of Sciences, African Journals OnLine, EMBASE, and Scopus. From the total number of retrieved studies, 623 papers were removed due to duplication, and 10,383 publications were excluded after being evaluated based on their titles and abstracts. The remaining 20 full-text publications were assessed for eligibility, with 15 articles further excluded based on the inclusion and exclusion criteria. Finally, only 5 publications were included in the final meta-analysis based on the predefined criteria and quality assessment (Figure 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow chart displaying the selection process of included studies.



Characteristics of Included Studies

This meta-analysis and systematic review included a total of 2439 health professionals. The number of studies with the smallest and largest sample sizes was 412 and 606, respectively. Among the included primary studies, 2 studies were undertaken in the eastern Ethiopia [2,20], 1 in northwest Ethiopia's Amhara region [11], 1 in the Tigray region [21], and 1 in Ethiopia's capital, Addis Ababa [7]. As shown in Table 1, these 5 original studies were published between early December 2014 and November 2021. All included studies used an institutional-based cross-sectional study design to estimate the use of EMR systems, as shown in Table 1.

Table 1. Summary of primary cross-sectional studies included in the meta-analysis of the use of electronic medical records among health professionals in Ethiopia, 2022.

Author and publication year	Region	Study area	Sample size	Magnitude	Quality[a]
Oumer et al [2], 2021	Harari region and Dire Dawa	Eastern Ethiopia	412	67.7	9
Mekonnen et al [20], 2021	Harari region	Harari Regional State	498	42.3	7
Biruk et al [11], 2014	Amhara	Northwest Ethiopia	606	46.5	9
Yehualashet et al [21], 2015	Tigray	Ayder Referral Hospital	501	71	8
Tilahun et al [7], 2015	Addis Ababa	Addis Ababa	422	31.7	9

[a]To assess the quality of each included study, we used the Joanna Briggs Institute quality assessment tool using 9 criteria.

The Pooled Utilization of EMR System in Ethiopia

The pooled estimate of EMR use in Ethiopia from 5 studies [2,7,11,20,21] was 51.85% (95% CI 37.14%-66.55%; Figure 2). The included studies were found to be heterogeneous (I^2>75% and *P*<.05) [22]. Subgroup analysis is done based on the study location and publication year due to high heterogeneity across the included studies (I^2=98.3% and *P*<.001; Figure 2).

According to the subgroup study, the northern Ethiopia region ranked highest in EMR use (58.75%), followed by the Eastern portions of Ethiopia (54.99%) and the Addis Ababa region (31.70%; Table 2). Furthermore, disparities in publication time were identified, with current publications on the use of EMRs being higher (54.99%) than the research published before 2016 (49.75%), as shown in Table 2.

Figure 2. Forest plot displaying a pooled estimate of electronic medical record use among health professionals in Ethiopia. ES: Effect Size.

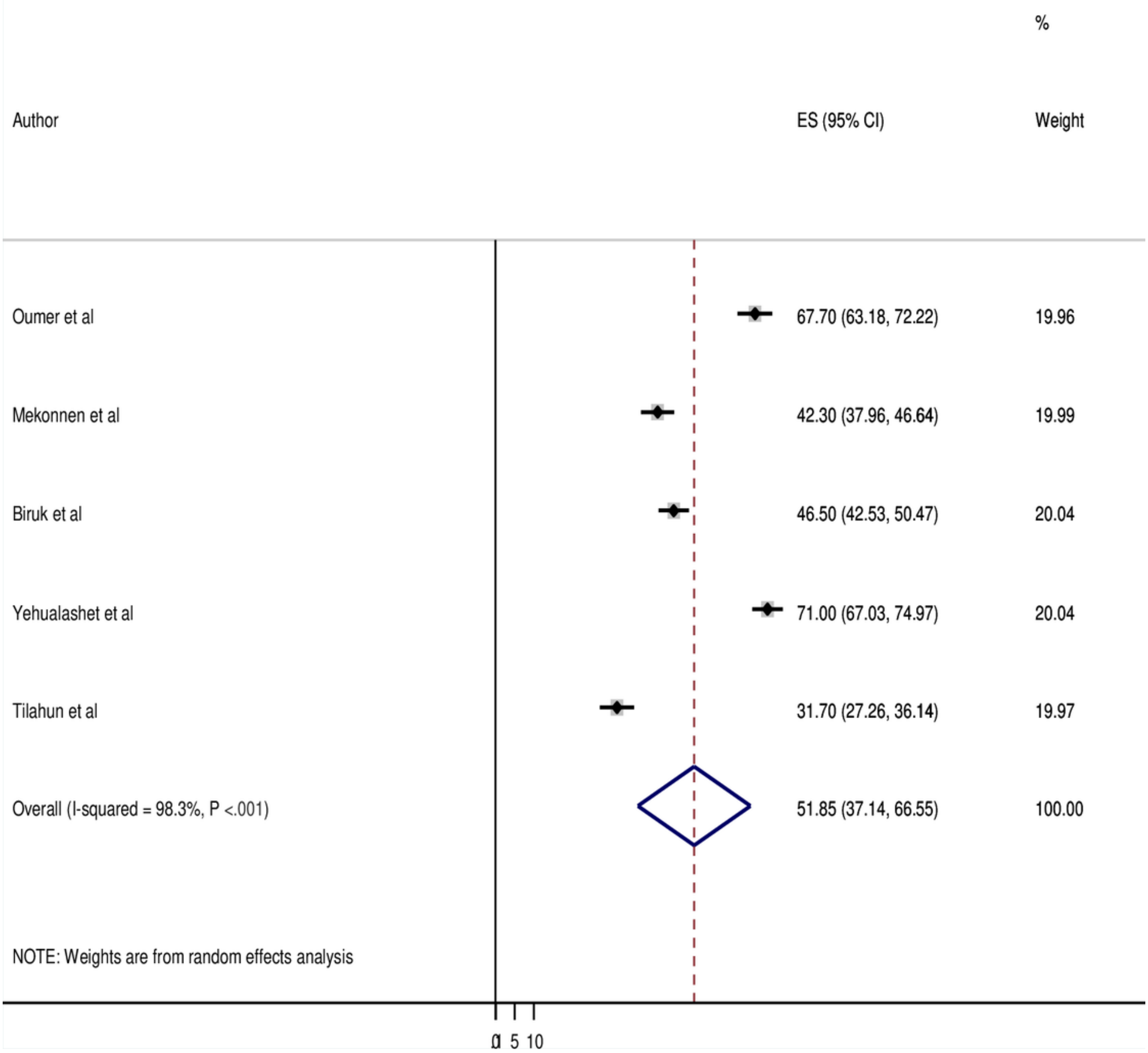


Table 2. Subgroup analysis by study location and publication year of electronic medical record use among health professionals in Ethiopia.

Variable and subgroup	Number of studies	Sample size, n	Prevalence (95% CI)	I^2 (%)	*P* value
Study location					
Eastern Ethiopia	2	910	54.99 (30.10-79.88)	98.4	<.001
Northern Ethiopia	2	1107	58.75 (34.74-82.76)	98.6	<.001
Addis Ababa	1	301	31.70 (27.26-66.55)	—[a]	—
Year of publication					
Before 2016	3	1529	49.75 (27.50-72.00)	98.9	<.001
After 2016	2	910	54.99 (30.10-79.88)	98.4	<.001

[a]Not applicable.

Sensitivity Analysis and Publication Bias

Sensitivity analysis revealed that the overall effect sizes remained stable with the deletion of any of the studies from the analysis without a notable improvement in heterogeneity (Table 3). A funnel plot and Egger's regression test were used to investigate potential publication bias. As a result, the funnel plot is symmetric, indicating no publishing bias because all of the research falls inside the triangular region (Figure 3). Furthermore, Egger's regression test results revealed no evidence of publication bias (*P*=.30; Table 4).

Table 3. Sensitivity analysis results for the 5 studies.

Study omitted	Estimates (95% CI)	Hetrogenity	
		I^2 (%)	*P* value
Oumer et al [2], 2021	47.89 (31.43-64.35)	98.4	<.001
Mekonnen et al [20], 2021	54.23 (36.36-72.10)	98.6	<.001
Biruk et al [11], 2014	53.18 (34.30-72.07)	98.7	<.001
Yehualashet et al [21], 2015	47.04 (32.82-61.27)	97.7	<.001
Tilahun et al [7], 2015	56.87 (42.56-71.19)	97.9	<.001
Combined	51.85 (37.14-66.55)	98.3	<.001

Figure 3. Funnel plot to test publication bias of the 5 included studies.

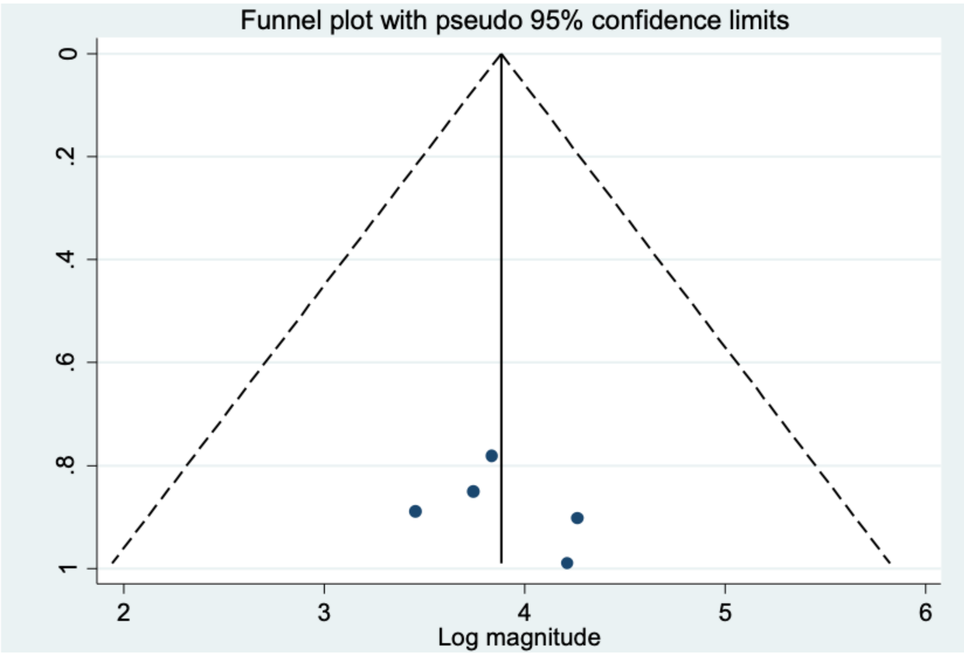


Table 4. Egger's test for publication bias of the 5 studies.

SE of the effect size	Coefficient	SE	*t* value	*P* value	95% CI
Slope	0.115799	0.4459117	0.26	.80	–0.9386145 to 1.170213
Bias	2.147455	1.909221	1.12	.30	–2.367135 to 6.662046

Factor Associated With the Use of EMR Systems

Some of the factors associated with the use of EMRs were quantitatively pooled in this systematic review and meta-analysis. In contrast, others were not because the independent variables were not consistently categorized about the use of EMRs.

Three studies indicated that health professionals who were younger (age groups <30 years) were 2.24 times (adjusted odds ratio [AOR]=2.24, 95% CI 1.36-3.68) more likely to use EMR compared to those whose age group was greater than or equal to 30 years. The included studies were characterized by the presence of heterogeneity (I^2=60.4%; P=.08). Hence a random effect model analysis was performed in this meta-analysis (Figure 4).

Two studies showed that the presence of an EMR manual has a significant association with the use of EMR systems. The odds of EMR use were 2.86 times (AOR=2.08, 95% CI 1.47-2.96) higher for health care professionals with EMR manuals compared to those without them. The included studies in this meta-analysis did not exhibit any heterogeneity (I^2=17.3%; P=.27). Consequently, a fixed-effect model analysis was performed (Figure 5).

Two studies showed that training related to EMRs has a significant association with the use of EMR systems. The odds of using EMRs were 3.41 times (AOR=3.41, 95%CI 1.25-9.29) higher for health professionals who routinely received EMR training compared to those who did not. Random effects model analysis was carried out in this meta-analysis because the included studies were characterized by the existence of heterogeneity (I^2=82.3%; P=.02; Figure 6).

Furthermore, 2 studies indicated a significant association between using an EMR system and receiving managerial support. Health care professionals who got managerial support were 2.86 times (AOR=1.70, 95% CI 1.21-2.38) more likely to use EMR systems compared to those who did not get managerial support. There was no heterogeneity among the papers included in this meta-analysis (I^2=0%; P=71). As a result, a fixed-effect model analysis was carried out (Figure 7).

Figure 4. Forest plot displaying the association between younger age group and use of electronic medical records among health professionals in Ethiopia. ES: Effect Size.

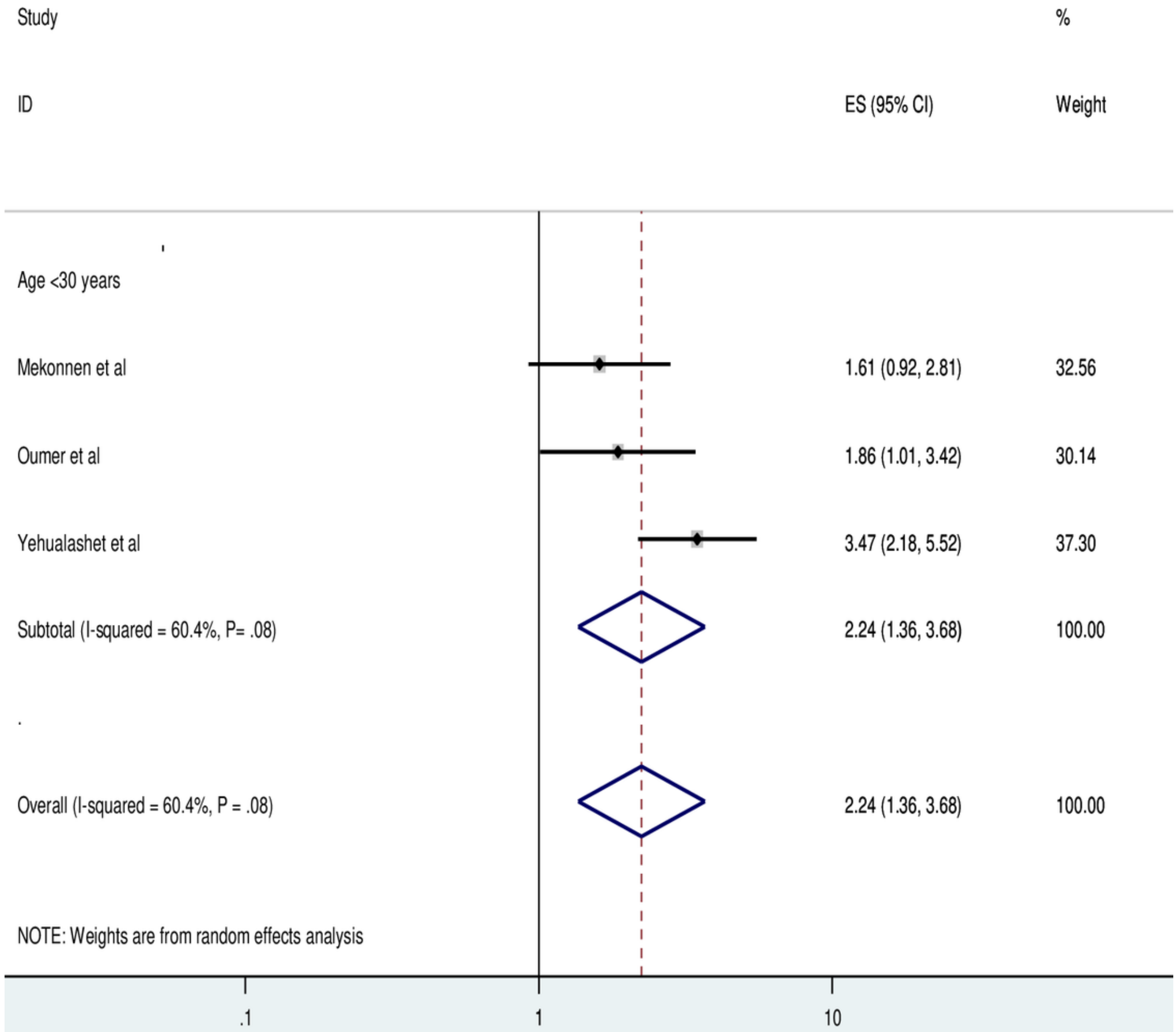


Figure 5. Forest plot displaying the association between availability of electronic medical record (EMR) manual and the use of EMR among health professionals in Ethiopia. ES: Effect Size.

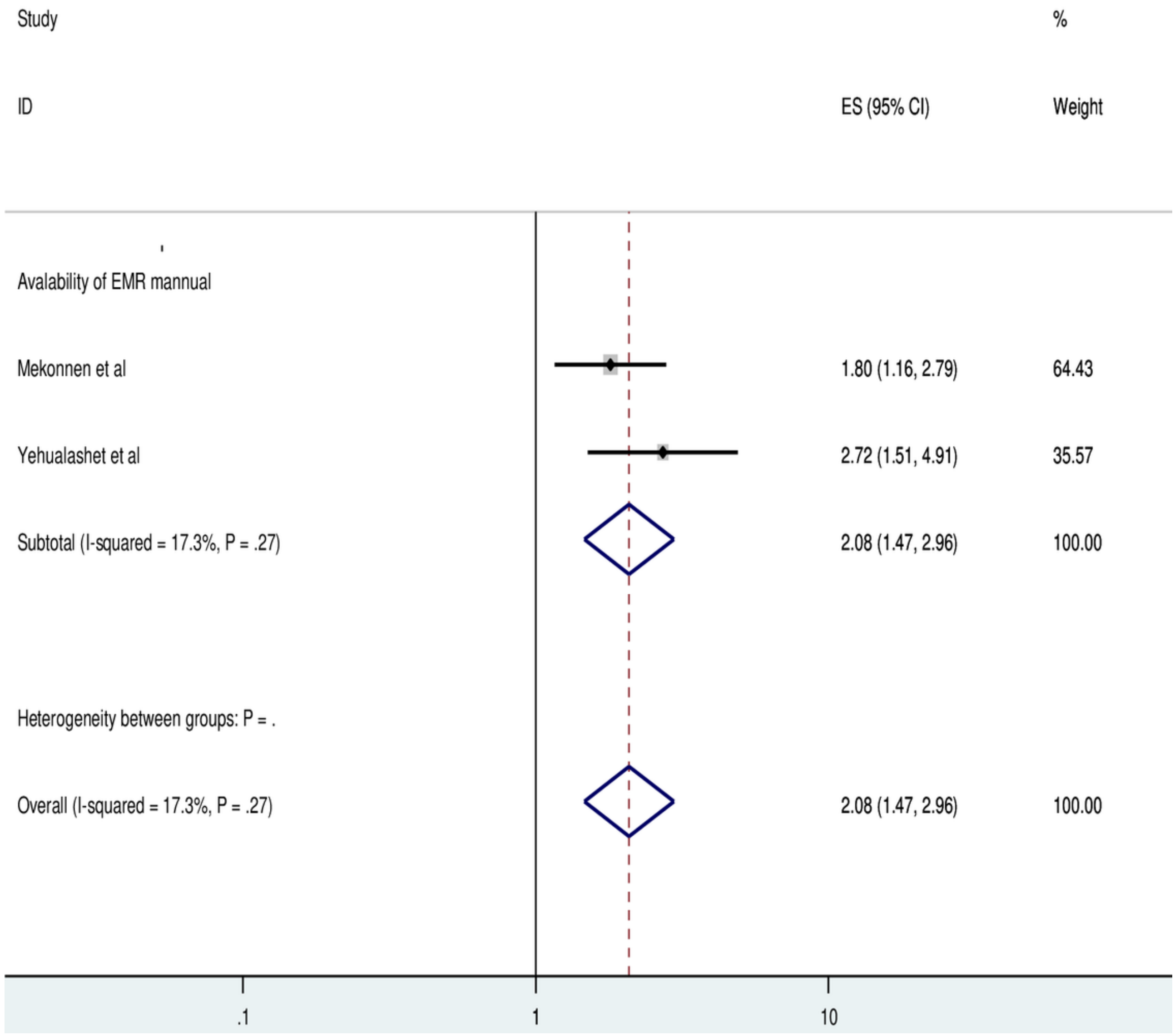


Figure 6. Forest plot displaying the association between electronic medical record (EMR) training and the use of EMR among health professionals in Ethiopia. ES: Effect Size.

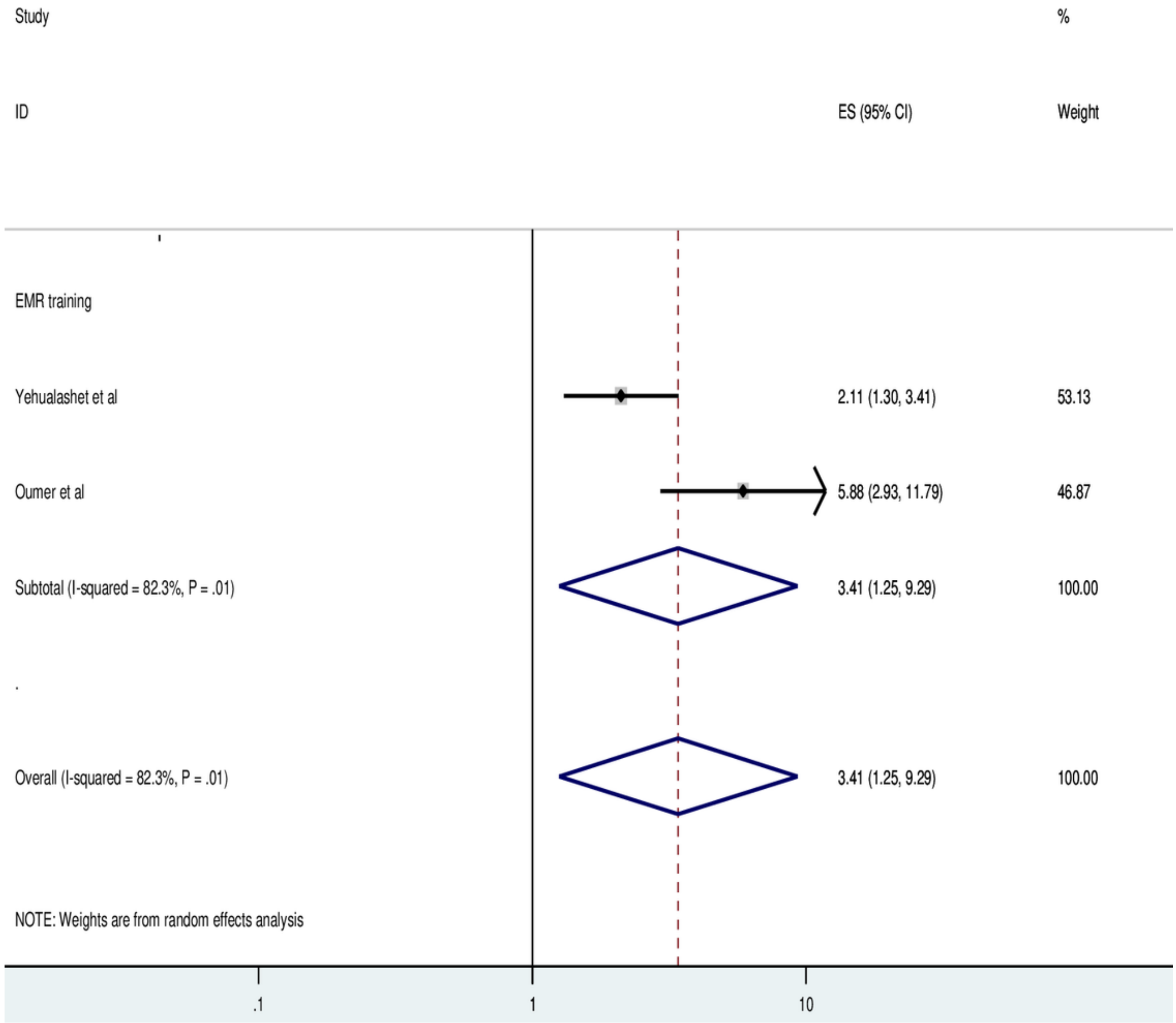
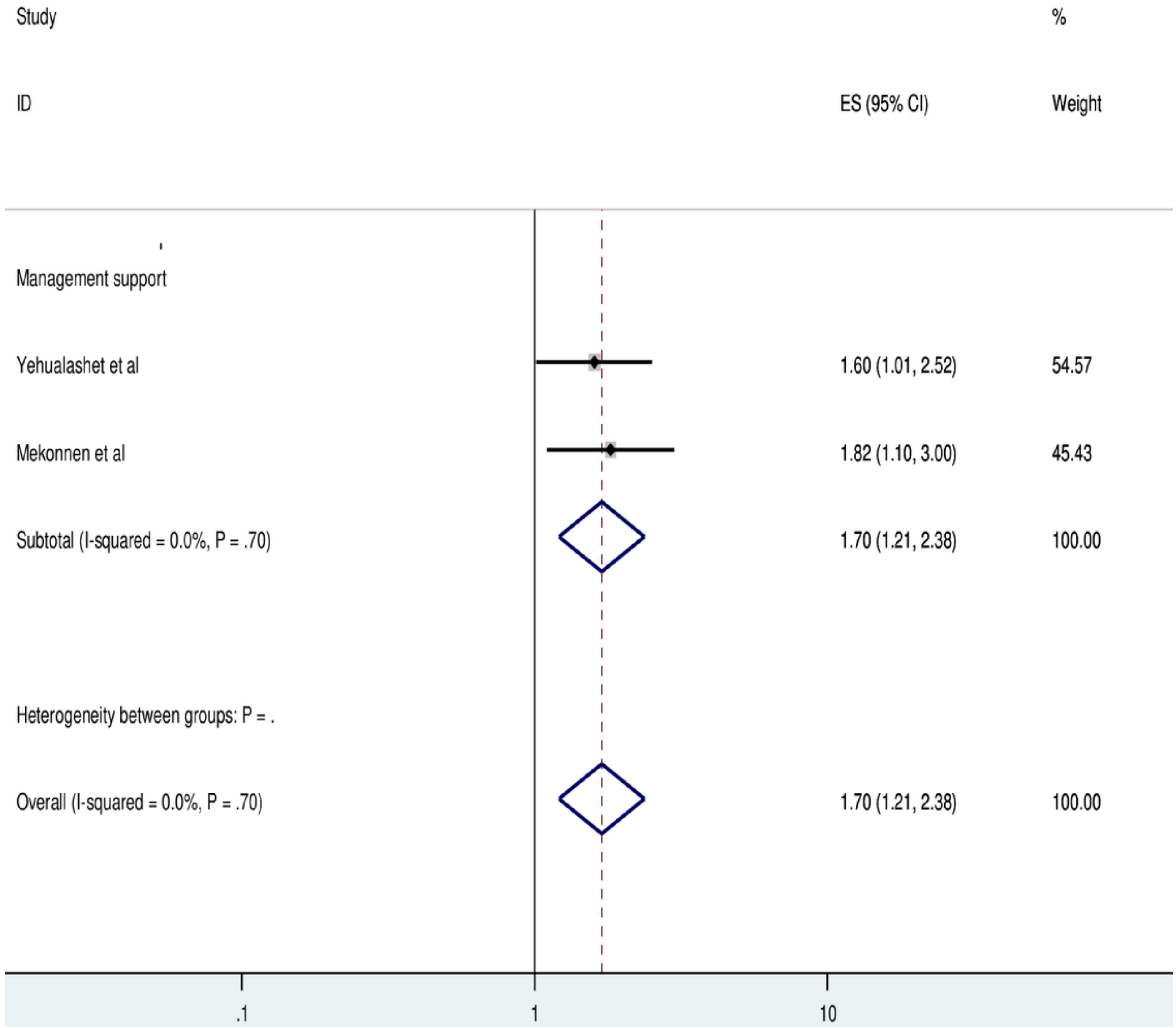


Figure 7. Forest plot displaying the association between availability of management support and the use of electronic medical records among health professionals in Ethiopia. ES: Effect Size.



Discussion

Principal Findings

This systematic review and meta-analysis investigated the use and determinants of the EMR system among health professionals in Ethiopia. Results revealed that the pooled estimate of EMR system use among health care professionals in Ethiopia was 51.85% (95% CI 37.14%-66.55%). We carried out a subgroup analysis based on the study site, where the studies were conducted. In the subgroup study, the northern Ethiopia region had the greatest rate of EMR utilization (58.75%), followed by the eastern parts of Ethiopia (54.99%). Similarly, we also carried out a subgroup analysis based on the year of publication of the original studies. We discovered disparities in publication timing, with current publications on the use of EMRs being higher in percentage (54.99%) than the studies published before 2016 (49.75%).

Furthermore, this analysis was conducted to identify determinants of EMR utilization in Ethiopia. The results showed that health care professionals younger than 30 years, health care professionals with access to an EMR manual, health care professionals with EMR-related training, and health care professionals with managerial support were found to have a positive association with the use of EMR in Ethiopia.

Comparison With Prior Work

Despite the lack of a meta-analysis on this topic of research, the use of the EMR system presented in this study is consistent with earlier individual studies conducted in Saudi Arabia (52%) [23]. Our results show a slightly lower rate of EMR use compared to those of studies done in Malawi on central hospitals, which showed that 68.8% of health workers used EMRs for collecting and analyzing clinical data [24]. However, the results of this study show a considerably lower rate of EMR use compared to those of studies conducted in industrialized nations, where the use of EMRs was 98% in Sweden, 88% in France, 88% in Germany, and 70% in Switzerland [25]. The discrepancy may be caused by disparities in information and communications technology infrastructure between lower- and higher-income nations, where in the case of lower-income countries, there is a power outage, limited access to standby generators, inadequate maintenance, and technical issues [21]. Furthermore, the lack of a standardized EMR system in Ethiopia and the health professionals' inadequate understanding and attitude toward EMRs may be contributing factors to Ethiopia's lower EMR use [26].

The northern Ethiopia region had a greater rate of EMR utilization, followed by the eastern part of Ethiopia, according to the subgroup analysis based on the study sites. This disparity could result from Ethiopia's northern region serving as a test site for implementing EMR systems. EMRs are used more frequently in the Ayder specialized hospital in northern Ethiopia than in the other studies that were taken into account [21]. Furthermore, the University of Gondar specialized hospital in northern Ethiopia served as a center of expertise for implementing EMRs [12].

The subgroup analysis based on publication year also revealed that studies published before 2016 showed lower EMR usage than recent studies. This could be attributed to the "Information Revolution," one of the transformative goals of the current Ethiopian federal ministry of health [27]. As a result, most hospitals have implemented EMRs and have mentorship and capacity-building programs for health care professionals regarding the use of health information and data quality [12,27].

This review showed that younger health professionals were more likely to use EMR than people in older age groups. Previous investigations have also supported these findings [2,21]. This might be because younger medical professionals are more open to adopting new technologies and have a better comprehension of information and communications technology than their older counterparts [28-30]. This implied that special attention should be paid to older health care professionals to boost the acceptance of the new health information technology.

This study also showed a strong link between using EMR systems and accessing an EMR manual. This suggests that health providers needed access to the EMR guidelines to promote the usage of EMR systems. Substantial evidence from various places supports this justification [28-30]. Our analysis also revealed that receiving management assistance has a significant association with the use of EMR systems. Previous research has shown that managerial support is the foundation for increasing the use of EMRs by health care workers [11,31]. This suggests that health administrators must work very hard to enhance the usage of EMRs and encourage their staff to use EMRs to make data-driven decisions that will raise the quality of health services.

Furthermore, our findings showed a strong correlation between receiving EMR training and using EMR systems. This result is consistent with earlier studies that discovered EMR system training positively impacted using the EMR system [32-34]. According to this finding, the health care system's adoption of health information technology may be strongly impacted by ongoing EMR and basic computer training. This recommends that the Ethiopian Ministry of Health should get ready to give thorough end-user training packages for medical staff to increase the level of EMR use and ensure its successful implementation.

Limitations

We are aware that there are certain limitations to this review. The review's primary challenge is the small number of included studies. Additionally, because of the varied categorization of factors in the included study, the pooled odds ratio for all variables associated with using EMRs by health workers was not evaluated. Furthermore, since all of the included papers were facility-based cross-sectional studies, the quality of the evidence and the generalization of the findings may be diminished. However, we tried to produce high-quality evidence by evaluating each included study's quality using 9 criteria from the Joanna Briggs Institute's quality evaluation tool.

Conclusions

The use of EMR systems in Ethiopia is relatively low. This study provides strong evidence for future implementers to pay close attention to improving health professionals' use of EMRs after implementation. This can be accomplished by making the EMR manual available to health practitioners, offering an EMR training program, and providing managerial support.

Acknowledgments

The authors recognize and appreciate the original publications included in this study and used them as a basis for this systematic review and meta-analysis.

Data Availability

The data analyzed during this meta-analysis and supplementary information are available in the published document.

Authors' Contributions

MDT was responsible for the conceptualization and design of the protocol, study selection, data extraction, statistical analysis, and preliminary paper versions. Data extraction, quality evaluation, and review were done by the authors in the following order: SMW, MSM, MHK, and HSD. BT and TMY both contributed to the conception and design of the protocol as well as the evaluation and modification of the paper. The paper's final draft, which was completed by MDT, was read and approved by all authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 19 KB - ijmr_v12i1e40721_app1.docx]

Multimedia Appendix 2
Quality score.
[DOCX File , 17 KB - ijmr_v12i1e40721_app2.docx]

References

1. Dutta B, Hwang HG. The adoption of electronic medical record by physicians: a PRISMA-compliant systematic review. Medicine (Baltimore) 2020 Feb;99(8):e19290 [FREE Full text] [doi: 10.1097/MD.0000000000019290] [Medline: 32080145]
2. Oumer A, Muhye A, Dagne I, Ishak N, Ale A, Bekele A. Utilization, determinants, and prospects of electronic medical records in Ethiopia. Biomed Res Int 2021 Nov 8;2021:2230618-2230611 [FREE Full text] [doi: 10.1155/2021/2230618] [Medline: 34790816]
3. Weed LL. Medical records that guide and teach. N Engl J Med 1968 Mar 14;278(11):593-600. [doi: 10.1056/nejm196803142781105]
4. Ariffin NABN, Ismail AB, Kadir IKA, Kamal JIA. IJARPED 2018 Aug 06;7(3):187-199. [doi: 10.6007/ijarped/v7-i3/4358]
5. Tegegne MD, Melaku MS, Shimie AW, Hunegnaw DD, Legese MG, Ejigu TA, et al. Health professionals' knowledge and attitude towards patient confidentiality and associated factors in a resource-limited setting: a cross-sectional study. BMC Med Ethics 2022 Mar 14;23(1):26 [FREE Full text] [doi: 10.1186/s12910-022-00765-0] [Medline: 35287659]
6. Grimson J. Delivering the electronic healthcare record for the 21st century. Int J Med Inform 2001 Dec;64(2-3):111-127. [doi: 10.1016/s1386-5056(01)00205-2]
7. Tilahun B, Fritz F. Comprehensive evaluation of electronic medical record system use and user satisfaction at five low-resource setting hospitals in ethiopia. JMIR Med Inform 2015 May 25;3(2):e22 [FREE Full text] [doi: 10.2196/medinform.4106] [Medline: 26007237]
8. Hamade N, Terry A, Malvankar-Mehta M. Interventions to improve the use of EMRs in primary health care: a systematic review and meta-analysis. BMJ Health Care Inform 2019 May 22;26(1) [FREE Full text] [doi: 10.1136/bmjhci-2019-000023] [Medline: 31142493]
9. Otieno GO, Hinako T, Motohiro A, Daisuke K, Keiko N. Measuring effectiveness of electronic medical records systems: towards building a composite index for benchmarking hospitals. Int J Med Inform 2008 Oct;77(10):657-669. [doi: 10.1016/j.ijmedinf.2008.01.002] [Medline: 18313352]
10. Mengesha T. Electronic solutions for Ethiopian health sectorlectronic medical record (EMR) system. Theseus.: theseus URL: https://www.theseus.fi/handle/10024/36264 [accessed 2022-12-25]
11. Biruk S, Yilma T, Andualem M, Tilahun B. Health Professionals' readiness to implement electronic medical record system at three hospitals in Ethiopia: a cross sectional study. BMC Med Inform Decis Mak 2014 Dec 12;14(1). [doi: 10.1186/s12911-014-0115-5]
12. Tilahun B, Gashu KD, Mekonnen ZA, Endehabtu BF, Asressie M, Minyihun A, et al. Strengthening the national health information system through a capacity-building and mentorship partnership (CBMP) programme: a health system and university partnership initiative in Ethiopia. Health Res Policy Syst 2021 Dec 09;19(1):141 [FREE Full text] [doi: 10.1186/s12961-021-00787-x] [Medline: 34886865]
13. MOH. Information revolution road map. Ethiopian Federal Ministry of Health. 2016. URL: http://repository.iifphc.org/bitstream/handle/123456789/316/Information%20Revolution%20Roadmap.pdf [accessed 2022-12-25]
14. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol 2009 Oct;62(10):e1-34 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.06.006] [Medline: 19631507]
15. Critical appraisal tools 2021. Institute JB. URL: https://jbi.global/critical-appraisal-tools [accessed 2022-12-25]
16. DerSimonian R, Kacker R. Random-effects model for meta-analysis of clinical trials: an update. Contemp Clin Trials 2007 Feb;28(2):105-114. [doi: 10.1016/j.cct.2006.04.004] [Medline: 16807131]
17. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002 Jun 15;21(11):1539-1558. [doi: 10.1002/sim.1186] [Medline: 12111919]
18. Sterne JA, Egger M. Funnel plots for detecting bias in meta-analysis. J Clin Epidemiol 2001 Oct;54(10):1046-1055. [doi: 10.1016/s0895-4356(01)00377-8]
19. Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Comparison of two methods to detect publication bias in meta-analysis. JAMA 2006 Feb 08;295(6):676-680. [doi: 10.1001/jama.295.6.676] [Medline: 16467236]
20. Mekonnen G, Tesfaye G, Zelalem B, Zelalem TMAT, Behailu H, Tesfaye D, et al. Utilization and Determinants of Electronic Medical RecordSystem among Health Professionals in Public Health Facilities of Harari Regional State, Eastern Ethiopia. J Health Med Informat 2021 Oct 23;12(10):1-6 [FREE Full text]

21. Yehualashet G, Asemahagn M, Tilahun B. The attitude towards and use of electronic medical record system by health professionals at a referral hospital in northern Ethiopia: Cross-sectional study. J Health Inform Afr 2015;3(1):19-29.
22. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 06;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
23. Nour El Din MM. Physicians' use of and attitudes toward electronic medical record system implemented at a teaching hospital in saudi arabia. J Egypt Public Health Assoc 2007;82(5-6):347-364. [Medline: 18706293]
24. Mkalira Msiska KE, Kumitawa A, Kumwenda B. Factors affecting the utilisation of electronic medical records system in Malawian central hospitals. Malawi Med J 2017 Sep 24;29(3):247-253 [FREE Full text] [doi: 10.4314/mmj.v29i3.4] [Medline: 29872515]
25. Share of primary care physicians in selected countries using electronic medical records (EMR) as of 2019. Statista. URL: https://www.statista.com/statistics/236985/use-of-electronic-medical-records/ [accessed 2022-06-22]
26. Yehualashet DE, Seboka BT, Tesfa GA, Demeke AD, Amede ES. Barriers to the adoption of electronic medical record system in Ethiopia: a systematic review. JMDH 2021 Sep;Volume 14:2597-2603. [doi: 10.2147/jmdh.s327539]
27. Taye G, Ayele W, Biruk E, Tassew B, Beshah T. The Ethiopian health information system: where are we? And where are we going? EJHD 2021 Jul 16;35(1):1-4.
28. Alwan K, Awoke T, Tilahun B. Knowledge and utilization of computers among health professionals in a developing country: a cross-sectional study. JMIR Hum Factors 2015 Mar 26;2(1):e4 [FREE Full text] [doi: 10.2196/humanfactors.4184] [Medline: 27025996]
29. Tegegne MD, Endehabtu BF, Guadie HA, Yilma TM. Health professionals' attitude toward the use of social media for covid-19 related information in northwest Ethiopia: a cross-sectional study. Front Public Health 2022 Jun 17;10:900293 [FREE Full text] [doi: 10.3389/fpubh.2022.900293] [Medline: 35784221]
30. Adedeji P, Irinoye O, Ikono R, Komolafe A. Factors influencing the use of electronic health records among nurses in a teaching hospital in Nigeria. JHIDC 2018 Jul 30;12(2):1-20.
31. Mohamed-Arraid A. Information needs and information seeking behaviour of Libyan doctors working in Libyan hospitals.: Loughborough University; 2011 Jan 01. URL: https://repository.lboro.ac.uk/articles/thesis/Information_needs_and_information_seeking_behaviour_of_Libyan_doctors_working_in_Libyan_hospitals/9414755 [accessed 2011-12-23]
32. Bredfeldt CE, Awad EB, Joseph K, Snyder MH. Training providers: beyond the basics of electronic health records. BMC Health Serv Res 2013 Dec 2;13(1). [doi: 10.1186/1472-6963-13-503]
33. Sharp K, Williams M, Aldrich A, Bogacz A, Denier S, McAlearney A. Conversion of provider EMR training from instructor-led training to eLearning at an academic medical center. Appl Clin Inform 2017 Dec 20;08(03):754-762. [doi: 10.4338/aci-2017-03-cr-0040]
34. Pantaleoni J, Stevens L, Mailes E, Goad B, Longhurst C. Successful physician training program for large scale EMR implementation. Appl Clin Inform 2017 Dec 19;06(01):80-95. [doi: 10.4338/aci-2014-09-cr-0076]

Abbreviations

AOR: adjusted odds ratio
EMR: electronic medical record
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by A Mavragani; submitted 02.07.22; peer-reviewed by Z Zandesh, M Behzadifar, J Shull; comments to author 18.10.22; revised version received 31.10.22; accepted 17.11.22; published 11.01.23.

Please cite as:
Tegegne MD, Wubante SM, Kalayou MH, Melaku MS, Tilahun B, Yilma TM, Dessie HS
Electronic Medical Record System Use and Determinants in Ethiopia: Systematic Review and Meta-Analysis
Interact J Med Res 2023;12:e40721
URL: https://www.i-jmr.org/2023/1/e40721
doi:10.2196/40721
PMID:36630161

Review

Cost-effectiveness of Digital Tools for Behavior Change Interventions Among People With Chronic Diseases: Systematic Review

Tun Lin Kyaw[1], MBBS, MPH; Nawi Ng[1,2], MD, MPH, PhD; Margarita Theocharaki[2], MPH; Patrik Wennberg[3], MD, PhD; Klas-Göran Sahlen[1], PhD

[1]Department of Epidemiology and Global Health, Faculty of Medicine, Umeå University, Umeå, Sweden

[2]School of Public Health and Community Medicine, Institution of Medicine, University of Gothenburg, Göteborg, Sweden

[3]Department of Public Health and Clinical Medicine, Family Medicine, Umeå University, Umeå, Sweden

Corresponding Author:
Tun Lin Kyaw, MBBS, MPH
Department of Epidemiology and Global Health
Faculty of Medicine
Umeå University
Universitetstorget 4
Umeå, SE-901 87
Sweden
Phone: 46 764229739
Email: tunlinnkyawrcs@gmail.com

Abstract

Background: Chronic diseases, including cardiovascular diseases, diabetes, chronic obstructive pulmonary disease, and cerebrovascular diseases, contribute to the most significant disease burden worldwide, negatively impacting patients and their family members. People with chronic diseases have common modifiable behavioral risk factors, including smoking, alcohol overconsumption, and unhealthy diets. Digital-based interventions for promoting and sustaining behavioral changes have flourished in recent years, although evidence of the cost-effectiveness of such interventions remains inconclusive.

Objective: In this study, we aimed to investigate the cost-effectiveness of digital health interventions for behavioral changes among people with chronic diseases.

Methods: This systematic review evaluated published studies focused on the economic evaluation of digital tools for behavioral change among adults with chronic diseases. We followed the Population, Intervention, Comparator, and Outcomes framework to retrieve relevant publications from 4 databases: PubMed, CINAHL, Scopus, and Web of Science. We used the Joanna Briggs Institute's criteria for economic evaluation and randomized controlled trials to assess the risk of bias in the studies. Two researchers independently screened, assessed the quality, and extracted data from the studies selected for the review.

Results: In total, 20 studies published between 2003 and 2021 fulfilled our inclusion criteria. All the studies were conducted in high-income countries. These studies used telephones, SMS text messaging, mobile health apps, and websites as digital tools for behavior change communication. Most digital tools for interventions focused on diet and nutrition (17/20, 85%) and physical activity (16/20, 80%), and a few focused on smoking and tobacco control (8/20, 40%), alcohol reduction (6/20, 30%), and reduction of salt intake (3/20, 15%). Most studies (17/20, 85%) used the health care payer perspective for economic analysis, and only 15% (3/20) used the societal perspective. Only 45% (9/20) of studies conducted a full economic evaluation. Most studies (7/20, 35%) based on full economic evaluation and 30% (6/20) of studies based on partial economic evaluation found digital health interventions to be cost-effective and cost-saving. Most studies had short follow-ups and failed to include proper indicators for economic evaluation, such as quality-adjusted life-years, disability-adjusted life-years, lack of discounting, and sensitivity analysis.

Conclusions: Digital health interventions for behavioral change among people with chronic diseases are cost-effective in high-income settings and can therefore be scaled up. Similar evidence from low- and middle-income countries based on properly designed studies for cost-effectiveness evaluation is urgently required. A full economic evaluation is needed to provide robust evidence for the cost-effectiveness of digital health interventions and their potential for scaling up in a wider population. Future studies should follow the National Institute for Health and Clinical Excellence recommendations to take a societal perspective, apply discounting, address parameter uncertainty, and apply a lifelong time horizon.

(Interact J Med Res 2023;12:e42396) doi:10.2196/42396

KEYWORDS

digital tools; chronic diseases; cost-effectiveness; lifestyle; behavior; systematic review; mobile phone

Introduction

Background

Chronic diseases are long-lasting conditions that do not improve or cure completely over time. Chronic diseases are the leading cause of death worldwide. According to the World Health Organization, ischemic heart disease, stroke, and chronic obstructive pulmonary disease (COPD) are the top 3 causes, whereas diabetes mellitus (DM) is the ninth leading cause of death globally [1]. In the Global Burden of Disease study (2016), disability-adjusted life-years (DALYs) because of ischemic heart disease, cerebrovascular disease, and lower respiratory infections accounted for 16.1% of all DALYs [2]. Approximately 10% of the adult population (≥40 years) had COPD [3]. In recent decades, the disease burden has shifted sharply toward noncommunicable diseases (NCDs) and injuries [4]. Between 1999 and 2019, ischemic heart disease, diabetes, stroke, chronic kidney disease, lung cancer, and age-related hearing loss showed the most substantial absolute increase in the number of DALYs, giving rise to the largest burden of disease in older age groups. Although there are several chronic diseases, this study focused on 4 major NCDs: cardiovascular diseases (CVDs), cerebrovascular diseases, COPD, and DM.

These chronic diseases share several risk factors, including tobacco use, unhealthy diet, physical inactivity, and excessive alcohol consumption [5]. The World Health Organization also highlighted that high systolic blood pressure (BP), tobacco use, dietary risks (eg, low intake of fruits and vegetables and high salt intake), air pollution, high fasting plasma glucose, high BMI, and high low-density lipoprotein cholesterol are the major risk factors responsible for millions of deaths worldwide [6]. Over the past decades, global exposure to several highly preventable risks has risen by >0.5% annually (obesity, high blood sugar, alcohol use, and drug use); these factors contribute not only to the growing burden of NCDs but also to the risk factors for a growing number of fatalities and highlight the necessity for investments in public health [7].

In addition to having direct consequences for persons with chronic diseases, chronic physical illnesses may also distort the lives of their families. A study in the Netherlands showed that chronic diseases negatively impact their partners in good health in 4 main areas: personal life, social relations, finance, and intrinsic rewards [8].

Today, smartphone use and internet access have increased significantly, providing the potential to improve health through the use of information technology. The term digital health intervention refers to interventions delivered using digital technologies such as smartphones, websites, and SMS text messages to provide effective, cost-effective, safe, and scalable interventions to enhance health and health care and promote healthy behaviors [9]. Developing complex health service interventions involves the use of behavior change techniques (BCTs). A BCT is "an observable, replicable, and irreducible component of an intervention designed to alter or redirect causal processes that regulate behavior, that is, a technique is proposed to be an active ingredient" [10]. In the National Institute for Health and Clinical Excellence (NICE) guidelines, interventions for changing unhealthy individual behaviors, such as unhealthy diet, physical inactivity, alcohol overconsumption, unsafe sexual practices, and smoking, are recommended to use evidence-based BCTs strategies such as goal-setting, feedback, and social support [11]. A previous systematic review and meta-analysis concluded that digital health interventions using smartphones, PCs, and wearable devices combined with technologies such as software, mobile apps, and the internet improve healthy behavioral factors such as physical activity (PA), diet, and medication compliance [12].

Knowledge Gap

Despite the well-established evidence of behavioral lifestyle interventions on chronic disease–related morbidity and mortality, particularly when implemented at a population level or in high-risk groups [13], evidence on the cost-effectiveness of digital-based health interventions for NCD prevention and control is inconclusive. Available studies on economic analyses of digital health interventions have shown conflicting evidence and inconsistent findings. One systematic review published in 2002 argued that telemedicine is not a cost-effective method of delivering health care [14], whereas the systematic review by Rojas and Gagnon in 2008 confirmed that telemedicine is cost-effective in general, as it reduces hospital use and improves patient compliance, satisfaction, and quality of life [15]. To the best of our knowledge, there is no cost-effectiveness study combining digital tools and behavioral changes for chronic diseases. This study aimed to determine whether digital tools are cost-effective for lifestyle behavior interventions.

Sustainable Behavior Change for Health Supported by Person-Tailored, Adaptive, Risk-Aware Digital Coaching in a Social Context Project

This study was part of the Sustainable Behavior Change for Health Supported by Person-Tailored, Adaptive, Risk-Aware Digital Coaching in a Social Context (STAR-C) project. It is an interdisciplinary research program aimed at developing and assessing a technical platform that can be used for behavior change interventions targeting CVD prevention through digital coaching. A team of researchers from complementary fields, such as public health, social science, computer science, cardiology, and health economics, designed and implemented this project [16]. The project will run in two phases from 2019 to 2024: (1) a formative intervention design and development phase and (2) an intervention evaluation phase. STAR-C will use gender and equity lenses in all phases of the program [17].

This study assessed the cost-effectiveness of digital health interventions for risk-reduction behavior and provided evidence-based recommendations regarding the health economic evaluation for the STAR-C project.

Methods

Overview

We conducted this review following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. We used the Population, Intervention, Comparator, and Outcomes (PICO) framework to develop the review question to ensure that the relevant components of the question are well defined [18]. This review considered (1) studies that included adults with one or more of the 4 chronic diseases (CVDs, DM, COPD, and cerebrovascular diseases); (2) studies with economic evaluations using digital tools (telemedicine, mobile health [mHealth] apps, web-based, SMS text messaging, telephone consultations in combination with other digital support); and (3) studies that included behavior change interventions (quitting smoking, exercising optimally, taking a healthy diet, and reducing alcohol consumption). The comparators were no intervention, usual care, current practice, counselor-based counseling, or pharmacologic therapy. The following 4 major risk factors for chronic diseases were considered in this study: smoking or tobacco, overconsumption of alcohol, physical inactivity, and unhealthy diet (low intake of fruits and vegetables and excessive salt intake).

Studies were excluded if they were (1) systematic reviews or meta-analyses; (2) irrelevant publication types (editorials, letters, conference papers, commentary, case reports, study protocols, pilot studies, descriptive studies, and ecologic studies); (3) wrong study design (animal and in vitro trials and guidelines); (4) not published in English; (5) no information on outcomes (eg, pure economic studies without clinical or behavioral outcomes) or intervention costs (eg, those with only gross economic benefits were estimated); and (6) interventions using mass media, in addition to any deviation from PICO criteria.

Types of Health Economic Evaluation

This review considers both partial and full health economic evaluations. According to Drummond, full economic evaluation is defined as a comparative analysis of alternative courses of action in terms of both their costs (resource use) and consequences (effectiveness), such as cost-benefit analysis (CBA), cost-effectiveness analysis (CEA), and cost-utility analysis (CUA) [19]. Partial economic evaluations either focus solely on costs or resource use without considering costs related to outcomes or focus on both costs and outcomes without comparing alternative interventions such as cost comparison or cost analysis, cost consequence analysis, cost description, outcome description, and cost of illness study [20].

Search Strategies

We developed search strategies based on the PICO framework to retrieve the relevant publications. Accordingly, we created 4 separate search blocks, each based on one of the 4 topics: cost-effectiveness, behavior change, digital health intervention, and chronic conditions under study. Controlled vocabulary, including Medical Subject Headings and keywords, was also used in the search to ensure that as many relevant articles as possible were identified using synonyms and truncations in every search block. We used a Boolean operator to expand, exclude, or join keywords, using "AND" and "OR." We searched the following 4 main bibliographic databases: PubMed, CINAHL, Scopus, and Web of Science. In addition to the web-based search, we manually conducted an extensive literature search using references from retrieved articles or recent results of ongoing studies identified from the database searches. Interested readers can find the detailed search blocks and terms in Multimedia Appendix 1.

Study Selection

Initially, retrieved articles from the 4 databases were imported into Endnote on the web, a citation manager, where we removed duplicates before exporting the search results to Rayyan [21], a web-based platform to facilitate collaborative systematic review processes. First, we screened the titles and abstracts of all the search results, guided by our inclusion and exclusion criteria. If a paper was rejected, we recorded the reasons for exclusion. We downloaded all included articles for full-text reviews after the first screening. The full-text papers were again reviewed against the eligibility criteria (Multimedia Appendix 2). Two independent reviewers thoroughly scanned the titles, abstracts, and full texts. Reviewers then compared their independent decisions for inclusion, and disagreements during the review processes were resolved by discussion between the reviewers. The "blind on" option on Rayyan made it impossible to see the decision of another reviewer on a particular abstract, which helped reduce the risk of selection bias during screening.

Data Extraction

We extracted data from each selected paper using a data-extraction form. These data included author, setting (country and year), inclusion and exclusion criteria, intervention and follow-up length of the study, number of participants in the intervention and control groups, economic perspective, uncertainty consideration (discounting and sensitivity analysis), outcomes, results, and type of behavioral interventions (Multimedia Appendix 3).

Quality Review (Risk of Bias)

We appraised the quality of all included papers using the Joanna Briggs Institute criteria for economic evaluation and randomized controlled trials (RCTs; Multimedia Appendices 4 and 5). The economic quality criteria considered were the type of economic study, appropriate valuation of economic and clinical outcomes, uncertainty consideration (discounting), appropriate conclusions, and conflicts of interest. For the RCT criteria, this study considered the similarity of both groups at baseline, the same outcome in both groups, and the appropriate analysis. In terms of economic study design, this study rated full economic evaluation (CEA, CUA, and CBA) as *high quality* and others as *low quality*. A study was rated high quality if it used actual costs rather than estimated costs. The economic outcome of the study should be feasible for full economic analysis (eg, cost per quality-adjusted life-years [QALY] or DALY, cost per life-year saved, cost per clinical outcome, etc) to produce good quality. If the study period was >1 year, discounting should be included. This study used the NICE scale from the lowest to highest risk of bias to provide a qualitative appraisal [22]. The review used 10 criteria (a combination of economic and RCT criteria) for quality appraisal. Studies with ≥3 unfavorable responses (eg, *no*) were considered a *high risk* of bias. In comparison, we considered studies with 2 unfavorable responses a *medium risk*, and studies with 1 or no unfavorable responses were considered a *low risk*.

Cost-effectiveness Appraisal

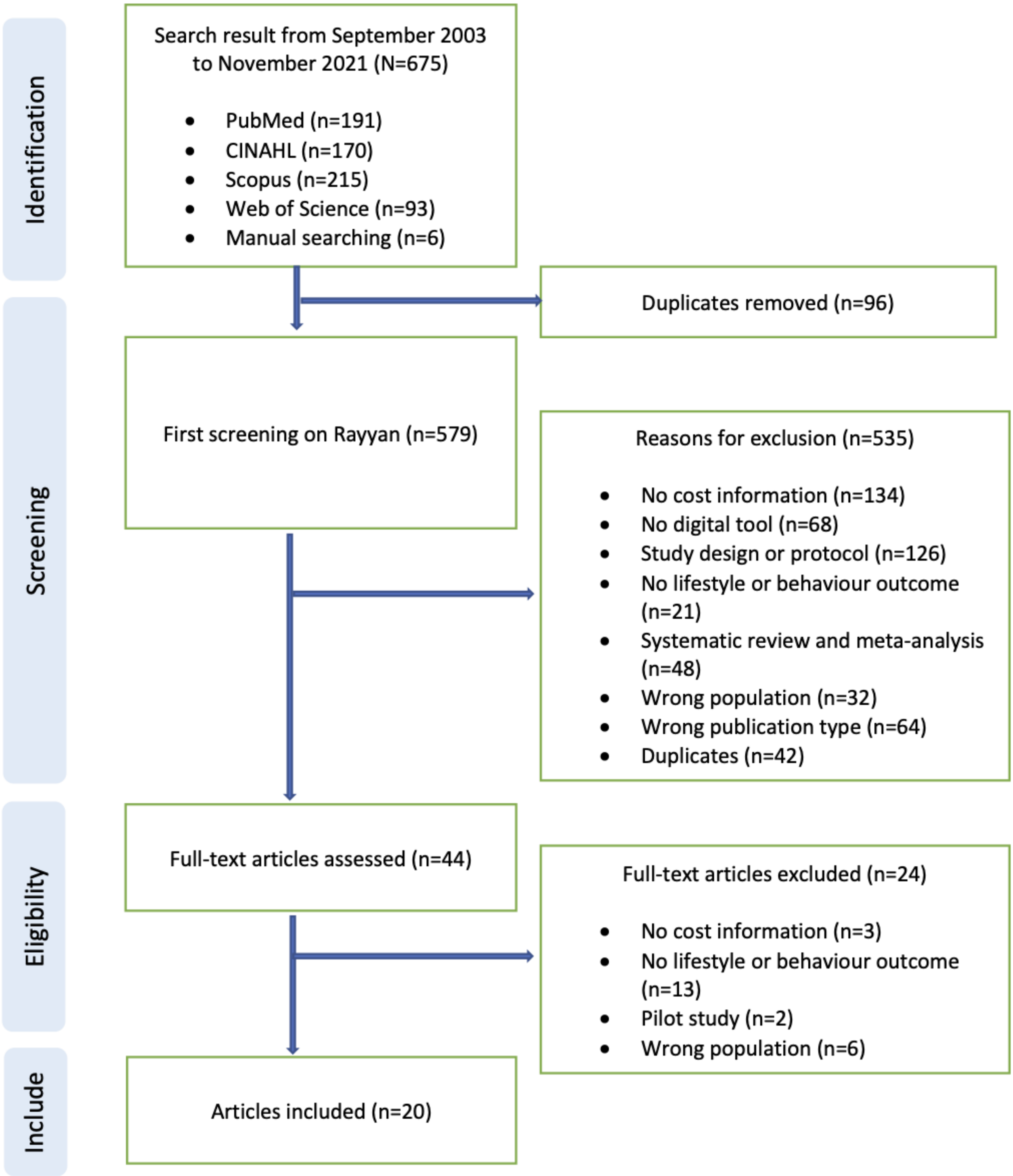
We assessed the cost-effectiveness of each study based on the cost-effectiveness threshold (CET) determined per country. Because it was impossible to determine the cost-effectiveness for partial economic studies, this study used the term *cost saving* or *not cost saving*, as stated in the respective study. The cost-effectiveness appraisal in this review was based entirely on conclusions of the respective studies.

Results

Study Selection and Characteristics

A total of 675 papers appeared in the initial search results, of which 44 (6.5%) papers were eligible for full-text review, and 20 (3%) papers were included. In general, studies were excluded if they had no cost data, had no digital tools, had no lifestyle or behavior outcomes, had an inappropriate study design, or were study protocols (Figure 1).

Figure 1. PRISMA (Preferred Reporting Item for Systematic Reviews and Meta-Analyses) flow diagram of studies showing reasons for exclusion.



Country of Origin

Most papers (12/20, 60%) in this review were from the United States [23-34]. The rest were from Australia (4/20, 20%) [35-38], New Zealand (1/20, 5%) [39], the United Kingdom (1/20, 5%) [40], Italy (1/20, 5%) [41], and 5% (1/20) of studies conducted in 3 countries (the Netherlands, Spain, and Taiwan) [42]. All the studies (20/20, 100%) were conducted in high-income countries. The period of publication of the studies ranged from 2003 to 2021; however, most studies (16/20, 80%) were published after 2010.

Disease Area and Patient Population

Most studies (12/20, 60% and 4/20, 20%) focused on CVD [24-26,29,32-35,38-40,42] (hypertension, ischemic heart disease, myocardial infarction [MI], and heart failure) and DM [28,30,31,37,41]. Overall, 15% (3/20) of other studies focused on CVD and DM [23,36,37], and only 5% (1/20) of studies focused on COPD [27]. No studies on cerebrovascular diseases were included, as none met the eligibility criteria. In most studies (18/20, 90%), the participants were those with one or more of the 4 chronic diseases. Furthermore, 10% (2/20) of studies [23,29] focused on people with a high risk of CVD and

DM, measured by Framingham Risk Score [43], which included scoring on age, blood lipid profiles, smoking status, and hypertension (which is one of the subcategories of CVD).

Studies on DM included patients with type 1 diabetes (1/20, 5%), type 2 diabetes (2/20, 10%), and DM of nonspecific type. Generally, studies on individuals with severe diseases, complications, comorbidities, or who cannot exercise or have no mobile phone or internet access were excluded. The participants in these studies were aged 18 to 89 years, but 10% (2/20) of studies focused on the older adult (≥60 years) population [23,27].

Comparator

We included studies comparing digital health interventions with an alternative strategy representing the existing method of providing health services to the study population or on intervention. Most studies (15/20, 75%) compared key interventions with usual care, home health care, or existing practices. Some studies (3/20, 15%) used health education at the clinic, counselor-delivered counseling, or pharmacological therapies as comparators [26,29,30], and only 10% (2/20) of studies compared interventions with no intervention [23,35].

Study Design

Of the 20 studies, 9 (45%) performed full health economics analysis [25,29,30,35,36,38-40,42] using CEA and CUA methods, whereas the remaining 11 (55%) were partial economics studies. Furthermore, 60% (12/20) of studies used RCT design and incorporated economic evaluation. In full health economics studies (7/20, 35%), CUA, which used the QALY as the outcome measure, was the most common method. Only 35% (7/20) of studies [23,25,30,33,35,36,42] used modeling methods such as Markov modeling, event-based simulation, and decision trees. Other studies (13/20, 65%) were embedded in RCT studies.

Economic Perspective

An evaluation must specify and justify the perspective taken to measure behavior or lifestyle change programs and health resource use. A societal perspective is recommended by NICE [22], as the goal of public health is to improve the health and well-being of the whole population. Most studies (17/20, 85%) used health care payers as study perspectives, and only 15% (3/20) used a societal perspective [29,41,42].

Time Horizon

As this review focuses on chronic diseases, a longer time horizon is needed to measure the effects of costs and health outcomes. UK NICE guidelines prefers a lifetime horizon [22]. All studies (20/20, 100%) had a range of time horizons from 6 months to lifelong. In only 10% (2/20) of studies [35], the time horizon was a lifetime; in 10% (2/20) others [23,36], it was 10 years. Most studies (9/20, 45%) did not mention the time horizon, while for 30% (6/20) of studies, it was between 1 and 5 years; 5% (1/20) of studies used 6 months as the time horizon [38].

Direct Costs Included

Program-specific costs, a measure of program administration, program delivery, and program capital costs (eg, the technology needed for web-based interventions), are required. Health care costs, that is, the cost of all relevant health care services, such as general practitioner visits, specialist visits, hospitalizations, diagnostic tests and investigations, medications, and specialized equipment, must be calculated. The actual cost should be based on invoices, receipts, administrative records, and the hospital register rather than patient-estimated costs. All studies (20/20, 100%) in this review used both programs and direct medical costs in their calculations. Program costs differed significantly depending on the country, type, and year of intervention [24]. To make reading easier, all currencies other than US $ are always accompanied by the conversation to US $ (converted values in parentheses).

Indirect Costs Included

Studying costs from a societal perspective requires indirect costs, which include the patient's or caregiver's productivity loss owing to disease or travel time of the patient to health care services, as well as other home care costs. Of the 20 studies, only 3 (15%) studies [29,41,42] that used a societal perspective included the indirect costs.

Economic Outcomes

The incremental costs and outcomes of each health care program must be assessed in an economic evaluation. Accordingly, of the 20 studies, 7 (35%) studies using CUA methods presented incremental cost-effectiveness ratio (ICER) values based on the cost per QALY gained to assess the cost-effectiveness of the intervention. Furthermore, 10% (2/20) of other CEA studies showed ICER values using cost per life-year saved and cost per mm Hg reduction in BP. Although the remaining 55% (11/20) of studies did not provide cost-effectiveness information, it is still valuable to determine whether a treatment is justified based on its cost. Reduced use of health care resources is interpreted as evidence of improved outcomes in these studies, and it is usually presented as the average cost-savings per patient.

Sensitivity Analysis

Economic assessments should consider at least one sensitivity analysis to determine the robustness of the study results [44]. Nearly half of this review's studies (9/20, 45%) performed sensitivity analyses, whereas the remaining studies (11/20, 55%) did not. Of the 9 studies with sensitivity analysis, 5 (56%) studies [25,35,36,39,40] performed probabilistic sensitivity analysis by the Monte Carlo simulation method.

Generalizability of the Result

Of the 20 studies, only 5 (25%) studies [28,35,37,40,42] discussed that their findings could be generalized to other populations, whereas the other 5 (25%) studies [25,26,29,34,36] did not. In the remaining studies (10/20, 50%), generalizability was not mentioned.

A description of the study characteristics, the economic perspective of the interventions, the results and the cost-effectiveness appraisal of selected studies can be found in Table 1.

Table 1. Study characteristics, economic perspective, results, and cost-effectiveness appraisal.

Study	Country and year	Study population	Follow-up[a]	Key intervention	Control	Perspective	Results	Cost-effectiveness
Bertuzzi et al [41]	Italy, 2017	Patients aged 5-50 years with type 1 DM[b] and internet access	1 year	Teleconsultation, tele-education (nutrition, medication, and self-management; n=35)	Usual care (n=39)	Societal	1. No difference in HbA_{1c}[c] 2. Reduced DM complications 3. Saving of €80 per visit (US $89 per visit)	Inconclusive
Burn et al [35]	Australia, 2017	Patients with CHD[d] or MI[e] or bypass graft surgery	5 years	SMS Text message for behavior change over 24 weeks (n=5000)	No intervention	Health care	1. Reduced occurrence of MI and strokes 2. ICER[f]: Aus $6123 per QALY[g] (US $4,648 per QALY)	Cost-effective
Chen et al [23]	United States, 2016	Overweight or obese older adults (≥65 years) with risks for DM or CVD[h] (by FRS[i])	10 years	16 weeks of web-based education for behavior change (n=997)	No intervention	Health care	1. Saving of US $13,240 per capita at 10 years for prediabetes 2. Saving of US $12,840 per capita at 10 years for pre-CVD	Cost-saving
Copeland et al [24]	United States, 2010	Patient ≥18 years with CHF[j]	1 year	Telephone coaching for behavior change (n=220)	Usual care (n=238)	Health care	1. No difference in clinical outcomes 2. Higher total cost in the intervention group (US $6165) 3. More regular exercise (OR[k] 1.94, 95% CI 1.08-3.49)	Not cost-saving
Datta et al [25]	United States, 2010	Patients with hypertension and taking antihypertensive medication	2 years	Telephone coaching for behavior change (n=294)	Usual care (n=294)	Health care	1. No group difference in BP[l] control 2. ICER: US $42,457-US $87,300 per life-year saved	Cost-effective
Dunagan et al [26]	United States, 2005	Patient aged ≥21 years, at least one sign and symptom of heart failure	1 year	A phone call to improve self-management (diet and adherence to therapy) plus education about signs and symptoms of heart failure (n=76)	Education for heart failure (n=75)	Health care	1. Time to hospitalization (HR[m] 0.67, 95% CI: 0.47-0.96) 2. Hospital readmission (HR 0.67, 95% CI 0.46-0.99) 3. Lowered hospital days and costs in the first 6 months only	Inconclusive

Study	Country and year	Study population	Follow-up[a]	Key intervention	Control	Perspective	Results	Cost-effectiveness
Finkelstein et al [27]	United States, 2006	Patients aged 60-96 years with CHF, chronic obstructive pulmonary disease, and chronic wound	2.5 years	Video group: HHC[n] + 2 video consultations, Monitoring group: HHC + 2 video consultations +monitoring symptoms (n=54)	HHC (n=19)	Health care	1. No difference in mortality 2. No difference in morbidity 3. Lower cost than the control group	Cost-saving
Fischer et al [28]	United States, 2012	People aged >17 years with diabetes	20 months	Telephone coaching for medication management and healthy behavior (n=381)	Usual care (n=381)	Health care	1. LDL[o] (AOR[p] 1.72, 95% CI 1.28-2.32) 2. Saving US $2433 per average patient cost 3. No difference in the number of admissions	Cost-saving
Graves et al [36]	Australia, 2009	Adults with type 2 DM or hypertension	10 years	Telephone counseling over 1 year for physical activity and diet (n=228)	Existing practice (n=206)	Health care	1. ICER: Aus $29,375 per QALY gained (US $23,466 per QALY) 2. CET[q]: 100% at a threshold of Aus $64,000 per QALY (US $51,126 per QALY)	Cost-effective
Hamar et al [37]	Australia, 2015	People aged 20-89 years with confirmed heart disease or DM; all under MGH[r] program coverage	4 years	Telephone coaching and web-based tool for self-management and behavior change (n=4948)	Usual care (n=28,520)	Health care	1. Hospital admission rate (AOR 0.73, 95% CI 0.69-0.78) 2. Readmission rate (AOR 0.55, 95% CI 0.48-0.63) 3. Hospitalization days (ARR[s] 0.83, 95% CI 0.77-0.90) 4. Saving Aus $3549 per patient per year (US $2732 per patient per y)	Cost-saving
Keyserling et al [29]	United States, 2014	Adults aged 35-79 years with moderate to high risk for CVD (by FRS)	1 year	Web-based counseling for healthy behavior and medication adherence (n=193)	Counselor-delivered counseling (n=192)	Societal	1. No difference in FRS, 2. ICER: US $2973 per QALY gained	Cost-effective
Maddison et al [39]	New Zealand, 2015	Patients with IHD aged ≥18 years and were able to perform the exercise	2 years	SMS text messaging and video messages via the website for exercise (n=85)	Usual care (n=86)	Health care	1. No difference in peaked O2 uptake 2. More physical activity, 3. More walking, 4. ICER: US $28,768 per QALY gained	Cost-effective

Study	Country and year	Study population	Follow-up[a]	Key intervention	Control	Perspective	Results	Cost-effectiveness
McManus et al [40]	United Kingdom, 2021	People with poorly controlled hypertension	1 year	Web-based counseling for self-monitoring, titration of drugs, and healthy behaviors (n=305)	Usual care (n=317)	Health care	1. No group difference in BP 2. ICER: £11 (US $13.27) per mm Hg reduction (95% CI £6-£29; [US $15] per mm Hg reduction)	Cost-effective
Nordyke et al [30]	United States, 2019	Patients aged 45-76 years with type 2 DM or hypertension	3 years	Digital therapeutic intervention using mobile phone app (n=2570)	Pharmacologic therapies (n=2575)	Health care	1. ICER: US $6468 per QALY for DM 2. ICER: US $6620 per QALY for hypertension	Cost-effective
Nundy et al [31]	United States, 2014	people ≥18 years with DM	6 months	Text message for self-care and 2 weeks web education on diet, exercise and medication (n=74)	Usual care (n=274)	Health care	1. HbA1c level: group difference: −0.4% (*P*=.01) 2. Cost-savings of US $437 per participant	Cost-saving
Piera-Jiménez et al [42]	The Netherlands, Spain, and Taiwan, 2020	Aged 18-75 years with hypertension or CHD or HF	5 years	SMS text messages and mobile apps for a healthy lifestyle over 6 months (n=120)	Usual care (n=118)	Societal	1. ICER: €124,489 per QALY (US $139,680 per QALY) in the Netherlands, €18,769 per QALY (US $21,059 per QALY) in Spain, €11,303 per QALY (US $12,682 per QALY) in Taiwan	Cost-effective for Spain, but not for the Netherlands and Taiwan
Southard et al [32]	United States, 2003	Patients with CHD or heart failure or both and access to the internet	6 months	Web-based education and email contact for exercise and diet over 6 months (n=53)	Usual care (n=51)	Health care	1. Fewer CVD events (15.7% reduction in intervention and 4.1% in the control group) 2. Saving of US $1418 per patient	Cost-saving
Wang et al [33]	United States, 2012	Patients with poorly controlled hypertension, and taking drugs	18 months	Telephone intervention for 1. healthy behavior, 2. medication management, and 3. both (n=444)	Usual care (n=147)	Health care	1. No difference in BP control 2. No difference in total costs	Not cost-saving
Maciejewski et al [34]	United States, 2014	Adults with hypertension medication and adults with poorly controlled hypertension	36 months	Telephone-delivered medication management, 2. software-assisted behavioral management, 3. combined over 18 months (n=444)	Usual care (n=147)	Health care		Not cost-saving

Study	Country and year	Study population	Follow-up[a]	Key intervention	Control	Perspective	Results	Cost-effectiveness
							1. BP control: (17.1% patients; 95% CI 6.9-27.4) and US $3237 saving in behavioral arm, 20.2% patients (95% CI: 9.7-30.6) and US $977 saving in medication arm, and 20.4% patients (95% CI 10-30.8) and US $303 saving in the combined arm. No difference in cost-saving	
Turkstra et al [38]	Australia, 2013	Patients aged 18-80 years with MI	1 years	Telephone coaching for self-monitoring, healthy behavior, and telemonitoring over 6 months (n=215)	Usual care (n=215)	Health care	1. No difference in HRQoL[t] 2. ICER: Aus $85,423 per QALY gained (US $82,072 per QALY)	Not cost-effective

[a]The follow-up time of trial.
[b]DM: diabetes mellitus.
[c]HbA_{1c}: hemoglobin A_{1c}.
[d]CHD: coronary heart disease.
[e]MI: myocardial infarction.
[f]ICER: incremental cost-effectiveness ratio.
[g]QALY: quality-adjusted life-years.
[h]CVD: cardiovascular disease.
[i]FRS: Framingham Risk Score.
[j]CHF: congestive heart failure.
[k]OR: odds ratio.
[l]BP: blood pressure.
[m]HR: hazard ratio.
[n]HHC: home health care.
[o]LDL: low-density lipoprotein.
[p]AOR: adjusted odds ratio.
[q]CET: cost-effectiveness threshold.
[r]MGH: My Health Guardian.
[s]ARR: adjusted risk ratio.
[t]HRQoL: health-related quality of life.

Evidence for Cost-effectiveness

Of the studies (9/20, 45%) with full economic evaluation, 7 (78%) studies concluded that using digital tools for behavior modification was cost-effective when the comparators were no intervention, usual care, counselor-delivered counseling, or pharmacologic therapies [25,29,30,35,36,39,40]; 6 (86%) studies concluded their cost-effectiveness from the health care payer perspective and 1 (14%) from the societal perspective [29]. Of the studies (11/20, 55%) with partial economic evaluations, 55% (6/11) of studies were cost-saving; 18% (2/11) of studies were inconclusive [26,41]; and 27% (3/11) of studies were not cost-saving [24,33,34].

Digital Tools for Intervention

The studies in this review used telephone, SMS text messaging, websites and software, mobile apps, and web-based video consultations as digital tools. The most cost-effective interventions (5/20, 25%) used telephone coaching, SMS text messaging, or health apps on mobile phones. Most studies (9/20, 45%) used telephones with other digital support as the tool for behavior change communication [24-26,28,33,34,36-38]. Typically, telephone interventions were provided by experienced nurses trained in motivational interviewing, but 22% (2/9) of these studies [36,38] used trained counselors and medical doctors.

Using a website to provide consultation or counseling for healthy behavior was the second most commonly used method in some studies (6/20, 30%) [23,29,32,34,40,41]. The studies involved a wide variety of health care professionals in web-based counseling. In addition, one study used email reminders to encourage exercise and incentives (key chains, athletic socks, book markers, and refrigerator magnets) to encourage active participation [32]; one study conducted in Italy used a website [45] for diabetes teleconsultation [41].

Another study used SMS text messaging for behavior change communication. Experts created automated messages that encouraged PA and a healthy diet for respective diseases and are typically sent out 3 to 5 times weekly [39]. In addition to behavior-related messages, they also reminded the patient about self-monitoring (eg, “time to check blood sugar”) [31].

The use of mobile apps, such as Moves, Vire, and Beddit, to encourage healthy behavior has been observed in 10% (2/20) of studies [30,42]. These apps were designed to integrate input from all monitoring devices, including pedometers that count steps, and the HORUS app collected pictures of the patients’ meals to provide dietary recommendations. These apps provided information to patients and create alerts for exercise [42]. Overall, 5% (1/20) of studies used video calls for internet-based visits and encouraged patients to exercise [27].

Types of Risk Behaviors Aimed by Interventions

Smoking and Tobacco Control

Of the 8 (40%) studies on smoking cessation interventions, 5 (62%) were conducted in the United States [25,28,29,33,34], 2 (25%) in Australia [35,38], and 1 (12%) in 3 countries (the Netherlands, Spain, and Taiwan) [42]. In total, 50% (4/8) of studies [25,29,35,42] concluded that smoking cessation interventions were cost-effective. In cost-effective interventions, the studies used web-based counseling, SMS text messaging, and telephone counseling as tools for behavior change. The SMS text messaging intervention (TEXT ME) was cost-effective in an Australian study using Markov simulation, with an ICER of Aus $6123 per QALY (US $4648 per QALY) gained when compared with no intervention with the CET of Aus $64,000 per QALY (US $51,125 per QALY) [35]. A study in the United States was cost-effective at an ICER of US $2973 per QALY gained when web-based counseling was compared with counselor-based counseling, given that the CET was US $100,000 per QALY. Another study in the United States compared telephone coaching for behavior change with usual care using life-year saved as an outcome measure and concluded that the intervention was cost-effective at an ICER of US $42,457 per life-year saved for women and US $87,300 per life-year saved for men [25]. One study in 3 countries showed that the intervention was cost-effective only in Spain with the ICER of €18,769 per QALY (US $21,059 per QALY) and not in the Netherlands and Taiwan [42].

Alcohol Reduction

The cost-effectiveness of alcohol reduction interventions was evaluated in only 30% (6/20) of studies that focused on people with MI, DM, or poorly controlled hypertension as study participants. Only 10% (2/20) of studies [25,40] confirmed that telephone coaching or web-based counseling for healthy behavior was more cost-effective than usual care. A study in the United Kingdom reported that the intervention was cost-effective at an ICER of £11 per mm Hg reduction (US $15 per mm Hg) in BP when the willingness-to-pay threshold was £20 per mm Hg reduction (US $28 per mm Hg) [40].

Salt Intake

In total, 15% (3/20) of studies considered salt intake control in their interventions and aimed at people with poorly controlled hypertension [33,34,40], and only 33% (1/3) of those studies showed that it was cost-effective [40].

PA Assessment

Most studies (16/20, 80%) included PA (exercise, walking, dancing, gardening, yoga, etc) in their interventions. Of the 16 studies, we found 7 (44%) studies to be cost-effective when we compared web-based counseling, SMS text messaging, and telephone counseling with no intervention, usual care, or counselor-led counseling [25,29,30,35,36,39,40]. Among cost-effective interventions, they used the telephone [25,36], SMS text messaging [35,39], websites [29,40], and mobile apps [30] as digital tools to encourage PA.

In an Australian study, PA improvement was measured as moderate PA engagement for ≥5 days per week for at least 150 minutes each time. It was also estimated that the total cost of telephone counseling was Aus $570 (US $460) for the first year and Aus $410 (US $330) per year for the next 10 years, and it was cost-effective with an ICER of Aus $29,375 per QALY (US $23,466 per QALY) gained, given that the willingness-to-pay threshold is Aus $64,000 per QALY (US $51,125 per QALY) [36]. In total, 10% (2/20) of studies used SMS text messaging to encourage PA, such as “the more you eat, the more you need to exercise” [35,39]. One study in New Zealand reported that SMS text messaging encouraged more leisure time PA (110.2 minutes per week) and more walking (151.4 minutes per week) in the intervention group and was cost-effective at an ICER of US $28,768 per QALY gained [39].

Using mHealth apps for PA was cost-effective, as measured by the ICER of US $6468 per QALY gained and US $6620 per QALY gained for digital interventions targeting people with DM and people with hypertension, respectively, when compared with pharmacological therapy [30]. According to an Australian study published in 2013, telephone coaching for PA was not cost-effective for patients with MI [38].

Of the 35% (7/20) of partial economic studies for PA, 6 (86%) studies [23,27,28,31,37] showed cost-savings with a wide range of values depending on the type of digital tools and country.

Diet and Nutrition

Most studies (17/20, 85%) evaluated the cost-effectiveness of diet and nutritional interventions, and 35% (6/17) of these studies found them to be cost-effective. These interventions used a web-based coaching [30,40], telephone coaching [25,36], mobile apps [30], and SMS text messaging [35] as digital tools targeting CVD (MI, ischemic heart disease, and hypertension) and type 2 DM. It had the same ICER values as those for PA.

Studies on behavioral interventions using telephone coaching for healthy diet and nutrition reported intervention costs of US $112 per participant in the United States [25] and Aus $570 (US $460) per participant in Australia [36]. Overall, 5% (1/20) of studies used mobile apps and SMS text messaging to promote a healthy diet. The HORUS application was designed to collect pictures of different meals of the patient to provide dietary recommendations [42].

According to studies with partial economic evaluations, 25% (5/20) of interventions for a healthy diet were cost-saving, and the value of the savings was the same as that for PA [23,28,31,32,37]. Overall, 10% (2/20) of studies reported that it was not cost-saving because of higher use of health care services among patients with heart failure and hypertension in intervention groups compared with usual care [24,33].

Risk of Bias Assessment

Table 2 presents the risk of bias across the selected studies. Four studies were deemed high risk [24,26,27,41], 6 medium risk [28,31-34,37], and 10 had a low risk of bias.

Nearly half of all studies (9/20, 45%) involved in this review had a potential conflict of interest because of stakeholder involvement in the analysis processes and unclear disclaimers [23,24,26,27,31-34,37]. Of these studies, 22% (2/9) had a serious risk of conflict of interest, as 1 author is the cofounder of mHealth Solutions company [31], and the other authors received consultation funds from pharmaceutical companies [34]. The remaining 55% (11/20) of studies were deemed to have no conflicts of interest. In total, 40% (8/20) of studies in this review showed cost-effective results without any conflicts of interest.

In this review, 20% (4/20) of studies [24,26,28,32] had unclear research questions regarding economic evaluation; 10% (2/20) of studies [27,41] had imprecise valuations, as they did not use actual costs in at least one of the cost categories.

Table 2. Quality appraisal (risk of bias assessment).

Study	Is the research question for economic evaluation?	Is the economic study design appropriate?	Are costs valued appropriately?	Are outcomes valued appropriately?	Was discounting applied?	Is the conclusion appropriate?	Is the conflict of interest disclosed?	Were groups similar at baseline?	Was the same outcome measured in both groups?	Is the analysis appropriate?	Risk of bias
Bertuzzi et al [41]	No	No	No	Yes	No	No	Yes	Yes	Yes	Yes	High
Burn et al [35]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	N/A[a]	N/A	N/A	Low
Chen et al [23]	Yes	No	Yes	Yes	Yes	Yes	No	N/A	Yes	N/A	Low
Copeland et al [24]	No	No	Yes	Yes	No	Yes	No	Yes	Yes	Yes	High
Datta et al [25]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	N/A	Low
Dunagan et al [26]	No	No	Yes	Yes	No	Yes	No	Yes	Yes	Yes	High
Finkelstein et al [27]	Yes	No	No	Yes	No	Yes	No	Yes	Yes	Yes	High
Fischer et al [28]	No	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Medium
Graves et al [36]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Low
Hamar et al [37]	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Medium
Keyserling et al [29]	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Low
Maddison et al [39]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Low
McManus et al [40]	Yes	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Low
Nordyke et al [30]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	N/A	N/A	N/A	Low
Nundy et al [31]	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Medium
Piera-Jiménez et al [42]	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Low
Southard et al [32]	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Medium
Wang et al [33]	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Medium
Maciejewski et al [34]	Yes	No	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Medium
Turkstra et al [38]	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes	Low

[a]N/A: not applicable.

Discussion

Principal Findings

In general, digital health interventions for healthy behavior in people with chronic diseases are cost-effective, as all studies with cost-effective results have a low risk of bias. Previous studies have shown that digital interventions positively affect smoking, alcohol consumption, diet, and PA [46]. However, it is impossible to know how this effect will sustain for many years, as many studies had considerably short follow-up periods.

Studies on digital interventions for reducing behavioral risks of CVD in nonclinical adult populations revealed that they were effective 6 months after the end of the intervention, and the interventions lost their effectiveness after 12 months, according to a scoping review [46]. It also concluded that the shorter duration of effect was due to a shorter follow-up period and intention-to-treat analysis.

In most cases, studies in this review used <2 years as a follow-up period, and only 10% (2/20) used lifelong time horizons for economic evaluation [25,35]. Except for 10% (2/20) of studies [27,37] that used >2 years as an intervention period, most studies used parameters from the short-term effects of interventions to construct cost-effectiveness estimates and extrapolation. The results could be misleading because some behaviors could relapse, such as smoking, PA, and eating habits, which could diminish the effectiveness of the intervention, and hence extrapolation could overestimate the effects. This problem is particularly prevalent in mathematical modeling that predicts the outcomes of interventions over a person's lifetime because their parameters of economic impacts are based on a model of behavioral changes beyond the intervention period.

Most studies (6/20, 30%) with cost-effective or cost-saving results were published after 2010 [25,29,30,35,39,40]. With technology costs likely to have decreased in the recent years, digital health intervention costs could have been higher in the studies published before 2010 than in more recent ones; therefore, the cost-effectiveness of digital health interventions could be confounded by the year of publication. In Australia, 2 studies used telephone coaching as the intervention method. One study conducted in 2009 showed that the intervention cost was Aus $570 (US $460) per participant [36], whereas the other study showed that it was Aus $33 (US $25) per participant in 2017 [35].

Only half of studies (10/20, 50%) used specific clinical indicators, such as hemoglobin A_{1c} level, low-density lipoprotein cholesterol, and BP in mmHg, to measure clinical outcomes concerning the program's effectiveness. Other studies (10/20, 50%) used more general indicators, such as hospital admission rates, readmission rates, length of hospital stay, mortality rates, morbidity rates, and health-related quality of life, and interpreted reductions in these indicators as well as reduced health care resource use as evidence of improved clinical outcomes. For instance, decreased hospital admission rates or reduced outpatient visits could be due to reasons other than the effectiveness of the program. In addition, except for 15% (3/20) of studies [29,36,39] that used specific behavioral indicators, improvement in behavior or lifestyle was usually measured by clinical outcomes in most studies. These findings could be problematic in interpreting the program's effectiveness, as the improvement in clinical outcomes may be due to pharmacologic effects (antihypertensive medication, for instance) rather than adoption of healthy behaviors.

Although UK NICE guidelines strongly recommends a societal perspective for economic evaluations, it was implemented in only 15% (3/20) of studies [29,41,42], whereas the others (17/20, 85%) used health care payer perspectives. The results of an economic evaluation could be more cost-effective when conducted from a societal perspective, partly because the inclusion of homecare costs and productivity loss owing to illness significantly impact economic benefits. Furthermore, nonhealth outcomes, such as waiting time, time to diagnosis, and improved education and reassurance, should also be considered when assessing the cost-effectiveness of an intervention program.

Some behavior change interventions are embedded in telemonitoring, tele-education, or teleconsultation services that act as internet-based visits and enhance patient self-monitoring [27,38,40,47]. As a result, physical access to health care services would be reduced, but this does not necessarily mean reduced demand because of a healthy lifestyle. Therefore, researchers should be aware of this pitfall and use more specific indicators to measure the outcomes of healthy behaviors.

Although 25% (5/20) of studies [28,35,37,40,42] concluded that their results could be generalized to other settings, this is only possible for populations with high chronic disease prevalence because none of these interventions were aimed at the entire population. Because of the need for more information from low- and middle-income countries (LMICs), evidence-based recommendations are challenging to develop; however, digital health interventions also have potential. Although all studies were conducted in high-income settings, scaling up the digital health intervention in LMICs is feasible because of the high NCD burden and high population in these countries. Labrique et al [48] discussed that scaling up the digital health interventions in LMICs is possible under 5 conditions: involvement of end user inputs, engagement of all stakeholders in the developmental process, a good technical profile (simplicity, interoperability, and adaptability), well-established policy, and availability of appropriate infrastructure for digital health. The mHealth platforms will be more effective than other eHealth platforms because mobile phone use is on the rise, and smartphone adoption and use is ubiquitous not only in high-income countries but also in LMICs [49]. In addition, a systematic review found that mHealth can significantly modify health behavior as smartphones become more accessible to underserved and minority communities [50].

Owing to the demand for remote health services resulting from COVID-19, health care systems have implemented digital health and telemedicine solutions. Although telemedicine and digital solutions cannot replace all components of the health care experience, they offer certain advantages, such as the convenience of care, technology-assisted remote interaction, and increased accessibility to care, which can be crucial in managing chronic diseases [51]. Cost-effectiveness, accessibility to specialty services, and the ability to assist in alleviating physician shortages are key benefits of telemedicine, especially during COVID-19 [52]. Although health care professionals' attitudes toward telemedicine were influenced by factors such as self-efficacy, performance expectations, and facilitating conditions, mHealth emerged as the most preferred mode of telemedicine, enabling health care systems to be integrated into telemedicine systems during pandemics in low-income countries [53].

Recommendations for Further Research

On the basis of the findings of this review, the following recommendations are suggested:

1. The research question should include a cost-effectiveness assessment of the interventions for economic evaluation. Future studies should follow NICE recommendations to take a societal perspective, apply discounting, address parameter uncertainty, and apply a lifelong time horizon.
2. A full economic evaluation (CEA, CBA, and CUA) is needed to evaluate the cost-effectiveness of digital health interventions.
3. Researchers should use behavior-specific indicators such as walking time (minutes per week) for PA, urine nicotine testing for smoking, daily serving of fruits and vegetables, or plasma carotenoid index for diet, in addition to clinical indicators for the respective diseases.
4. Future research should be conducted on more diverse populations with chronic diseases to identify populations that can benefit the most from these interventions.
5. Assessment of the cost-effectiveness of digital interventions for behavioral change should include all stakeholders, including policy makers, implementers, and end users, to ensure that the final product is acceptable, scalable, feasible, and sustainable for wider implementation.

Limitations

First, because most studies in this review sought to determine the effectiveness of digital health interventions based on clinical outcomes, economic evaluations were embedded in RCTs. Thus, most studies have many weaknesses in economic evaluations, such as not using QALY or DALY, no discounting, and no sensitivity analysis, which lead to uncertainty in decision making regarding cost-effectiveness. Moreover, this review contains no studies on LMICs, making it difficult to generalize the findings to broader regions because many LMICs have a poor infrastructure for digital health, such as an unstable internet connection. Finally, this review has limited conclusions owing to the heterogeneity of the interventions and diseases examined and the short follow-up periods. Furthermore, the heterogeneity of the results makes a meta-analysis difficult.

Conclusions

Digital health interventions for behavioral change among people with chronic diseases are cost-effective in high-income settings and can therefore be scaled up. Similar evidence from LMICs based on properly designed studies for cost-effectiveness evaluation is urgently needed. A full economic evaluation is required to provide robust evidence of the cost-effectiveness of digital health interventions and their potential for scaling up in the broader population.

Acknowledgments

The Sustainable Behavior Change for Health Supported by Person-Tailored, Adaptive, Risk-Aware Digital Coaching in a Social Context project, funded by FORTE (Dnr 2018-01461), supported this research.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search blocks.
[DOCX File , 31 KB - ijmr_v12i1e42396_app1.docx]

Multimedia Appendix 2
Eligibility checklist.
[DOCX File , 23 KB - ijmr_v12i1e42396_app2.docx]

Multimedia Appendix 3
Study characteristics.
[DOCX File , 23 KB - ijmr_v12i1e42396_app3.docx]

Multimedia Appendix 4
Joanna Briggs Institute’s checklist for economic evaluations.
[DOCX File , 684 KB - ijmr_v12i1e42396_app4.docx]

Multimedia Appendix 5
Joanna Briggs Institute’s checklist for randomized controlled trials.
[DOCX File , 688 KB - ijmr_v12i1e42396_app5.docx]

References

1. The top 10 causes of death. World Health Organization. 2020 Dec 9. URL: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death [accessed 2022-04-11]
2. GBD 2016 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life-years (DALYs) for 333 diseases and injuries and healthy life expectancy (HALE) for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet 2017 Sep 16;390(10100):1260-1344 [FREE Full text] [doi: 10.1016/S0140-6736(17)32130-X] [Medline: 28919118]
3. Halbert RJ, Natoli JL, Gano A, Badamgarav E, Buist AS, Mannino DM. Global burden of COPD: systematic review and meta-analysis. Eur Respir J 2006 Sep;28(3):523-532 [FREE Full text] [doi: 10.1183/09031936.06.00124605] [Medline: 16611654]
4. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Oct 17;396(10258):1204-1222 [FREE Full text] [doi: 10.1016/S0140-6736(20)30925-9] [Medline: 33069326]
5. Guilbert JJ. The world health report 2002 - reducing risks, promoting healthy life. Educ Health (Abingdon) 2003 Jul;16(2):230. [doi: 10.1080/1357628031000116808] [Medline: 14741909]
6. Global health risks : mortality and burden of disease attributable to selected major risks. World Health Organization. 2009. URL: https://apps.who.int/iris/handle/10665/44203 [accessed 2022-04-11]
7. The Lancet: latest global disease estimates reveal perfect storm of rising chronic diseases and public health failures fuelling COVID-19 pandemic. Institute for Health Metrics and Evaluation. 2020 Oct 15. URL: https://www.healthdata.org/news-release/lancet-latest-global-disease-estimates-reveal-perfect-storm-rising-chronic-diseases-and [accessed 2022-04-12]
8. Baanders AN, Heijmans MJ. The impact of chronic diseases: the partner's perspective. Fam Community Health 2007;30(4):305-317. [doi: 10.1097/01.FCH.0000290543.48576.cf] [Medline: 17873637]
9. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
10. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95 [FREE Full text] [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
11. Behaviour change: individual approaches - Public Health Guideline (PH 49). National Institute for Health and Care Excellence (NICE). 2014 Jan 2. URL: https://www.nice.org.uk/guidance/ph49/resources/behaviour-change-individual-approaches-pdf-1996366337989 [accessed 2022-04-16]
12. Akinosun AS, Polson R, Diaz-Skeete Y, De Kock JH, Carragher L, Leslie S, et al. Digital technology interventions for risk factor modification in patients with cardiovascular disease: systematic review and meta-analysis. JMIR Mhealth Uhealth 2021 Mar 03;9(3):e21061 [FREE Full text] [doi: 10.2196/21061] [Medline: 33656444]
13. Gaziano TA, Galea G, Reddy KS. Scaling up interventions for chronic disease prevention: the evidence. Lancet 2007 Dec 08;370(9603):1939-1946. [doi: 10.1016/S0140-6736(07)61697-3] [Medline: 18063028]
14. Whitten PS, Mair FS, Haycox A, May CR, Williams TL, Hellmich S. Systematic review of cost effectiveness studies of telemedicine interventions. BMJ 2002 Jun 15;324(7351):1434-1437 [FREE Full text] [doi: 10.1136/bmj.324.7351.1434] [Medline: 12065269]
15. Rojas SV, Gagnon MP. A systematic review of the key indicators for assessing telehomecare cost-effectiveness. Telemed J E Health 2008 Nov;14(9):896-904 [FREE Full text] [doi: 10.1089/tmj.2008.0009] [Medline: 19035798]
16. Lindgren H, Guerrero E, Jingar M, Lindvall K, Ng N, Richter Sundberg L, et al. The STAR-C intelligent coach: a cross-disciplinary design process of a behavior change intervention in primary care. Stud Health Technol Inform 2020 Sep 04;273:203-208. [doi: 10.3233/SHTI200640] [Medline: 33087613]
17. Ng N, Eriksson M, Guerrero E, Gustafsson C, Kinsman J, Lindberg J, et al. Sustainable Behavior Change for Health Supported by Person-Tailored, Adaptive, Risk-Aware Digital Coaching in a Social Context: Study Protocol for the STAR-C Research Programme. Front Public Health 2021 Mar 1;9:593453 [FREE Full text] [doi: 10.3389/fpubh.2021.593453] [Medline: 33732674]
18. Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions version 6.3. 2nd edition. Hoboken, NJ, USA: John Wiley & Sons; Oct 2019.
19. Drummond MF, Sculpher MJ, Torrance GW, O'Brien BJ, Stoddart GL. Methods for the Economic Evaluation of Health Care Programmes. 3rd edition. Oxford, UK: Oxford University Press; 2005.
20. Veličković V, Višnjić A, Mihajlović J. Quick reference guide for critical appraisal of economics evaluations for busy decision makers. Acta Fac Med Naissensis 2015;32(1):23-30 [FREE Full text] [doi: 10.1515/afmnai-2015-0002]
21. Rayyan homepage. Rayyan. URL: https://www.rayyan.ai [accessed 2022-02-01]
22. NICE Process and Methods Guides. Methods for the development of NICE public health guidance. 3rd edition. National Institute for Health and Care Excellence. 2012 Sep 26. URL: https://www.nice.org.uk/process/pmg4/chapter/introduction [accessed 2022-04-19]

23. Chen F, Su W, Becker SH, Payne M, Castro Sweet CM, Peters AL, et al. Clinical and economic impact of a digital, remotely-delivered intensive behavioral counseling program on Medicare beneficiaries at risk for diabetes and cardiovascular disease. PLoS One 2016 Oct 5;11(10):e0163627 [FREE Full text] [doi: 10.1371/journal.pone.0163627] [Medline: 27706216]
24. Copeland LA, Berg GD, Johnson DM, Bauer RL. An intervention for VA patients with congestive heart failure. Am J Manag Care 2010 Mar;16(3):158-165 [FREE Full text] [Medline: 20225911]
25. Datta SK, Oddone EZ, Olsen MK, Orr M, McCant F, Gentry P, et al. Economic analysis of a tailored behavioral intervention to improve blood pressure control for primary care patients. Am Heart J 2010 Aug;160(2):257-263. [doi: 10.1016/j.ahj.2010.05.024] [Medline: 20691830]
26. Dunagan WC, Littenberg B, Ewald GA, Jones CA, Emery VB, Waterman BM, et al. Randomized trial of a nurse-administered, telephone-based disease management program for patients with heart failure. J Card Fail 2005 Jun;11(5):358-365. [doi: 10.1016/j.cardfail.2004.12.004] [Medline: 15948086]
27. Finkelstein SM, Speedie SM, Potthoff S. Home telehealth improves clinical outcomes at lower cost for home healthcare. Telemed J E Health 2006 Apr;12(2):128-136. [doi: 10.1089/tmj.2006.12.128] [Medline: 16620167]
28. Fischer HH, Eisert SL, Everhart RM, Durfee MJ, Moore SL, Soria S, et al. Nurse-run, telephone-based outreach to improve lipids in people with diabetes. Am J Manag Care 2012 Feb;18(2):77-84 [FREE Full text] [Medline: 22435835]
29. Keyserling TC, Sheridan SL, Draeger LB, Finkelstein EA, Gizlice Z, Kruger E, et al. A comparison of live counseling with a web-based lifestyle and medication intervention to reduce coronary heart disease risk: a randomized clinical trial. JAMA Intern Med 2014 Jul;174(7):1144-1157 [FREE Full text] [doi: 10.1001/jamainternmed.2014.1984] [Medline: 24861959]
30. Nordyke RJ, Appelbaum K, Berman MA. Estimating the impact of novel digital therapeutics in type 2 diabetes and hypertension: health economic analysis. J Med Internet Res 2019 Oct 09;21(10):e15814 [FREE Full text] [doi: 10.2196/15814] [Medline: 31599740]
31. Nundy S, Dick JJ, Chou CH, Nocon RS, Chin MH, Peek ME. Mobile phone diabetes project led to improved glycemic control and net savings for Chicago plan participants. Health Aff (Millwood) 2014 Feb;33(2):265-272 [FREE Full text] [doi: 10.1377/hlthaff.2013.0589] [Medline: 24493770]
32. Southard BH, Southard DR, Nuckolls J. Clinical trial of an internet-based case management system for secondary prevention of heart disease. J Cardiopulm Rehabil 2003;23(5):341-348. [doi: 10.1097/00008483-200309000-00003] [Medline: 14512778]
33. Wang V, Smith VA, Bosworth HB, Oddone EZ, Olsen MK, McCant F, et al. Economic evaluation of telephone self-management interventions for blood pressure control. Am Heart J 2012 Jun;163(6):980-986. [doi: 10.1016/j.ahj.2012.03.016] [Medline: 22709750]
34. Maciejewski ML, Bosworth HB, Olsen MK, Smith VA, Edelman D, Powers BJ, et al. Do the benefits of participation in a hypertension self-management trial persist after patients resume usual care? Circ Cardiovasc Qual Outcomes 2014 Mar;7(2):269-275. [doi: 10.1161/CIRCOUTCOMES.113.000309] [Medline: 24619321]
35. Burn E, Nghiem S, Jan S, Redfern J, Rodgers A, Thiagalingam A, et al. Cost-effectiveness of a text message programme for the prevention of recurrent cardiovascular events. Heart 2017 Jun;103(12):893-894 [FREE Full text] [doi: 10.1136/heartjnl-2016-310195] [Medline: 28235776]
36. Graves N, Barnett AG, Halton KA, Veerman JL, Winkler E, Owen N, et al. Cost-effectiveness of a telephone-delivered intervention for physical activity and diet. PLoS One 2009 Sep 25;4(9):e7135 [FREE Full text] [doi: 10.1371/journal.pone.0007135] [Medline: 19779611]
37. Hamar GB, Rula EY, Coberley C, Pope JE, Larkin S. Long-term impact of a chronic disease management program on hospital utilization and cost in an Australian population with heart disease or diabetes. BMC Health Serv Res 2015 Apr 22;15:174 [FREE Full text] [doi: 10.1186/s12913-015-0834-z] [Medline: 25895499]
38. Turkstra E, Hawkes AL, Oldenburg B, Scuffham PA. Cost-effectiveness of a coronary heart disease secondary prevention program in patients with myocardial infarction: results from a randomised controlled trial (ProActive Heart). BMC Cardiovasc Disord 2013 May 01;13:33 [FREE Full text] [doi: 10.1186/1471-2261-13-33] [Medline: 23634982]
39. Maddison R, Pfaeffli L, Whittaker R, Stewart R, Kerr A, Jiang Y, et al. A mobile phone intervention increases physical activity in people with cardiovascular disease: results from the HEART randomized controlled trial. Eur J Prev Cardiol 2015 Jun;22(6):701-709. [doi: 10.1177/2047487314535076] [Medline: 24817694]
40. McManus RJ, Little P, Stuart B, Morton K, Raftery J, Kelly J, HOME BP investigators. Home and Online Management and Evaluation of Blood Pressure (HOME BP) using a digital intervention in poorly controlled hypertension: randomised controlled trial. BMJ 2021 Jan 19;372:m4858 [FREE Full text] [doi: 10.1136/bmj.m4858] [Medline: 33468518]
41. Bertuzzi F, Stefani I, Rivolta B, Pintaudi B, Meneghini E, Luzi L, et al. Teleconsultation in type 1 diabetes mellitus (TELEDIABE). Acta Diabetol 2018 Feb;55(2):185-192. [doi: 10.1007/s00592-017-1084-9] [Medline: 29209814]
42. Piera-Jiménez J, Winters M, Broers E, Valero-Bover D, Habibovic M, Widdershoven JW, et al. Changing the health behavior of patients with cardiovascular disease through an electronic health intervention in three different countries: cost-effectiveness study in the do cardiac health: advanced new generation ecosystem (Do CHANGE) 2 randomized controlled trial. J Med Internet Res 2020 Jul 28;22(7):e17351 [FREE Full text] [doi: 10.2196/17351] [Medline: 32720908]
43. Empana JP, Ducimetière P, Arveiler D, Ferrières J, Evans A, Ruidavets JB, PRIME Study Group. Are the Framingham and PROCAM coronary heart disease risk functions applicable to different European populations? The PRIME Study. Eur Heart J 2003 Nov;24(21):1903-1911. [doi: 10.1016/j.ehj.2003.09.002] [Medline: 14585248]

44. Walker D, Fox-Rushby J. Allowing for uncertainty in economic evaluations: qualitative sensitivity analysis. Health Policy Plan 2001 Dec;16(4):435-443. [doi: 10.1093/heapol/16.4.435] [Medline: 11739369]
45. Telediabete homepage. Telediabete. URL: http://www.telediabete-fid.it/ [accessed 2022-04-25]
46. Gold N, Yau A, Rigby B, Dyke C, Remfry EA, Chadborn T. Effectiveness of digital interventions for reducing behavioral risks of cardiovascular disease in nonclinical adult populations: systematic review of reviews. J Med Internet Res 2021 May 14;23(5):e19688 [FREE Full text] [doi: 10.2196/19688] [Medline: 33988126]
47. Dixon P, Hollinghurst S, Edwards L, Thomas C, Gaunt D, Foster A, et al. Cost-effectiveness of telehealth for patients with raised cardiovascular disease risk: evidence from the Healthlines randomised controlled trial. BMJ Open 2016 Aug 26;6(8):e012352 [FREE Full text] [doi: 10.1136/bmjopen-2016-012352] [Medline: 27566642]
48. Labrique AB, Wadhwani C, Williams KA, Lamptey P, Hesp C, Luk R, et al. Best practices in scaling digital health in low and middle income countries. Global Health 2018 Nov 03;14(1):103 [FREE Full text] [doi: 10.1186/s12992-018-0424-z] [Medline: 30390686]
49. Thomas Craig KJ, Morgan LC, Chen CH, Michie S, Fusco N, Snowdon JL, et al. Systematic review of context-aware digital behavior change interventions to improve health. Transl Behav Med 2021 May 25;11(5):1037-1048 [FREE Full text] [doi: 10.1093/tbm/ibaa099] [Medline: 33085767]
50. Anderson-Lewis C, Darville G, Mercado RE, Howell S, Di Maggio S. mHealth technology use and implications in historically underserved and minority populations in the United States: systematic literature review. JMIR Mhealth Uhealth 2018 Jun 18;6(6):e128 [FREE Full text] [doi: 10.2196/mhealth.8383] [Medline: 29914860]
51. Seixas AA, Olaye IM, Wall SP, Dunn P. Optimizing healthcare through digital health and wellness solutions to meet the needs of patients with chronic disease during the COVID-19 era. Front Public Health 2021 Jul 12;9:667654 [FREE Full text] [doi: 10.3389/fpubh.2021.667654] [Medline: 34322469]
52. Kichloo A, Albosta M, Dettloff K, Wani F, El-Amir Z, Singh J, et al. Telemedicine, the current COVID-19 pandemic and the future: a narrative review and perspectives moving forward in the USA. Fam Med Community Health 2020 Aug;8(3):e000530 [FREE Full text] [doi: 10.1136/fmch-2020-000530] [Medline: 32816942]
53. Shiferaw KB, Mengiste SA, Gullslett MK, Zeleke AA, Tilahun B, Tebeje T, et al. Healthcare providers' acceptance of telemedicine and preference of modalities during COVID-19 pandemics in a low-resource setting: an extended UTAUT model. PLoS One 2021 Apr 22;16(4):e0250220 [FREE Full text] [doi: 10.1371/journal.pone.0250220] [Medline: 33886625]

Abbreviations

BCT: behavior change technique
BP: blood pressure
CBA: cost-benefit analysis
CEA: cost-effectiveness analysis
CET: cost-effectiveness threshold
COPD: chronic obstructive pulmonary disease
CUA: cost-utility analysis
CVD: cardiovascular disease
DALY: disability-adjusted life-year
DM: diabetes mellitus
ICER: incremental cost-effectiveness ratio
LMICs: low- and middle-income countries
mHealth: mobile health
MI: myocardial infarction
NCD: noncommunicable disease
NICE: National Institute for Health and Clinical Excellence
PA: physical activity
PICO: Population, Intervention, Comparator, and Outcomes
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QALY: quality-adjusted life-years
RCT: randomized controlled trial
STAR-C: Sustainable Behavior Change for Health Supported by Person-Tailored, Adaptive, Risk-Aware Digital Coaching in a Social Context

Edited by A Mavragani; submitted 02.09.22; peer-reviewed by H Moghaddasi, L Brennan, N Archer; comments to author 22.12.22; revised version received 11.01.23; accepted 12.01.23; published 16.02.23.

Please cite as:
Kyaw TL, Ng N, Theocharaki M, Wennberg P, Sahlen KG
Cost-effectiveness of Digital Tools for Behavior Change Interventions Among People With Chronic Diseases: Systematic Review
Interact J Med Res 2023;12:e42396
URL: https://www.i-jmr.org/2023/1/e42396
doi:10.2196/42396
PMID:36795470

Original Paper

Big Data and Infectious Disease Epidemiology: Bibliometric Analysis and Research Agenda

Lateef Babatunde Amusa[1,2], PhD; Hossana Twinomurinzi[1], PhD; Edith Phalane[3,4], PhD; Refilwe Nancy Phaswana-Mafuya[3,4], PhD

[1]Centre for Applied Data Science, University of Johannesburg, Johannesburg, South Africa

[2]Department of Statistics, University of Ilorin, Ilorin, Nigeria

[3]Pan African Centre for Epidemics Research (PACER) Extramural Unit, South African Medical Research Council/University of Johannesburg, Johannesburg, South Africa

[4]Department of Environmental Health, Faculty of Health Sciences, University of Johannesburg, Johannesburg, South Africa

Corresponding Author:
Lateef Babatunde Amusa, PhD
Centre for Applied Data Science
University of Johannesburg
Bunting Road Campus, Auckland Park
Johannesburg, 2006
South Africa
Phone: 27 659904686
Email: amusasuxes@gmail.com

Abstract

Background: Infectious diseases represent a major challenge for health systems worldwide. With the recent global pandemic of COVID-19, the need to research strategies to treat these health problems has become even more pressing. Although the literature on big data and data science in health has grown rapidly, few studies have synthesized these individual studies, and none has identified the utility of big data in infectious disease surveillance and modeling.

Objective: The aim of this study was to synthesize research and identify hotspots of big data in infectious disease epidemiology.

Methods: Bibliometric data from 3054 documents that satisfied the inclusion criteria retrieved from the Web of Science database over 22 years (2000-2022) were analyzed and reviewed. The search retrieval occurred on October 17, 2022. Bibliometric analysis was performed to illustrate the relationships between research constituents, topics, and key terms in the retrieved documents.

Results: The bibliometric analysis revealed internet searches and social media as the most utilized big data sources for infectious disease surveillance or modeling. The analysis also placed US and Chinese institutions as leaders in this research area. Disease monitoring and surveillance, utility of electronic health (or medical) records, methodology framework for infodemiology tools, and machine/deep learning were identified as the core research themes.

Conclusions: Proposals for future studies are made based on these findings. This study will provide health care informatics scholars with a comprehensive understanding of big data research in infectious disease epidemiology.

(Interact J Med Res 2023;12:e42292) doi:10.2196/42292

KEYWORDS

big data; bibliometrics; infectious disease; COVID-19; disease surveillance; disease; pandemic; data; surveillance; hotspot; epidemiology; social media; utility; electronic health records

Introduction

Globally, the infectious disease burden continues to be substantial in countries with low and lower-middle income, while morbidity and mortality related to neglected tropical diseases and HIV infection, tuberculosis, and malaria remain high. Tuberculosis and malaria are endemic to many areas, imposing substantial but steady burdens. At the same time, other infections such as influenza fluctuate in pervasiveness and intensity, disrupting the developing and developed settings alike when an outbreak and epidemic occurs. Additionally, deaths have persisted over the 21st century due to emerging and reemerging infectious diseases compared with seasonal and endemic infections. This portrays a new era of infectious disease, defined by outbreaks of emerging, reemerging, and endemic pathogens that spread quickly with the help of global mobility and climate change [1].

Moreover, the risk from infectious diseases is globally shared. While infectious diseases thrive in underresourced settings, inequalities and inequities in accessing health and health care create a favorable environment for infectious diseases to spread [2,3]. Addressing inequalities and inequities in accessing health care, and improving surveillance and monitoring of infectious diseases should be prioritized to minimize the emergence and spread of infections.

Recent years have witnessed the rapid emergence of big data and data science research, propelled by the increasing availability of digital traces [4]. The growing availability of electronic records and passive data generated by social media, the internet, and other digital sources can be mined for pattern discoveries and knowledge extraction. Like most buzz words, *big data* has no straightforward meaning and its definition is evolving. Broadly, big data refers to a large volume of structured or unstructured data, with *largeness* itself associated with three major terms known as the "3 Vs": volume (large quantity), velocity (coming in at unprecedented real-time speeds), and variety (increasing collection from different data sources). Additional characteristics of big data include veracity, validity, volatility, and value [5]. For epidemiology and infectious diseases research, this means that in the last decade, there has been a significant spike in the number of studies with considerable interest in using digital epidemiology and big data tools to enhance health systems in terms of disease surveillance, modeling, and evidence-based responses [4,6-8]. Digital epidemiology uses digital data or online sources to gain insight into disease dynamics and health equity, and to inform public health programs and policies [9,10].

The success of infectious disease control relies heavily on surveillance systems tracking diseases, pathogens, and clinical outcomes [11]. However, conventional surveillance systems are known to frequently have severe time lags and limited spatial resolution; therefore, surveillance systems that are robust, local, and timely are critically needed. It is crucial to monitor and forecast emerging and reemerging infections [12] such as severe acute respiratory syndrome, pandemic influenza, Ebola, Zika, and drug-resistant pathogens, especially in resource-limited settings such as low-middle–income countries. Using big data to strengthen surveillance systems is critical for future pandemic preparedness. This approach provides big data streams that can be triangulated with spatial and temporal data. These big data streams include digital data sources such as mobile health apps, electronic health (or medical) records, social media, internet searches, mobile phone network data, and GPS mobile data. Many studies have demonstrated the usefulness of real-time data in health assessments [13-18]. Some of these studies have been used explicitly for the monitoring and forecasting of epidemics such as COVID-19 [19], Zika [13], Ebola [16], and influenza [14].

The body of extant literature at the nexus of big data, epidemiology, and infectious diseases is rapidly growing. However, despite its growth and dispersion, there has been a limited synthesis of the applications. A previous study [20] performed a bibliometric analysis focusing on only HIV. A bibliometric analysis is a statistical or quantitative analysis of large-scale bibliographic metadata (or metrics of published studies) on a given topic. These quantitative analyses detect patterns, networks, and trends among the bibliographic metadata [21,22]. Thus, the aim of this study was to address the evolution of big data in epidemiology and infectious diseases to identify gaps and opportunities for further research. The study findings reveal interesting patterns and can inform trending research focus and future directions in big data–driven infectious diseases research.

Methods

Study Design

A bibliometric analysis was performed to understand and explore research on big data in infectious disease modeling and surveillance. The adopted bibliometric methodology involved three main phases: data collection, data analysis, and data visualization and reporting [23].

Search Strategy

Regarding data collection, which entails querying and exporting data from selected databases, we queried the Web of Science (WoS) core databases for publications using specific inclusion and exclusion criteria. Compared to other databases, the WoS has been shown to have better quality bibliometric information [23,24] and more excellent coverage of high-impact journals [25]. With the aid of domain knowledge experts from the fields of both big data and epidemiology, we iteratively developed a search strategy and selected the following search terms. The following search string queried all documents' titles, abstracts, and keywords, and generated 3235 publications in the WoS collection:

> *(Epidemic* OR "infectious disease*" OR "Disease surveillance" OR "disease transmission" OR "disease outbreak*" OR ("communicable disease*" NOT "non-communicable disease") OR syndemic* OR HIV OR AIDS OR "human immunodeficiency virus" OR coronavirus* OR SARS-CoV-2 OR COVID-19 OR Influenza OR flu OR Zika OR Ebola OR MERS OR "Middle East respiratory syndrome" OR Tuberculosis OR "Monkey Pox" OR "Dengue virus" OR Hepatitis*)*

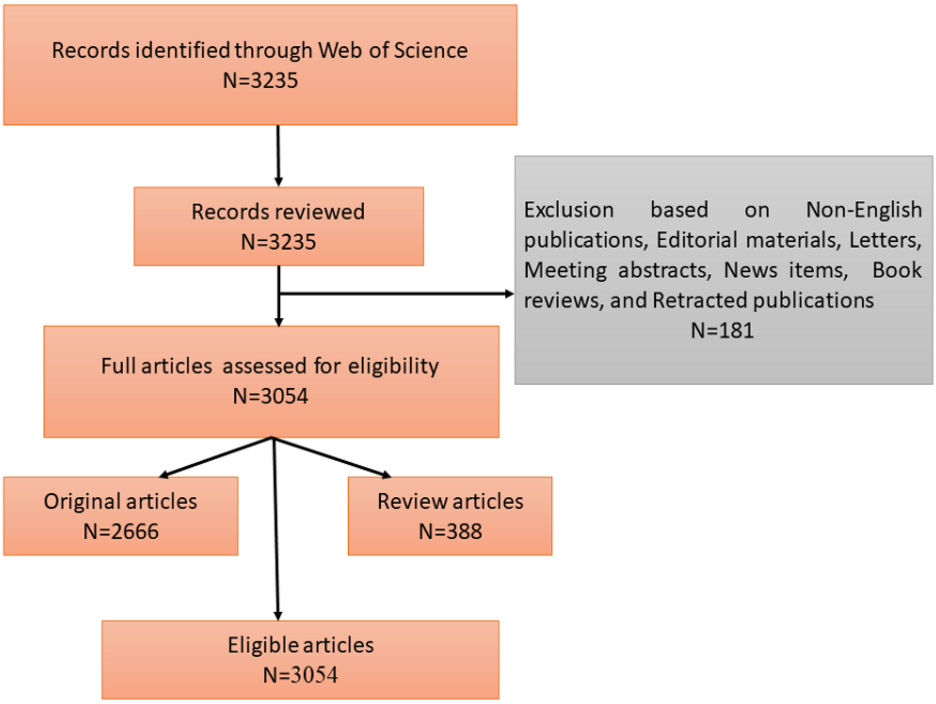
AND

> *("BIG DATA" OR "web mining" OR "opinion mining" OR "Google Trend*" OR "Google search*" OR "Google quer*" OR "Internet search*" OR "Internet quer*" OR "search engine quer*" OR "Digital traces" OR "electronic health records" OR "Digital epidemiology")*

Screening Strategy

Documents not written in English and not peer-reviewed, including editorial materials, letters, meeting abstracts, news items, book reviews, and retracted publications, were removed from the data set given the focus on bibliometric analysis, leaving 3054 documents for the analytic sample (Figure 1).

Figure 1. Flow chart of the literature selection process.



Analysis

The 3054 bibliographic data were exported into the R package *bibliometrix* [23] for analysis. This package was specifically used to conduct performance analysis and science mapping of big data in infectious disease epidemiology. Performance mapping evaluates the production and impact of research constituents, including authors, institutions, countries, and journals. Science mapping examines the relationships between the research constituents by analyzing the topic's conceptual, intellectual, and social structure.

There are several metrics available for bibliometric analysis. In this study, the primary metrics used for evaluating productivity and influence were the H-index and M-index. The H-index represents the number of published papers *h*, such that the citation number is at least *h* [26]. The H-index can be computed for different bibliometric units of analysis: authors, journals, institutions, and countries. The M-index simply adjusts the H-index for the academic age (ie, the number of years since the researcher's first publication). Other utilized performance analysis metrics were obtained from yearly research output and citation counts. These metrics also contribute to identifying the main themes and the key actors in the research area.

In terms of science mapping, network maps were constructed for some selected bibliographic units of analysis [27]. These networks exhibit frequency distributions of the involved bibliographic data over time. For instance, international collaborations can be explored by assessing same-country publications. A cocitation network analysis was also used to analyze publication references. In addition, using the Louvain clustering algorithm and a greedy optimization technique [28], a co-occurrence analysis was used to understand the conceptual structure of the research area. The basic purpose of co-occurrence analysis is to investigate the link between keywords based on the number of times they appear together in a publication. Notable research topics and over-time trends were detected by generating clusters for author-provided keywords [29]. VOSviewer [30] was used to construct the network visualizations. Each network node represents a research constituent (eg, author, country, institution, article, document source, keyword). The node's size is proportional to the occurrence frequency of the relevant parameters. The degree of association is represented by the thickness of the link between nodes, and the various colors reflect distinct clusters.

Results

Descriptive Summary

The bibliographic data set comprises 3054 documents from 1600 sources, 14,351 authors, and 121,726 references. From the 3054 documents, 2666 (87.30%) were original research articles and the remaining 388 (12.70%) were review papers. The research output before 2009 was relatively low. The annual publication output during the 27 years (1995-2022) grew steadily, with a yearly growth rate of 26.5%. The publication growth increased steeply between 2013 and 2020 (Figure 2). Table 1 presents the summary statistics of the primary characteristics of these 3054 publications, including the time span and information about documents and authors.

As shown in Table 2, the most productive and influential sources publishing on topics related to big data and infectious diseases epidemiology were *Journal of Medical Internet Research* and

PLoS One (H-index=18), followed by *IEEE Access* (H-index=13). In terms of productivity, *Journal of Medical Internet Research* produced a slightly higher number of publications (n=61) than the next best journal *PLoS One* (n=56). *PLoS One* had the highest number of total citations at 1893.

As shown in Table 3, the most productive and influential author was Zhang Y (H-index=17), followed by Li X (H-index=13) and Wang J (H-index=12). Wang L had the highest total citations (n=1072), which was substantially higher than the next most impactful author Wang J (total citations=861).

Figure 2. Annual growth of publications related to big data in infectious diseases research.

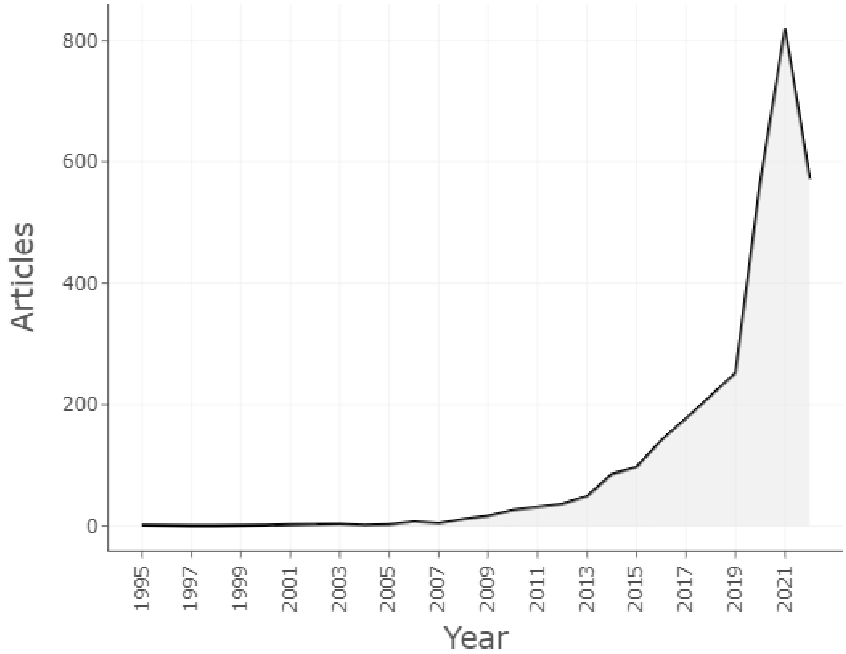


Table 1. Main descriptive summary of the extracted bibliographic records from 1995 to 2022.

Description	Results
Time span (years)	1995-2022
Sources, n	1600
Documents, n	3054
Annual growth rate, %	26.52
Document average age (years)	2.86
Average citations per document, n	18.52
References, n	121,726
Authors, n	
Total	14,351
Single-authored documents	225
Author collaborations	
Single-authored documents, n	236
Coauthors per document, n	5.55
International coauthorships, %	28.04

Table 2. Top 10 productive and influential publication sources ranked by H-index.

Journal	Aim and scope	H-index	M-index	Total citations, n	Publications, n	Publication year
Journal of Medical Internet Research	Digital health, data science, health informatics, and emerging technologies for health, medicine, and biomedical research	18	1.13	1705	61	2007
PLoS One	Multidisciplinary	18	1.39	1893	56	2010
IEEE Access	Multidisciplinary, comprising all IEEE fields of interest, emphasizing applications-oriented and interdisciplinary articles	13	1.63	983	32	2015
Scientific Reports	Publishes from all areas of the natural sciences, psychology, medicine, and engineering	13	1.63	389	23	2015
Journal of the American Medical Informatics Association	Biomedical and health informatics, including clinical care, clinical research, translational science, implementation science, imaging, education, consumer health, public health, and policy	12	0.86	569	33	2009
BMJ Open	Medical journal considering papers in clinical medicine, public health, and epidemiology	11	1.10	310	32	2013
JMIR Public Health & Surveillance	Multidisciplinary journal with a unique focus on the intersection of innovation and technology in public health	11	2.20	724	23	2018
International Journal of Medical Informatics	Medical informatics, including information systems and computer-aided medical support decision systems	11	0.65	450	16	2006
International Journal of Infectious Diseases	Original clinical and laboratory-based research, together with reports of clinical trials, reviews, and some case reports dealing with the epidemiology, clinical diagnosis, treatment, and control of infectious diseases	10	0.91	530	17	2012
BMC Medical Informatics & Decision Making	Relating to the design, development, implementation, and evaluation of health information technologies and decision-making for human health	10	0.91	208	15	2012

Table 3. Top 10 productive and influential authors ranked by H-index and total citations.

Author	H-index	M-index	Total citations, n	Publications, n	Publication year
Zhang Y	17	—[a]	776	35	—
Li X	13	—	544	35	—
Wang J	12	1.33	861	24	2014
Wang L	12	—	1072	22	—
Wang Y	10	1.25	342	21	2015
Li Z	10	1.67	366	14	2017
Brownstein JS	10	0.77	748	11	2010
Wang Z	9	1.00	427	18	2014
Zhang W	9	1.13	556	12	2015
Zhang X	9	1.29	371	12	2016

[a]Not available.

The aim and scope of the top 10 most influential journals, as listed in Table 2, is to publish medical research, medical informatics, or multidisciplinary studies. It can thus be inferred that major future breakthroughs regarding big data in infectious diseases epidemiology will likely appear in these journals.

Figure 3 displays the top 20 most productive institutions. Institutional contributions were assessed by affiliations with at least one author in the publication. Except for the University of California, the top three institutions, which account for 21.3% of the number of publications in the top 20, were medical schools: Harvard Medical School (7.9%) and Icahn School of

Medicine at Mount Sinai (6.4%). The other institutions, each accounting for more than 6% of the total, included Columbia University and Oxford University in the top 5, whereas others in the top 20 are research universities: London School of Hygiene and Tropical Medicine focuses on global and public health, Taipei Medical University is medical-based, and Huazhong University of Science and Technology is focused on science and technology. The United States produced the majority of the top 10 most productive institutions, which were in the top 5.

The 20 most productive countries (Figure 4) are led by the United States and China, accounting for more than half (57.3%) of the total publication output. The United States alone accounted for 41.1% of the productivity in this field. The other countries in the top five were the United Kingdom (9.4%), India (4.4%), and Canada (3.3%).

Computer science was the most productive research domain in the bibliographic collection (Figure 5), accounting for 17.6% of the top 10 subject areas. In order of productivity, the other research subjects in the top 5 were public environmental and occupational health (11.4%), health care services (9.6%), medical informatics (9.0%), and engineering (8.8%).

Figure 3. Top 20 institutions by number of publications. CALIF: California; HARVARD MED SCH: Harvard Medical School; ICAHN SCH MED MT SINAI: Icahn School of Medicine at Mount Sinai; LONDON SCH HYG AND TROP MED: London School of Hygiene & Tropical Medicine; PENN: Pennsylvania; UNIV: University.

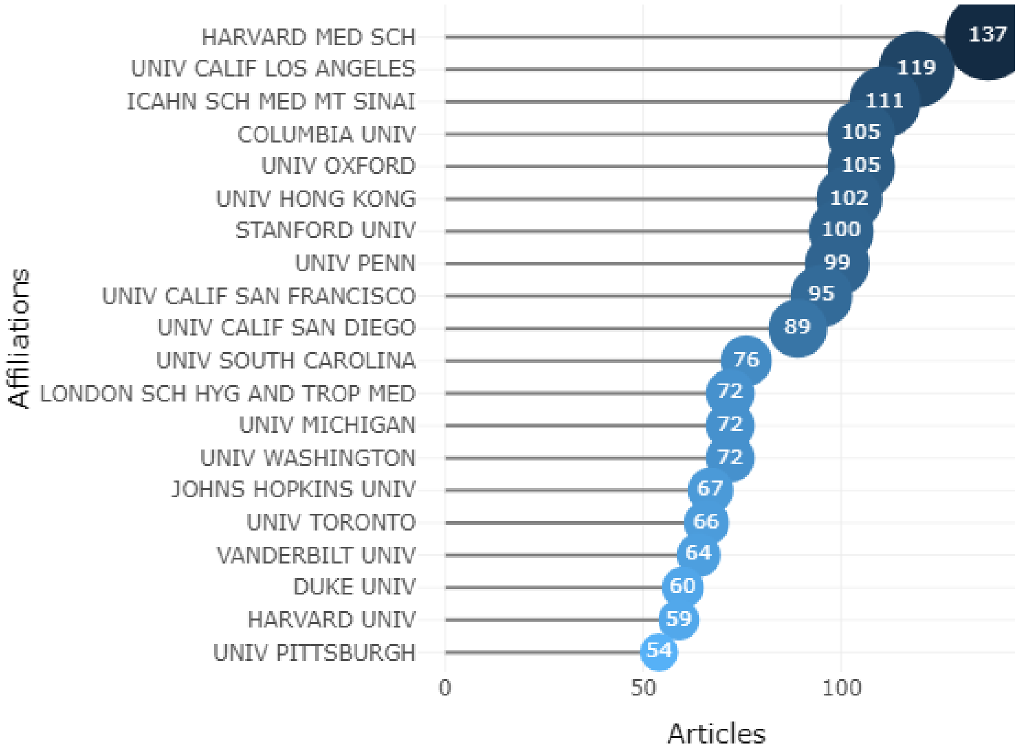


Figure 4. Top 20 productive countries by number of publications.

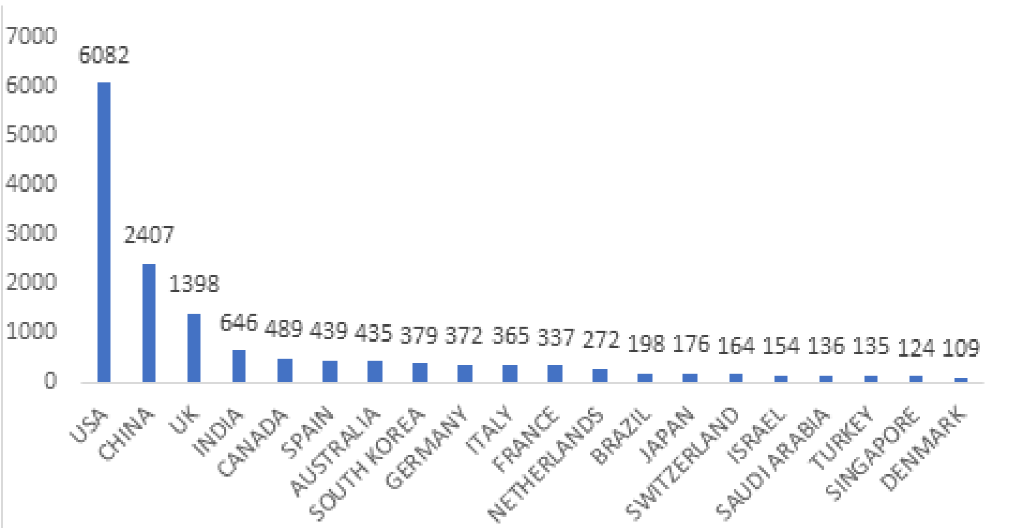
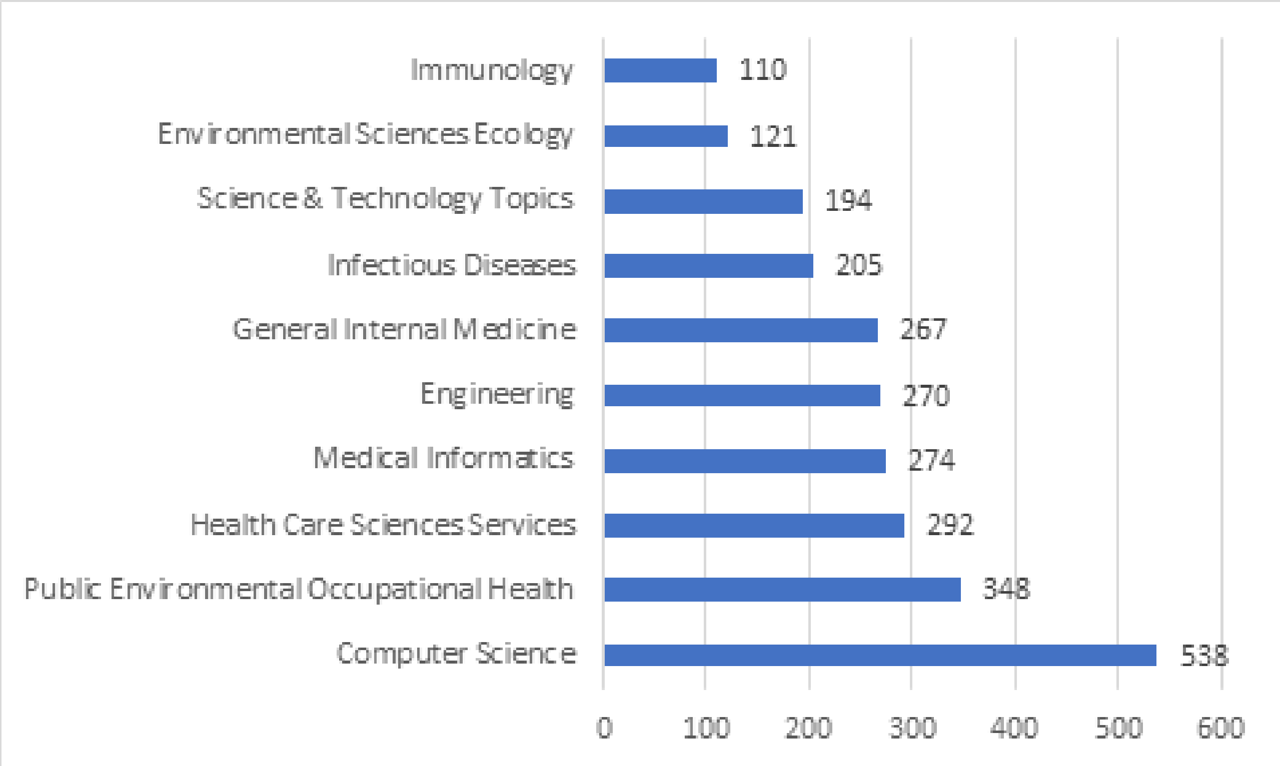


Figure 5. Top 10 key subject areas by number of publications.



Two major clusters of countries represent the collaboration patterns of the most productive countries (Figure 6). The network was set to include only countries with at least 10 documents, resulting in 50 productive countries. The clustering results demonstrated a demarcation of European countries from the others. For instance, cluster 1 (red) represented most countries from Europe, with England, Germany, and Spain being the core countries. Non-European countries constituted the second cluster (green). The United States and China were the core countries of this group.

Regarding collaboration strength, the United States, with a total link strength of 570, featured the highest number of partners (48), accounting for almost all 50 countries in the network (96%). China, which distantly followed the United States, featured 38 partners and a total link strength of 304. This implies that collaboration is mainly regional.

Figure 7 shows a network map of cocited references in this research area, wherein the node's size represents the citation strength of the individual studies. The network was set to include only studies with at least 25 citations, resulting in 37 studies. Ginsberg et al [31] published the most highly cited article (185 citations). This 13-year-old study presented a method that used Google search queries to track flu-like illnesses in a population.

The second most cited study by Eysenbach [9] introduced the concept of infodemiology, the science of using the internet (eg, social media, search engines, blogs, and websites) to inform public health and public policy. Table 4 further summarizes the top 15 most cited references, including the title, year of publication, number of citations, type of disease, and data source.

The 37 studies in the network map of cocited references produced four thematic clusters (Figure 7); disease monitoring and surveillance (cluster 1), utility of electronic health (or medical) records (cluster 2), methodology framework for infodemiology tools (cluster 3), and machine learning and deep learning methods (cluster 4) were the main topics discussed.

Keyword co-occurrence analysis serves as a supplement to enrich the understanding of the thematic clusters derived from the reference cocitation analysis and helps identify the core topics and contents [29]. As shown in Figure 8, the co-occurrence network displayed 100 relevant keywords after assigning a selection threshold of 10 for the number of keyword occurrences. The top 5 most frequently used keywords were *COVID-19*, *big data*, *machine learning*, *coronavirus*, and *electronic health records.*

Figure 6. Network of country collaborations (≥10 documents, 50 countries, 2 clusters).

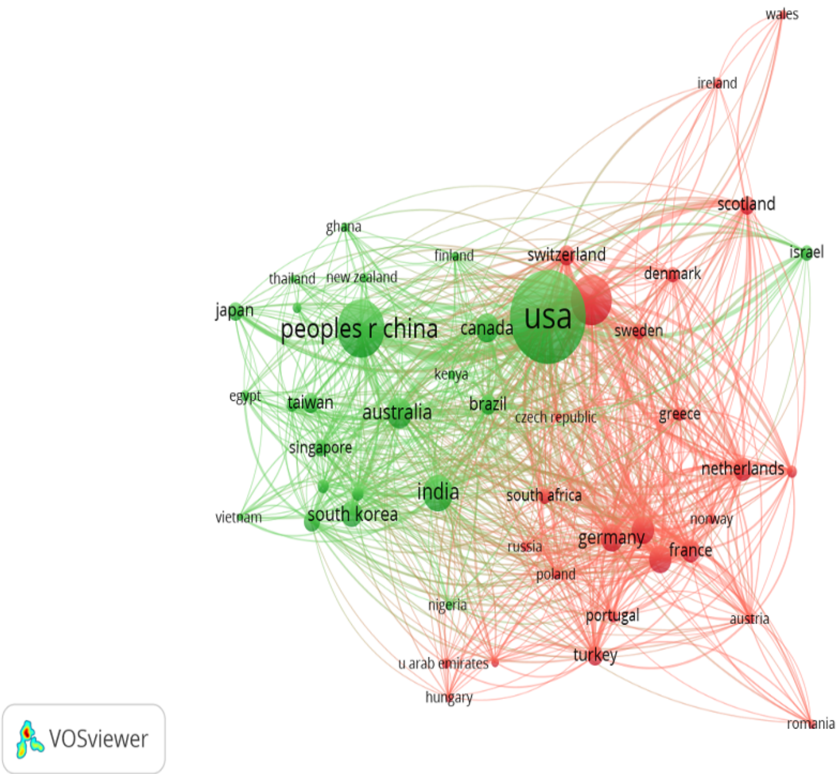


Figure 7. Network of cocited references.

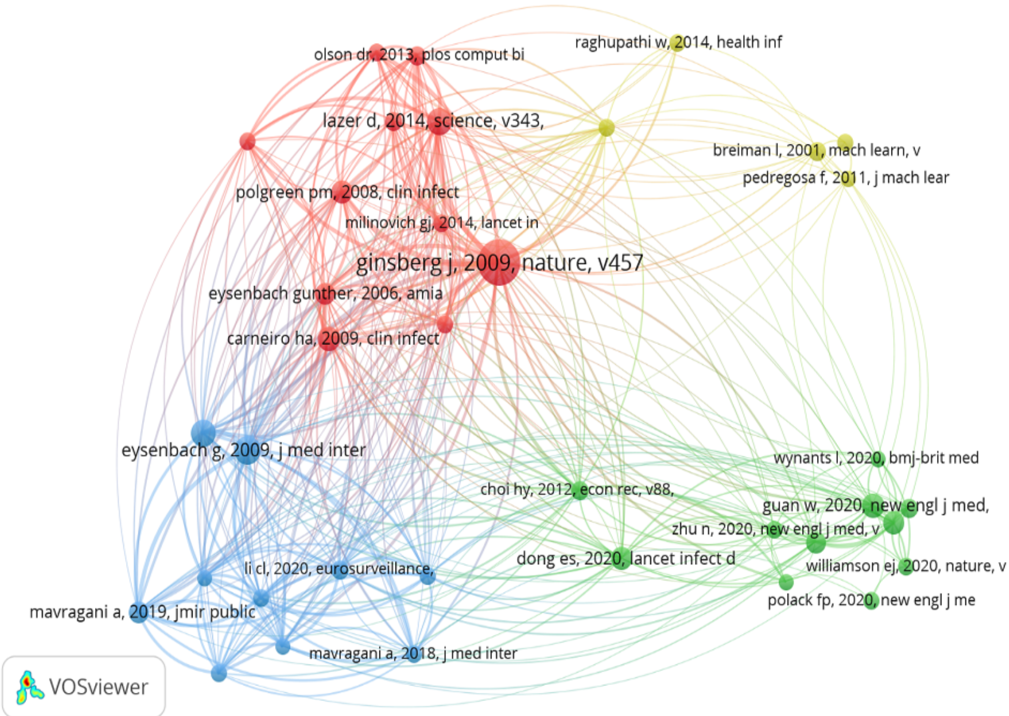
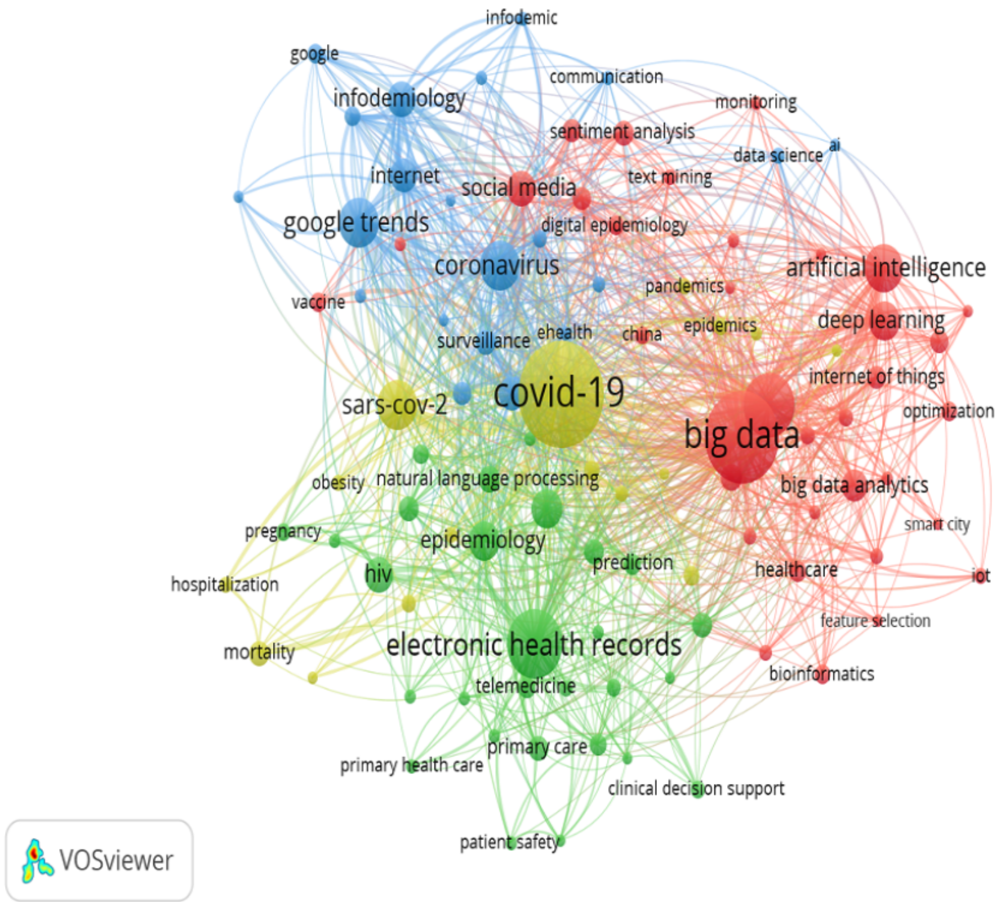


Table 4. Summary of the top 15 most cited references.

Reference	Citations, n	Disease	Data source
Ginsberg et al [31]	185	Influenza	Google Trends
Eysenbach [9]	74	Influenza	NA[a]
Nuti et al [32]	69	NA	Google Trends
Lazer et al [33]	67	Influenza	Google Flu
Carneiro and Mylonaki [34]	54	NA	Google Trends
Zhou et al [35]	49	COVID-19	Electronic health records
Dong et al [36]	49	COVID-19	Twitter feeds and DXY[b]
Polgreen et al [37]	48	Influenza	Yahoo searches
Mavragani and Ochoa [38]	43	NA	Google Trends
Huang et al [39]	42	COVID-19	Electronic medical records
Eysenbach [40]	41	Influenza	Google Trends
Wu et al [41]	34	COVID-19	Electronic medical records
Li et al [42]	33	COVID-19	Internet searches[c] and Weibo index[d]
Santillana et al [43]	31	Influenza	Twitter and Google Trends
Signorini et al [44]	30	Influenza	Twitter

[a]NA: not applicable (eg, a review paper, no particular disease or data source for a case study).
[b]Online platform of real-time COVID-19 cases in China.
[c]Internet searches include Google Trends and Baidu Index.
[d]Weibo is a China-based social media platform.

Figure 8. Co-occurrence networks of author keywords.



The 100 author-derived keywords produced four clusters from the coword analysis (Figure 8). Cluster 1 (yellow-green) is related to public health and infectious diseases, with top keywords such as *COVID-19, SARS-CoV-2, epidemiology,* and *epidemics*. Cluster 2 (green) is related to electronic storage and delivery of health care, with top keywords including *electronic health records*, *clinical decision support*, *primary care*, *epidemiology,* and *telemedicine*. Cluster 3 (blue) involves infodemiology tools, with top keywords including *coronavirus*, *google trends*, *social media*, *infodemiology*, and *surveillance*. Cluster 4 (red) is more coherent and broadly related to big data and artificial intelligence, including top keywords *big data, machine learning, artificial intelligence, deep learning,* and *big data analytics.*

Systematic Review of the Top 20 Papers

Further filtering of the top 20 papers was performed to determine if they met the following criteria: (1) addressed at least one infectious disease and (2) utilized a big data source. A review of these 20 papers (summarized in Table 5) was then performed. These selected studies were mainly characterized by papers that utilized novel data sources, including internet search engine data (Google Trends: n=11; Baidu or Weibo index: n=2; Yahoo: n=1) and social media data (Twitter: n=5). Other data sources included electronic health or medical records (n=3) and Tencent migration data (n=1). The most frequently studied diseases were COVID-19 (n=10) [35,36,39,42,45-50], followed by influenza (n=8) [37,40,43,44,51-54]. Only one study considered the Zika virus [55], and another considered the trio of meningitis, legionella pneumonia, and Ebola [56].

Table 5. Summary of top 20 studies that addressed an infectious disease and utilized a big data source.

Rank	Study	Research objective and key findings	Infectious disease(s)	Big data source
1	Polgreen et al [37]	Used internet search engine data for infectious diseases epidemiology and examined the relationship between Yahoo search queries for influenza and actual influenza occurrence. They estimated linear models, using searches with 1-10–week lead times as explanatory variables to predict the percentage of cultures positive for flu and deaths attributable to pneumonia and influenza in the United States. The fitted models predicted an increase in cultures positive for influenza 1-3 weeks in advance of when they occurred (P<.001), and similar models predicted an increase in mortality attributable to pneumonia and influenza up to 5 weeks in advance (P<.001).	Influenza	Internet search engine
2	Walker et al [45]	The research explored internet activity related to loss of smell in the United States and seven European countries. Spearman rank correlation was used to assess the relationship between loss-of-smell relative search volumes (RSVs), with the daily confirmed cases of COVID-19 and deaths. Strong and significant correlations (P<.05) between daily RSVs related to loss of smell, daily COVID-19 cases, and deaths were found, ranging from 0.633 to 0.952.	COVID-19	Google Trends
3	Effenberger et al [46]	Studied correlations between RSVs and the official COVID-19 cases reported by the European Centre for Disease Control (ECDC) for some selected countries. They opted for time-lag correlation analysis and observed a time lag of –11.5 days being the highest correlation across all investigated countries.	COVID-19	Google Trends
4	Ayyoubzadeh et al [47]	Opted for machine/deep learning with Google Trends data. Linear regression and long short-term memory (LSTM) models were used to estimate COVID-19 cases. They found that the linear regression model had the smaller root mean square error (RMSE) and was the better predictive model. They also found the most predictive factors of the model to be search terms of handwashing, hand sanitizer, and antiseptic topics.	COVID-19	Google Trends
5	Husnayain et al [48]	Considered smaller spatial coverages in their Google Trends analysis. They retrieved data from specific locations and subregions in Taiwan nationwide using defined search terms related to the coronavirus, handwashing, and face masks. Their findings suggest high to moderate correlations between RSVs and COVID-19 cases in Taipei (lag –3), New Taipei (lag –2), Taoyuan (lag –2), Tainan (lag –1), Taichung (lag 0), and Kaohsiung (lag 0).	COVID-19	Google Trends
6	Eysenbach [40]	Found a strong correlation (Pearson r=0.91) between the number of clicks on a keyword-triggered link in Google with epidemiological data from Canada's flu season of 2004-2005.	Influenza	Google Trends
7	Yang et al [51]	To improve the existing Google Flu Trends (GFT), they proposed an influenza tracking model, ARGO (AutoRegression with Google search data), that uses publicly available online search data. Besides having a rigorous statistical foundation, ARGO outperforms the latest GFT version. Not only does ARGO incorporate seasonality in influenza epidemics but it also captures changes in online search behavior over time.	Influenza	ARGO
8	Cook et al [52]	Evaluated the accuracy of each US GFT model by comparing weekly estimates of influenza-like illness (ILI) activity with the US Outpatient Influenza-like Illness Surveillance Network (ILINet). They calculated the correlation and RMSE between model estimates and ILINet for four seasons: pre-H1N1, Summer H1N1, Winter H1N1, and H1N1 overall. Both models' estimates were highly correlated with ILINet pre-H1N1 and over the entire surveillance period, although the original model underestimated the magnitude of ILI activity during the pre-H1N1 phase. The updated model was more correlated with ILINet than the original model during Summer H1N1 (r = 0.95 and 0.29, respectively).	Influenza	Google Trends
9	Yuan et al [53]	Used Baidu, a popular Chinese search index, to model and monitor influenza activity in China. A comprehensive technique was presented for (1) keyword selection, (2) keyword filtering, (3) index composition, and (4) modeling and detection of influenza activity in China. Sequential time series for the selected composite keyword index was significantly correlated with official Chinese influenza cases. Further, 1-month-ahead prediction of flu cases had a considerably small prediction error (mean absolute percent error<11%).	Influenza	Baidu search index
10	Dong et al [36]	Used DXY, an online platform of the Chinese medical community, as a primary data source to develop an online interactive dashboard. The dashboard is hosted by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University, United States. They monitored various Twitter feeds, online news services, and direct communication sent through the dashboard to identify new cases.	COVID-19	DXY, Twitter feeds, online news services

Rank	Study	Research objective and key findings	Infectious disease(s)	Big data source
11	Signorini et al [44]	To explore public concerns regarding rapidly evolving H1N1 activity, used Twitter data and support vector regression to show that estimates of ILI accurately tracked reported disease levels. They retrieved a large sample of public tweets that matched a set of flu-related search terms.	Influenza	Twitter
12	Chew and Eyesenbach [54]	Proposed and evaluated a complementary infoveillance approach using Twitter during the 2009 H1N1 pandemic. They performed a content analysis of tweets and validated Twitter as a real-time content and sentiment tracking tool. Infovigil, an infoveillance technology, was used to record more than 2 million Twitter postings with the terms "swine flu" and "H1N1." According to content analysis, resource-related posts were most commonly shared (52.6%). Misinformation occurred in 4.5% of the cases. News websites were the most popular sources (23.2%), while government and health agencies were linked only 1.5% of the time.	Influenza	Twitter
13	Zhou et al [35]	Used logistic regression models to explore the risk factors associated with in-hospital deaths. They utilized a retrospective, multicenter cohort study that included all adult inpatients with laboratory-confirmed COVID-19.	COVID-19	Electronic health (medical) records
14	Huang et al [39]	This descriptive study detailed descriptive statistics of clinical features of patients infected with COVD-19 in China as extracted from electronic medical records.	COVID-19	Electronic health (medical) records
15	Williamson et al [49]	Developed OpenSAFELY, a secure health analytics platform that maintains patient data in the current data center of a significant provider of primary care electronic health records and serves 40% of all patients in England. OpenSAFELY was used to examine factors associated with COVID-19–related deaths; 10,926 COVID-19–related deaths were pseudonymously linked to primary care records of 17,278,392 persons.	COVID-19	Electronic health (medical) records
16	Wu et al [50]	Used data on monthly airline bookings from the Official Aviation Guide and data on human mobility across more than 300 prefecture-level Chinese cities from the Tencent database. The reports released by the Chinese Center for Disease Control and Prevention provided information on confirmed cases. A susceptible-exposed-infectious-recovered (SEIR) model was used to simulate the epidemics in China's main cities. They concluded that, with a lag time of roughly 1-2 weeks behind the Wuhan outbreak, epidemics were already expanding exponentially in several large cities throughout China, assuming the transmissibility of SARS-CoV-2 was identical domestically and over time.	COVID-19	Tencent Migration data
17	Santillana et al [43]	Presented an ensemble-based machine learning method that leverages data from various sources, including Google searches, Twitter microblogs, and near real-time hospital visit records, to provide nowcast and forecast estimates of influenza activity in the United States. Their method combines multiple ILI activity estimates, generated independently with each data source, into a single prediction of ILI. Evaluation of the predictive ability of their method suggests that it outperforms every prediction using each data source independently. Additionally, it generated estimates 2 and 3 weeks ahead of time with comparable accuracy to real-time forecasts from an autoregressive model and predictions 1 week ahead of GFT's real-time estimates.	Influenza	Google searches, Twitter microblogs, and near real-time hospital visit records
18	Li et al [42]	Evaluated the predictive value of search data from Google Trends and two Chinese social media platforms, Weibo index and Baidu index, for the COVID-19 epidemic in China. They observed that the peak of internet searches and social media data about the COVID-19 outbreak occurred 10-14 days earlier than the peak of daily incidences in China. Internet searches and social media data were highly correlated (r>0.89) with daily incidences.	COVID-19	Google Trends, Weibo index, Baidu index
19	Cervellin et al [56]	Compared the reliability of Google Trends in different clinical settings for common diseases with lower media coverage and for less common diseases attracting major media coverage. They carried out a Google Trends search using the keywords "renal colic," "epistaxis," and "mushroom poisoning." Additionally, a second search was carried out for three clinical conditions (ie, "meningitis," "*Legionella pneumophila* pneumonia," and "Ebola fever"). No correlation was observed between Google Trends and epidemiology of renal colics, epistaxis, and mushroom poisoning. When searching the term "mushroom" alone, the Google Trends search generated a seasonal pattern, almost overlapping with the epidemiological profile.	Meningitis, *Legionella* pneumonia, and Ebola	Google Trends

Rank	Study	Research objective and key findings	Infectious disease(s)	Big data source
20	Teng et al [55]	Developed a dynamic forecasting model for Zika virus (ZIKV) based on Google Trends. A strong correlation was found between Zika-related Google Trends and the cumulative numbers of reported cases (confirmed, suspected, and total cases P<.001). Further, an autoregressive integrated moving average (ARIMA) model (0,1,3) was fitted for the dynamic estimation of ZIKV outbreaks. The forecasting results indicated that the ARIMA model, which used the online search data as the external regressor to enhance the forecasting model, is quite similar to the actual data during the Zika virus epidemic.	Zika	Google Trends

Discussion

Principal Findings

Novel big data streams have created interesting opportunities for infectious disease monitoring and control. The review of the top 20 papers suggests the domination of high-volume electronic health records and digital traces such as internet searches and social media. Of note is the relatively increased use of Google Trends. Most studies used Google Trends data by correlating them with official data on disease occurrence, spread, and outbreaks. Some of these studies further adopted nowcasting for disease surveillance. However, using Google Trends for forecasts and predictions in infectious diseases epidemiology fills a gap in the extant literature. Few studies have gone as far as predicting incidents and occurrences, even though data on reported cases of various health concerns and the associated Google Trends data have been correlated in many studies. Predicting the future is hard; hence, more reliable and efficient methodologies are needed for forecasting infectious disease outbreaks.

There are a few drawbacks to digital trace data that should be considered. Many of these data streams miss demographic information such as age and gender, which is essential in almost any epidemiological study. Besides, they represent a growing but still limited population segment, with infants unfeatured and fewer older adults than younger people. Geographic heterogeneity in coverage exists, with underrepresentation in developing countries, although these biases tend to fade and are arguably less pronounced than those found in traditional global surveillance systems. Further, the retrieved data are subject to spatial and temporal uncertainty. Accordingly, hybrid systems that supplement rather than replace conventional surveillance systems as well as improve prospects for accurate infectious disease models and forecasts should be developed.

Most studies, except for those in the United States and China, were conducted in the European context. Thus, more studies need to test the utility of these big data streams for infectious disease epidemiology in the context of more countries, especially in Africa. Future research questions should ask if any cross-cultural differences between countries affect the adoption and use of big data in infectious disease epidemiology.

The vast majority of infectious diseases have a global distribution. Apart from the coronavirus, influenza, Zika, and Ebola virus outbreaks that are featured in our review, the utility of these big data sources for more infectious diseases should be studied.

Limitations

A few limitations were inherent in our study. First, like any bibliometric study, we are limited by the search terms and database used. This study utilized English publications from the WoS core collection; therefore, relevant publications may have been missed. However, our choice of WoS was informed by its greater coverage of high-impact journals. Second, some studies may have been published after we concluded document extraction. Accordingly, this study does not claim to be exhaustive but rather extensive.

Future Research Agenda and Conclusions

The bibliometric study identified the United States and China as research leaders in this field, with most affiliations from the Harvard Medical School and the University of California. Top authors were Zhang Yi and Li Xingwang. *Journal of Medical Internet Research* and *PLoS One* are the most productive and influential journals in this field. Internet searches and social media data are the most utilized data sources. COVID-19 and influenza were the most studied infectious diseases. The main research themes in this area of research were disease monitoring and surveillance, utility of electronic health (or medical) records, methodology framework for infodemiology tools, and machine/deep learning. Most research papers on big data in infectious diseases epidemiology were published in outlets related to computer science, public health, and health care services.

Opportunities for future research are revealed directly from the results of this study. Integrating multiple surveillance platforms, including big data tools, are critical to better understanding pathogen spread. It is also paramount for the research needs to align with a global view of disease risk. The risk of infectious disease is globally shared in an increasingly connected world. The COVID-19 pandemic, including the rapid global circulation of evolved strains, has emphasized the need for an interdisciplinary, collaborative, global framework for infectious disease research and control. There is a need to empower epidemiologists and public health scientists to leverage insights from big data for infectious disease prevention and control.

Conflicts of Interest

None declared.

References

1. Baker RE, Mahmud AS, Miller IF, Rajeev M, Rasambainarivo F, Rice BL, et al. Infectious disease in an era of global change. Nat Rev Microbiol 2022 Apr 13;20(4):193-205 [FREE Full text] [doi: 10.1038/s41579-021-00639-z] [Medline: 34646006]
2. Hotez PJ. Globalists versus nationalists: Bridging the divide through blue marble health. PLoS Negl Trop Dis 2019 Jul 11;13(7):e0007156 [FREE Full text] [doi: 10.1371/journal.pntd.0007156] [Medline: 31295247]
3. Metcalf CJE, Tatem A, Bjornstad ON, Lessler J, O'Reilly K, Takahashi S, et al. Transport networks and inequities in vaccination: remoteness shapes measles vaccine coverage and prospects for elimination across Africa. Epidemiol Infect 2015 May;143(7):1457-1466 [FREE Full text] [doi: 10.1017/S0950268814001988] [Medline: 25119237]
4. Kasson PM. Infectious disease research in the era of big data. Annu Rev Biomed Data Sci 2020 Jul 20;3(1):43-59. [doi: 10.1146/annurev-biodatasci-121219-025722]
5. Big data. Wikipedia. URL: https://en.wikipedia.org/w/index.php?title=Big_data&oldid=1089502025 [accessed 2022-05-26]
6. Hay SI, George DB, Moyes CL, Brownstein JS. Big data opportunities for global infectious disease surveillance. PLoS Med 2013 Apr 2;10(4):e1001413 [FREE Full text] [doi: 10.1371/journal.pmed.1001413] [Medline: 23565065]
7. Milinovich GJ, Soares Magalhães RJ, Hu W. Role of big data in the early detection of Ebola and other emerging infectious diseases. Lancet Glob Health 2015 Jan;3(1):e20-e21 [FREE Full text] [doi: 10.1016/S2214-109X(14)70356-0] [Medline: 25539964]
8. Wong ZS, Zhou J, Zhang Q. Artificial intelligence for infectious disease big data analytics. Infect Dis Health 2019 Feb;24(1):44-48. [doi: 10.1016/j.idh.2018.10.002] [Medline: 30541697]
9. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
10. Salathé M. Digital epidemiology: what is it, and where is it going? Life Sci Soc Policy 2018 Jan 04;14(1):1 [FREE Full text] [doi: 10.1186/s40504-017-0065-7] [Medline: 29302758]
11. Thacker SB, Stroup DF. Future directions for comprehensive public health surveillance and health information systems in the United States. Am J Epidemiol 1994 Sep 01;140(5):383-397. [doi: 10.1093/oxfordjournals.aje.a117261] [Medline: 8067331]
12. Woolhouse MEJ, Rambaut A, Kellam P. Lessons from Ebola: Improving infectious disease surveillance to inform outbreak management. Sci Transl Med 2015 Sep 30;7(307):307rv5 [FREE Full text] [doi: 10.1126/scitranslmed.aab0191] [Medline: 26424572]
13. Farhadloo M, Winneg K, Chan MS, Hall Jamieson K, Albarracin D. Associations of topics of discussion on Twitter with survey measures of attitudes, knowledge, and behaviors related to Zika: probabilistic study in the United States. JMIR Public Health Surveill 2018 Feb 09;4(1):e16 [FREE Full text] [doi: 10.2196/publichealth.8186] [Medline: 29426815]
14. Lu FS, Hou S, Baltrusaitis K, Shah M, Leskovec J, Sosic R, et al. Accurate influenza monitoring and forecasting using novel internet data streams: a case study in the Boston metropolis. JMIR Public Health Surveill 2018 Jan 09;4(1):e4 [FREE Full text] [doi: 10.2196/publichealth.8950] [Medline: 29317382]
15. Mavragani A, Sampri A, Sypsa K, Tsagarakis KP. Integrating smart health in the US health care system: infodemiology study of asthma monitoring in the Google era. JMIR Public Health Surveill 2018 Mar 12;4(1):e24 [FREE Full text] [doi: 10.2196/publichealth.8726] [Medline: 29530839]
16. van Lent LG, Sungur H, Kunneman FA, van de Velde B, Das E. Too far to care? Measuring public attention and fear for Ebola using Twitter. J Med Internet Res 2017 Jun 13;19(6):e193 [FREE Full text] [doi: 10.2196/jmir.7219] [Medline: 28611015]
17. Wongkoblap A, Vadillo MA, Curcin V. Researching mental health disorders in the era of social media: systematic review. J Med Internet Res 2017 Jun 29;19(6):e228 [FREE Full text] [doi: 10.2196/jmir.7215] [Medline: 28663166]
18. Xu C, Yang H, Sun L, Cao X, Hou Y, Cai Q, et al. Detecting lung cancer trends by leveraging real-world and internet-based data: infodemiology study. J Med Internet Res 2020 Mar 12;22(3):e16184 [FREE Full text] [doi: 10.2196/16184] [Medline: 32163035]
19. Mavragani A, Gkillas K. COVID-19 predictability in the United States using Google Trends time series. Sci Rep 2020 Nov 26;10(1):20693. [doi: 10.1038/s41598-020-77275-9] [Medline: 33244028]
20. Liang C, Qiao S, Olatosi B, Lyu T, Li X. Emergence and evolution of big data science in HIV research: Bibliometric analysis of federally sponsored studies 2000-2019. Int J Med Inform 2021 Oct;154:104558 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104558] [Medline: 34481301]
21. Hassan NR, Loebbecke C. Engaging scientometrics in information systems. J Inf Technol 2017 Mar 01;32(1):85-109. [doi: 10.1057/jit.2015.29]

22. Chen C, Song M. Visualizing a field of research: a methodology of systematic scientometric reviews. PLoS One 2019 Oct 31;14(10):e0223994 [FREE Full text] [doi: 10.1371/journal.pone.0223994] [Medline: 31671124]
23. Aria M, Cuccurullo C. bibliometrix : An R-tool for comprehensive science mapping analysis. J Informetr 2017 Nov;11(4):959-975. [doi: 10.1016/j.joi.2017.08.007]
24. Zhang Y, Huang Y, Porter AL, Zhang G, Lu J. Discovering and forecasting interactions in big data research: a learning-enhanced bibliometric study. Technol Forecast Soc Change 2019 Sep;146:795-807. [doi: 10.1016/j.techfore.2018.06.007]
25. Harzing A, Alakangas S. Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. Scientometrics 2015 Nov 26;106(2):787-804. [doi: 10.1007/s11192-015-1798-9]
26. Hirsch JE. An index to quantify an individual's scientific research output that takes into account the effect of multiple coauthorship. Scientometrics 2010 Mar 13;85(3):741-754. [doi: 10.1007/s11192-010-0193-9]
27. Batagelj V, Cerinšek M. On bibliographic networks. Scientometrics 2013 Jan 20;96(3):845-864. [doi: 10.1007/s11192-012-0940-1]
28. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech Theory Exp 2008;10:P10008. [doi: 10.1088/1742-5468/2008/10/P10008]
29. Su H, Lee P. Mapping knowledge structure by keyword co-occurrence: a first look at journal papers in Technology Foresight. Scientometrics 2010 Jun 22;85(1):65-79. [doi: 10.1007/s11192-010-0259-8]
30. van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 2010 Aug 31;84(2):523-538 [FREE Full text] [doi: 10.1007/s11192-009-0146-3] [Medline: 20585380]
31. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
32. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of google trends in health care research: a systematic review. PLoS One 2014 Oct 22;9(10):e109583 [FREE Full text] [doi: 10.1371/journal.pone.0109583] [Medline: 25337815]
33. Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. Science 2014 Mar 14;343(6176):1203-1205. [doi: 10.1126/science.1248506] [Medline: 24626916]
34. Carneiro H, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. Clin Infect Dis 2009 Nov 15;49(10):1557-1564. [doi: 10.1086/630200] [Medline: 19845471]
35. Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet 2020 Mar 28;395(10229):1054-1062 [FREE Full text] [doi: 10.1016/S0140-6736(20)30566-3] [Medline: 32171076]
36. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020 May;20(5):533-534 [FREE Full text] [doi: 10.1016/S1473-3099(20)30120-1] [Medline: 32087114]
37. Polgreen P, Chen Y, Pennock D, Nelson F. Using internet searches for influenza surveillance. Clin Infect Dis 2008 Dec 01;47(11):1443-1448. [doi: 10.1086/593098] [Medline: 18954267]
38. Mavragani A, Ochoa G. Google trends in infodemiology and infoveillance: methodology framework. JMIR Public Health Surveill 2019 May 29;5(2):e13439 [FREE Full text] [doi: 10.2196/13439] [Medline: 31144671]
39. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
40. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. AMIA Annu Symp Proc 2006;2006:244-248 [FREE Full text] [Medline: 17238340]
41. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA 2020 Apr 07;323(13):1239-1242. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
42. Li C, Chen L, Chen X, Zhang M, Pang C, Chen H. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. Euro Surveill 2020 Mar;25(10):2000199 [FREE Full text] [doi: 10.2807/1560-7917.ES.2020.25.10.2000199] [Medline: 32183935]
43. Santillana M, Nguyen AT, Dredze M, Paul MJ, Nsoesie EO, Brownstein JS. Combining search, social media, and traditional data sources to improve influenza surveillance. PLoS Comput Biol 2015 Oct 29;11(10):e1004513 [FREE Full text] [doi: 10.1371/journal.pcbi.1004513] [Medline: 26513245]
44. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. PLoS One 2011 May 04;6(5):e19467 [FREE Full text] [doi: 10.1371/journal.pone.0019467] [Medline: 21573238]
45. Walker A, Hopkins C, Surda P. Use of Google Trends to investigate loss-of-smell-related searches during the COVID-19 outbreak. Int Forum Allergy Rhinol 2020 Jul 15;10(7):839-847 [FREE Full text] [doi: 10.1002/alr.22580] [Medline: 32279437]

46. Effenberger M, Kronbichler A, Shin JI, Mayer G, Tilg H, Perco P. Association of the COVID-19 pandemic with Internet Search Volumes: A Google Trends Analysis. Int J Infect Dis 2020 Jun;95:192-197 [FREE Full text] [doi: 10.1016/j.ijid.2020.04.033] [Medline: 32305520]
47. Ayyoubzadeh SM, Ayyoubzadeh SM, Zahedi H, Ahmadi M, R Niakan Kalhori S. Predicting COVID-19 incidence through analysis of Google Trends data in Iran: data mining and deep learning pilot study. JMIR Public Health Surveill 2020 Apr 14;6(2):e18828 [FREE Full text] [doi: 10.2196/18828] [Medline: 32234709]
48. Husnayain A, Fuad A, Su EC. Applications of Google Search Trends for risk communication in infectious disease management: a case study of the COVID-19 outbreak in Taiwan. Int J Infect Dis 2020 Jun;95:221-223 [FREE Full text] [doi: 10.1016/j.ijid.2020.03.021] [Medline: 32173572]
49. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors associated with COVID-19-related death using OpenSAFELY. Nature 2020 Aug 08;584(7821):430-436 [FREE Full text] [doi: 10.1038/s41586-020-2521-4] [Medline: 32640463]
50. Wu J, Leung K, Leung G. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. Lancet 2020 Feb 29;395(10225):689-697 [FREE Full text] [doi: 10.1016/S0140-6736(20)30260-9] [Medline: 32014114]
51. Yang S, Santillana M, Kou SC. Accurate estimation of influenza epidemics using Google search data via ARGO. Proc Natl Acad Sci U S A 2015 Nov 24;112(47):14473-14478 [FREE Full text] [doi: 10.1073/pnas.1515373112] [Medline: 26553980]
52. Cook S, Conrad C, Fowlkes AL, Mohebbi MH. Assessing Google flu trends performance in the United States during the 2009 influenza virus A (H1N1) pandemic. PLoS One 2011 Aug 19;6(8):e23610 [FREE Full text] [doi: 10.1371/journal.pone.0023610] [Medline: 21886802]
53. Yuan Q, Nsoesie EO, Lv B, Peng G, Chunara R, Brownstein JS. Monitoring influenza epidemics in china with search query from baidu. PLoS One 2013 May 30;8(5):e64323 [FREE Full text] [doi: 10.1371/journal.pone.0064323] [Medline: 23750192]
54. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
55. Teng Y, Bi D, Xie G, Jin Y, Huang Y, Lin B, et al. Dynamic forecasting of Zika epidemics using Google Trends. PLoS One 2017 Jan 6;12(1):e0165085 [FREE Full text] [doi: 10.1371/journal.pone.0165085] [Medline: 28060809]
56. Cervellin G, Comelli I, Lippi G. Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. J Epidemiol Glob Health 2017 Sep;7(3):185-189 [FREE Full text] [doi: 10.1016/j.jegh.2017.06.001] [Medline: 28756828]

Abbreviations

WoS: Web of Science

Edited by A Mavragani; submitted 30.08.22; peer-reviewed by Z Wang, P Berchialla; comments to author 10.10.22; revised version received 21.10.22; accepted 29.11.22; published 31.03.23.

Please cite as:
Amusa LB, Twinomurinzi H, Phalane E, Phaswana-Mafuya RN
Big Data and Infectious Disease Epidemiology: Bibliometric Analysis and Research Agenda
Interact J Med Res 2023;12:e42292
URL: https://www.i-jmr.org/2023/1/e42292
doi:10.2196/42292
PMID:36913554

Review

Nondrug Intervention for Opportunistic Infections in Individuals With Hematological Malignancy: Systematic Review

Nor Asiah Muhamad[1], MBBS, MPH; Nur Hasnah Ma'amor[1], BSc, MSc, PhD; Normi Mustapha[2], BSc, MSc; Fatin Norhasny Leman[1], BSc, MSc; Izzah Athirah Rosli[1], BSc; Marilyn Umar[3], MBBS, MPH; Tahir Aris[4], MD, MPH; Nai Ming Lai[5], MBBS, MRCP, MRCPCH

[1]Sector for Evidence-based Healthcare, National Institutes of Health, Ministry of Health, Shah Alam, Malaysia

[2]Faculty Science and Technology, Open University, Kuala Lumpur, Malaysia

[3]Non-Communicable Disease Section, Sarawak State Health Department, Ministry of Health, Sarawak, Malaysia

[4]Director's Office, Institutes for Medical Research, National Institutes of Health, Ministry of Health, Shah Alam, Malaysia

[5]School of Medicine, Faculty of Health & Medical Sciences, Taylor's University, Subang Jaya, Malaysia

Corresponding Author:
Nor Asiah Muhamad, MBBS, MPH
Sector for Evidence-based Healthcare
National Institutes of Health
Ministry of Health
1, Persiaran Setia Murni, Setia Alam
Shah Alam, 40170
Malaysia
Phone: 60 123390723
Email: norasiahdr@gmail.com

Abstract

Background: Hematological malignancies disturb the blood, lymph nodes, and bone marrow. Taking medications for treating opportunistic infections (OIs) in these individuals may enhance the risk of medication interaction as well as adverse drug reactions.

Objective: This review aims to evaluate the effectiveness of nondrug interventions in reducing OIs among patients with hematological cancers.

Methods: The PubMed, CENTRAL (Cochrane Central Register of Controlled Trials), and Embase databases were searched on December 26, 2022, for all randomized controlled trials (RCTs). The primary endpoint was OIs. The quality of included studies was assessed by the Cochrane Risk-of-Bias tool.

Results: A total of 6 studies were included in this review with 4 interventions: (1) types of mouthwash received, (2) presence of coating on central venous catheters (CVCs), (3) use of well-fitted masks, and (4) types of diet consumed. The results were presented in 8 different comparisons: (1) chlorhexidine-nystatin versus saline mouth rinse, (2) chlorhexidine versus saline mouth rinse, (3) nystatin versus saline mouth rinse, (4) chlorhexidine silver sulfadiazine–coated CVCs versus uncoated catheters, (5) well-fitted masks versus no mask, (6) amine fluoride-stannous fluoride versus sodium fluoride mouthwash, (7) low-bacterial diet versus standard hospital diet, and (8) herbal versus placebo mouthwash. No clear differences were reported in any of the outcomes examined in the first 3 comparisons. There were also no clear differences in the rate of catheter-related bloodstream infection or insertion site infection between the use of chlorhexidine silver sulfadiazine–coated CVCs versus uncoated catheters in the patients. Further, no significant differences were seen between patients who used a well-fitted mask and those without a mask in the incidence of OI. The all-cause mortality and mortality due to OI were similar between the 2 groups. There was no clear difference in all-cause mortality, although common adverse effects were reported in patients who used sodium fluoride mouthwash compared with those using amine fluoride-stannous fluoride mouthwash. There was no evidence of any difference in the incidence of possible invasive aspergillosis or candidemia between patients who consumed a low-bacterial diet and a standard diet. For the last comparison, no significant difference was seen between patients who received herbal and placebo mouthwash.

Conclusions: Very limited evidence was available to measure the effectiveness of nondrug interventions in hematological cancers. The effectiveness of the interventions included in this review needs to be evaluated further in high-quality RCTs in a dedicated setting among patients with hematological malignancies.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42020169186; https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=169186

(Interact J Med Res 2023;12:e43969) doi:10.2196/43969

KEYWORDS

nondrug; intervention; opportunistic infection; hematological malignancies

Introduction

Background

Hematological cancer or malignancies are cancers of the blood, bone marrow, and lymph nodes that come from either lymphoid cell lines or myeloid. Megakaryocytes, macrophages, granulocytes, mast cells, and erythrocytes are produced by the myeloid cell line. The myeloid cell line is responsible for acute and myelodysplastic syndromes, myeloproliferative disorders, and chronic myelogenous leukemia [1]. Other cells such as plasma, T and B cells, as well as natural killer cells or large granular lymphocytes are produced by the lymphoid cell line. This lymphoid cell line is responsible for lymphomas, lymphocytic leukemia, and myelomas [1]. In the United States and United Kingdom, 16 (13.6%) leukemia cases are reported per 100,000 people [2,3]. The clinical course and prognosis vary as it may depend on the existence or the type of genetic mutation itself. Chemotherapy, radiation, immunotherapy, and bone marrow transplantation are examples of active treatment options.

A healthy individual with a strong immune system is usually resistant to opportunistic infections (OIs) caused by pathogens such as protozoa, fungi, viruses, or bacteria [1]. However, if the immune system is weakened, the pathogens have a better chance of infecting the individual. Cancer therapy to treat hematological cancer can weaken a patient's immune system, making them vulnerable to OIs. OIs affect 5%-60% of individuals with hematological malignancies [4-7], with the mortality rate ranging between 15% and 41% [5]. Broad-spectrum intravenous antimicrobials are used to treat OIs, which are first used empirically and then adapted to the patient's specific needs [8]. Viruses causing the OIs can be treated with antiviral therapies.

Most individuals with hematological cancers are more susceptible to OIs due to both immunological impairment of cancer therapies and disturbances caused by the disease itself [9,10]. In a situation where patients with cancer become neutropenic, they are isolated as a precautionary measure to prevent OIs, while the length of stay in the isolation room varies depending on their medical condition. This may cause wide range of psychological burdens such as depression, anxiety, and stress as a result of staying in the isolation room [11]. It has been reported that the psychological well-being of patients with cancer influences their treatment response and long-term prognosis [12-14]. Nondrug therapies such as isolation have been used to reduce the OIs. The relation between OIs and cancer treatment is bidirectional. Treatment for various types of cancers causes immunosuppression and that may cause OIs among patients with cancer. Treatment knowledge, understanding, and adherence will improve the overall prognosis [15,16]. Therefore, the objective of this review is to evaluate the efficiency of nondrug interventions in preventing OIs in individuals with hematological cancers. As a result, this study was performed to assess the safety and efficacy of nondrug interventions for the prevention of OIs in individuals with hematological cancer or malignancies.

Types of Nondrug Interventions

There are three major types of nondrug interventions. The first type of intervention is the barrier method to prevent potential transmission such as wearing protective equipment or gloves, cleaning of bed sheets and clothes, and the use of masks. The second type of intervention is complete elimination of the causal agents such as fumigation on a regular basis, regular cleaning of potential microbe-harboring goods including toys and carpets, elimination of houseplants that are likely to be a reservoir for microorganisms, and the use of mouthwash as a personal hygiene modification [17]. The third type of intervention is the physical method such as applying positive pressure (controls the air quality in-flow) or using a high-efficiency particulate absorption (HEPA) filtered room to improve the hospital environment.

In patients with neutropenia, primary infections can be a result of mild injuries caused by venous and vascular catheters, which can spread through the bloodstream and eventually result in soft tissue and skin infection [18]. Neutropenic diet is also a part of the barrier method to reduce the risk of infection. It consists of a low-bacteria diet, such as meals cooked thoroughly or with boiled water. Despite its widespread use, the effectiveness of a neutropenic diet in patients undergoing chemotherapy remains debatable [19-21].

Physical methods such as using an HEPA filter may help prevent contact with pathogens that exist in soil or plants and can reduce OIs in vulnerable individuals [22-26].

Personal hygiene modifications, such as the use of mouthwash [27], chlorhexidine baths [28], and frequent cleaning of surfaces, may decrease the microorganisms and completely eliminate them [29].

How the Intervention Might Work

Targeted nondrug interventions to treat OIs act in 3 ways: removing the cause of illness, reducing the contact with infectious agents, and decreasing the risk of microorganism invasion.

The use of personal protective equipment (PPE) and HEPA filters are examples of barrier methods to reduce one's exposure to an infective agent. HEPA filters are reported to eliminate 99.97% of particles with a diameter over 0.3 m, which includes most microorganisms [30]. The use of PPE such as surgical and N95 masks has demonstrated beneficial effect in reducing the transmission of influenza virus [31,32]. Hand hygiene has been shown to prevent the transfer of harmful microorganisms, such as central line–acquired bloodstream infections and methicillin-resistant *Staphylococcus aureus* [33,34]. The use of gloves may limit the quantity of microbes' transmission through skin contact. Alcohol-based antiseptics have been suggested to reduce pathogen transmission by denaturing the microorganisms' proteins [1].

Justification for This Review

OIs contribute to mortality and morbidity in individuals with hematological cancer. In these patients, nondrug interventions, such as the use of PPE and modifying the environment, are commonly used, as they may lessen the need for medication prophylaxis and therapy; however, they are not without hazards and expenses. As a result, synthesizing the existing evidence on the safety and efficacy of these interventions is critical. In that regard, this study was performed to assess the safety and efficacy of nondrug interventions for OI prevention in individuals with hematological cancer or malignancies.

Methods

Overview

We conducted a systematic review by following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [35]. The PRISMA checklist was used while writing this report [36].

Eligibility Criteria

For this review, we followed the PICOS mnemonic, where "P" represents adult patients with hematological cancer, and we included all individuals or patients regardless of the stage of disease, type of hematological cancer or malignancy, morbidity, and the received treatment modality; "I" is any nondrug intervention included either alone or in combination with other therapies such as alteration of patient/caregiver behavior, alteration of the home-based environment, and hospital-based environmental control measures. We also included other interventions defined by the study authors as nondrug intervention. The control "C" group is defined as individuals or patients who did not receive the nondrug intervention, or those who obtained prophylactic pharmacological medications or therapies. The outcomes "O" are defined as either OIs or bacterial OIs, mortality due to OIs, all-cause mortality, hospitalization duration in days, chemotherapy interruption (number of episodes or duration of interruption), quality of life, and adverse effects related to the intervention. The included study design "S" is all randomized controlled trials (RCTs) or cluster RCTs that are published in full texts or abstracts. We excluded studies when the intervention contained pharmacological measures or any studies with the crossover design due to concerns of the "carryover" effects.

Search Methods for Identification of Studies

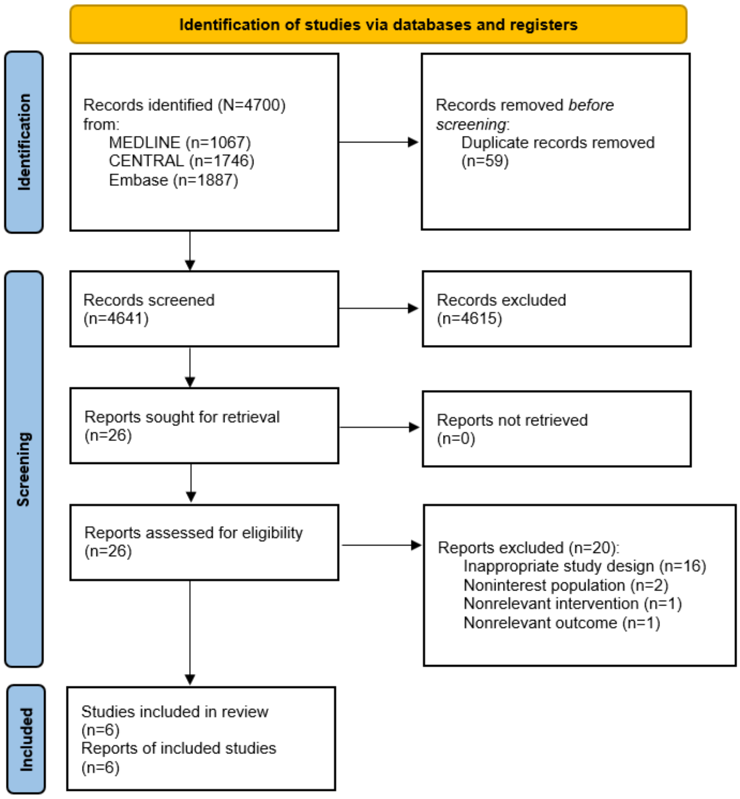
A comprehensive systematic literature search was performed on various electronic databases, Embase, Cochrane Central Register of Controlled Trials (CENTRAL), WHO (World Health Organization) International Clinical Trials Registry Platform (ICTRP), and PubMed, to identify relevant studies from inception to December 2022. A strategy search for PubMed was established and modified for use in the other databases. During the search, keywords and equivalent MeSH (Medical Subject Headings) phrases were combined when applicable, with no language or publication year restrictions. The search strategies for MEDLINE are presented in Multimedia Appendix 1 and CENTRAL in Multimedia Appendix 2. We also searched the abstract records from the following conferences organized by societies that are related to blood cancer or malignancies: The European Society for Medical Oncology Annual Congress, The European Hematology Association (EHA) Conference, American Association for Cancer Research and American Society of Clinical Oncology Conference, American Society of Hematology Meeting, and Virginia Association of Hematologists and Oncologists (VAHO) Spring Membership Conference. We also searched the following databases for ongoing studies: metaRegister of Controlled Trials [37], International Clinical Trials Registry Platform [38], and ClinicalTrials.gov [39]. During the searches, there was no limitation or restriction on the language of the article.

Strategy for Data Collection and Analysis

Selection of Studies

Two authors (NAM and NHM) independently screened all the study titles and abstracts and excluded studies that were not eligible. We resolved any discrepancies through discussion or by consultation with the third review author (TA). We followed the Cochrane Handbook for Systematic Reviews of Interventions [40] for reporting biases. We retrieved the full-text study reports/publications and 2 other review authors (NM and FNL) independently screened the full text to identify studies for inclusion, as well as identifying and recording reasons for exclusion of the ineligible studies. We resolved any disagreement through discussion or consulted a third review author (NML) to make the final judgment. We identified and excluded duplicates and collated multiple reports of the same study such that each study rather than each report is the unit of interest in the review. We recorded the selection process in sufficient detail to complete a PRISMA flow diagram (Figure 1) and tabulated the characteristics of the included and excluded studies [35].

Figure 1. PRISMA flow diagram of studies selection. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.



Data Extraction and Management

Two authors (NAM and IAR) independently extracted the data and completed the data extraction using a standardized data collection form for study characteristics and outcome data. The form contained information on methods, total participants, interventions, comparisons, outcomes, and study design. We resolved disagreements by consensus.

Assessment of the Risk of Bias in Included Studies

Two review authors (NAM and MU) independently assessed the quality of the included studies using the Cochrane Risk-of-Bias Tool [40]. The risk of bias is assessed using the following 6 domains: (1) random sequence generation, (2) allocation concealment, (3) blinding of individuals or participants and personnel, (4) blinding of the outcome assessment, (5) incomplete data outcome, and (6) selective reporting and other bias. We summarized the risk-of-bias judgments for each of the domains listed in the "risk-of-bias" table included in Multimedia Appendix 3 for the 6 included studies and present our overall assessment of the risk of bias using a "risk-of-bias" graph (Figure 2) and "risk-of-bias" summary (Figure 3). Any disagreement among the review authors was resolved by discussion to achieve a consensus.

Figure 2. Risk of bias graph on review authors' judgments about each risk of bias item presented as percentages across all included studies.

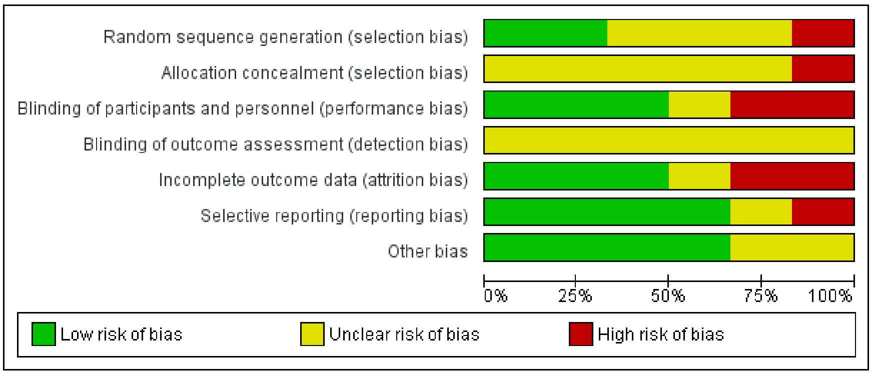
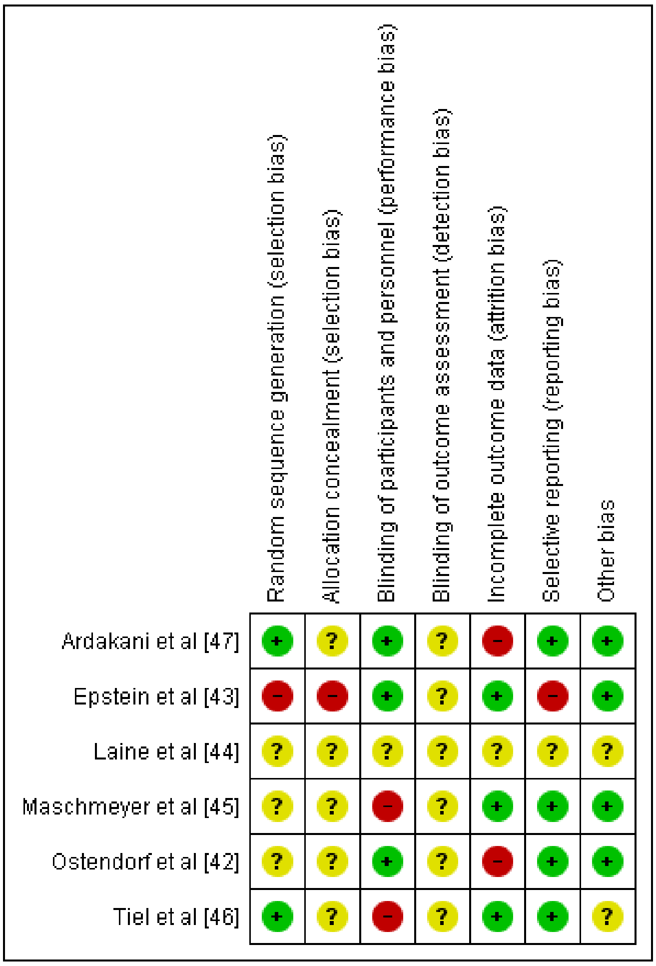


Figure 3. Risk of bias summary on review authors' judgments about each risk of bias item for each included study.



Data Synthesis

Dichotomous data were determined as risk ratio (RR) and reported with the 95% CI, whereas continuous data were observed as mean difference (MD) and the respective 95% CI [40]. Heterogeneity of treatment effects was measured using the χ^2 test and the degree of heterogeneity was assessed using the I^2 statistic, with the value of 75% or higher indicating substantial heterogeneity [40]. Two authors (NAM and NM) performed data analysis using Review Manager version 5.4 [41]. Meta-analysis was not possible because we could not include more than 1 study that provided usable data in a single comparison.

We created "summary of findings" (SOF) tables using the software GRADEpro (McMaster University and Evidence Prime Inc.; Tables S1-S8 in Multimedia Appendix 4). In the SOF tables, we included the following major outcomes, regardless of whether the outcome data were available [40]: (1) OI, as reported variously by the study authors; (2) all-cause death or mortality; (3) death or mortality that is associated with OI; (4) duration of hospitalization; (5) quality of life; and (6) adverse effects (either chemotherapy associated or attributable to the intervention examined).

In a comparison that evaluated chlorhexidine silver sulfadiazine–coated central venous catheters (CVCs) versus noncoated catheters [42], the major outcome reported was that the catheters were associated with various infections (catheter colonization, catheter-associated bloodstream infection, and insertion site infection). For this review, we have grouped these outcomes together with our predefined outcome of "OI" and have displayed these outcomes in the SOF tables.

Results

Findings From The Search Strategies

A total of 4700 records (1746 records from CENTRAL [Cochrane Central Register of Controlled Trials], 1067 records from MEDLINE, and 1887 records from Embase) were successfully retrieved. No additional records were identified through other sources, such as online conference archives and clinical trial registries. After removing duplicates, 4641 records remained. Subsequently, 4615 records were excluded. We obtained 26 records to be assessed for eligibility. Of these, 20 articles were excluded. Ultimately, 6 eligible articles or studies were included. The PRISMA flow diagram of the selection of studies is depicted in Figure 1.

Included Studies

The 6 included studies were published as full papers [42-47]. Of these 6 studies, 2 were performed in Germany [42,45], 1 was conducted in Canada [43], 1 in Finland [44], 1 in The Netherlands [46], and 1 in Iran [47]. Table 1 describes in detail the characteristics of the studies included in this review.

Table 1. Characteristics of the included articles or studies.

Study	Study design	Participants	Interventions	Findings or outcomes	Notes
Epstein et al [43]	A single-center 4-arm randomized controlled trial conducted in Canada (published in 1992)	A total of 99 adult (>18 years old) patients diagnosed with leukemia with severe neutropenia. All patients were treated with chemotherapy with or without bone marrow transplantation. No exclusion criteria were stated.	Chlorhexidine, chlorhexidine + nystatin, nystatin, and saline rinse groups.	Chemotherapy-related oral complications such as oral mucositis, ulceration, gingivitis, and dental plaque.	This study was conducted for eligible individuals or patients who were admitted under the Leukemia/Bone Marrow Transplantation Service, Vancouver General Hospital, Canada. The study obtained the ethical review board approval of the Vancouver General Hospital. There was no conflicts of interest reported between the authors.
Laine et al [44]	A single-center randomized controlled trial conducted in Finland (study period was not stated)	A total of 76 adult patients who had been diagnosed with Hodgkin disease/non-Hodgkin lymphoma, which was confirmed by histological analysis. These patients received chemotherapy with curative intent. Their estimated life expectancy was <1 year. Patients were eligible if they did not have other concomitant disease, were receiving cancer therapy medication only, and with a Karnofsky Performance Status score of ≥60. Exclusion criteria were not explicitly stated.	Using a mouthwash containing 0.025% fluoride known as amine fluoride-stannous fluoride or 0.05% of sodium fluoride solution.	All-cause mortality and adverse effects such as stinging pain near the mouth, staining teeth, nausea, bad taste, and combined adverse effects, as well as salivary microbial count and salivary secretion rate.	This study was partly supported by the pharmaceutical industry (Gaba International Ltd., Basle, Switzerland) and partly by the Linda Gadd Foundation of the Finska Lakaresallskapet.
Maschmeyer et al [45]	A prospective randomized study conducted in Germany from February 2004 to October 2005	A total of 80 hospitalized patients (>18 years old) who had received either chemotherapy/intensive myelosuppressive therapy or allogeneic stem cell transplantation for acute leukemia. Exclusion criteria were clearly stated.	Patients received the standard or routine prophylaxis with or without a well-fitted FFP2[a] mask.	The primary outcome was the occurrence of a possible, probable, or proven aspergillosis. Secondary outcomes were tolerability, patient compliance with wearing masks and other procedures related to infection prevention, mortality, administration of systemic antifungal agents for empirical or targeted treatment of invasive mycoses, and diagnosis of fungal infection within 2 weeks after the study.	The protocol of the study was approved by the Human Ethics Committee of the Charité University of Berlin, Germany. This study was performed in collaboration with 3M Germany, which provided the masks for free.
Ostendorf et al [42]	A single-center randomized controlled trial conducted in Germany	Individuals or patients with hematological cancer or malignancy. These patients needed or were on a CVC[b] for a minimum of 7 days. A total of 184 CVCs were evaluated. The exclusion criteria were not mentioned in this study.	Chlorhexidine silver sulfadiazine–impregnated CVC versus nonimpregnated CVC.	Catheter colonization (mentioned as “catheter-related bacteremia”), mortality rate, and catheter-associated local infection.	This study received funding from the industrial partners (eg, from the distributor and manufacturer of catheter).

Study	Study design	Participants	Interventions	Findings or outcomes	Notes
Tiel et al [46]	A randomized and controlled pilot study conducted in The Netherlands from February to December 2003	About 20 individuals or patients (>18 years old) with acute leukemia who were on chemotherapy treatment or remission induction. No exclusion criteria were stated.	The authors discussed 2 categories of diet: low bacterial diet versus standard or normal hospital diet. Patients in both groups received antimicrobial prophylaxis.	Colonization of feces with *Candida* species or Gram-negative bacilli. Secondary findings were infection parameters and the total societal costs.	Data on stool colonization were very skewed and may not be analyzable. The study was financially supported by the Dutch Board and Profileringsfonds of the University Hospital Maastricht. The Medical Ethics Committee of the University Hospital Maastricht, The Netherlands approved the study protocol.
Ardakani et al [47]	A single-center, double-blind, randomized, placebo-controlled clinical trial conducted in Iran from April 2011 to August 2012	A total of 60 patients were enrolled (nonsmokers, aged ≥15 years, able to gargle the mouthwash solution, and capable of reading and communicating with staff). Exclusion criteria were no cooperation during the study, allergic reactions to herbal mouthwash, and failure to adhere to the oral health protocol due to any changes of their health condition.	A herbal mouthwash containing 1% dried extract of *Matricaria recutita*, 1% peppermint oil, and 99% ethanol. By contrast, the placebo mouthwash had similar taste, smell, and color, but contained 0.02% edible red color, 0.5% chlorophyllin color, 13% ethanol, and 71.5% distilled sterile water.	Development of oral mucositis assessed using the NCI-CTC[c], with an evaluation of its duration, which was assessed by the number of days with the infection.	The study was registered with the Iranian Registry of Clinical Trials and approved by the Ethical Committee of Shahid Beheshti University of Medical Sciences.

[a]FFP2: filtering face piece.

[b]CVC: central venous catheter.

[c]NCI-CTC: National Cancer Institute-Common Terminology Criteria for Adverse Events (formerly Common Toxicity Criteria).

Excluded Studies

We excluded 20 out of the short-listed 26 studies [48-67] as depicted in Multimedia Appendix 5. The studies were excluded due to the following reasons:

- Inappropriate study design: 5 nonrandomized comparative studies [50,53,55,58,64], 3 cohort studies [54,59,67], 3 case-control studies [51,52,56], 2 cross-sectional studies [62,65], 1 before-and-after study [63], 1 crossover study [49], and 1 mixed method study [61].
- Noninterest population: 2 studies [60,66].
- Nonrelevant intervention: 1 study [57].
- Nonrelevant outcome: 1 study [48].

We did not identify any on-going study for this review.

Risk of Bias of the Included Studies

Overall, the risk-of-bias profile of the included studies varied, with insufficient information in most studies to enable a meaningful assessment of the risk-of-selection bias, and the high risk of performance bias in half of the included studies was due to a lack of blinding of individuals and personnel. The distribution of risk of bias in different aspects for the included studies are shown in Figures 2 and 3 and Multimedia Appendix 3.

Effects of Interventions

Overview

All data presented were extracted from published reports. There were 8 comparisons included in the analysis, each only represented by a single study. The summarized analysis for the comparisons is tabulated in Multimedia Appendix 6. In the following sections, findings of the analysis are reported according to the comparison.

Comparison 1: Chlorhexidine and Nystatin Versus Saline Mouth Rinse

A single study was included [43], with 52 patients analyzed under this comparison out of the total of 99 patients from all 4 arms.

Only chemotherapy-related oral mucosal adverse effects were assessed in the study included [43]. However, there was no clear difference in mucositis score (grade 0-3, with higher scores indicating a more severe mucositis) between patients who received the chlorhexidine-nystatin mouth rinse and those who used the saline mouth rinse (MD 0.96, 95% CI –0.09 to 2.01; number of patients=52; analysis 1.1: quality of evidence was very low for both findings, reduced 1 level on the basis of risk of bias and another 2 levels due to severe serious concerns on imprecision). There was no clear difference in the average oral mucosal ulcer size between patients who received the chlorhexidine-nystatin mouth rinse and those who received the saline mouth rinse (MD 1.65 mm, 95% CI –7.48 to 10.78; number of patients=52; analysis 1.2: quality of evidence was very low for both outcomes, downgraded 1 level on the basis of risk of bias and another 2 levels due to severe serious concerns on imprecision). See also Multimedia Appendix 6 for the analysis.

Comparison 2: Chlorhexidine Versus Saline Mouth Rinse

A single study [43] was included, with 36 patients analyzed under this comparison out of the total of 99 patients from all 4 arms.

Only chemotherapy-related oral mucosal adverse effects were assessed in the included study [43]. However, there was no clear difference in mucositis score (grade 0-3, with higher scores indicating a more severe mucositis) between patients who received the chlorhexidine mouth rinse and patients who received the saline mouth rinse (MD 0.56, 95% CI –0.59 to 1.71; number of patients=36; analysis 2.1: quality of evidence was very low for both findings, reduced 1 level on the basis of risk of bias and another 2 levels due to severe serious concerns on imprecision). There was no clear difference in the oral mucosal ulcer size among patients who received the chlorhexidine mouth rinse and patients who received the saline mouth rinse (MD 2.17 mm, 95% CI –8.17 to 12.51; number of patients=36; analysis 2.2: quality of evidence was very low for both outcomes, reduced 1 level on the basis of risk of bias and another 2 levels due to severe serious concerns on imprecision).

Comparison 3: Nystatin Versus Saline Mouth Rinse

A single study [43] was included, with 34 patients analyzed under this comparison out of the total of 99 patients from all 4 arms.

Only chemotherapy-related oral mucosal adverse effects were assessed in the included study [43]. However, there was no clear difference in mucositis score (grade 0-3, with higher scores indicating a more severe mucositis) between patients who received the nystatin mouth rinse and patients who received the saline mouth rinse (MD 0.90, 95% CI –0.23 to 2.03; number of patients=34; analysis 3.1: quality of evidence was low, decreased 1 level on the basis of risk of bias and another 2 levels due to severe serious concerns on imprecision).

Comparison 4: Chlorhexidine Silver Sulfadiazine–Coated CVCs Versus Uncoated Catheters

A single study [42] of 184 catheters was included under this comparison. The study did not evaluate the risk of OI, but assessed CVC colonization, catheter-related bloodstream infection, and insertion site infection, which we classified for the purpose of this review as secondary findings. For catheter colonization the evidence showed that patients with cancer who received chlorhexidine silver sulfadiazine–coated CVCs appeared less likely to develop catheter colonization (RR 0.37, 95% CI 0.20 to 0.69; number of catheters=184; analysis 4.1: quality of evidence was moderate, downgraded 1 level based on indirectness of the outcome assessed).

For the association between catheters and bloodstream infections, the evidence showed no significant difference for the rate of catheter-related bloodstream infection between patients with cancer who received chlorhexidine silver sulfadiazine–coated CVCs and those who received standard, uncoated catheters (RR 0.45, 95% CI 0.12 to 1.68; number of catheters=184; analysis 4.2: quality of evidence was low, which was decreased 2 levels due to severe serious concerns on imprecision). For insertion site infection, there was no clear difference observed for the rate of insertion site infection between patients with cancer who received chlorhexidine silver sulfadiazine–coated CVCs and those who received standard, uncoated catheters (RR 0.94, 95% CI 0.66 to 1.33; number of catheters=184; analysis 4.3: quality of evidence observed was moderate, reduced 1 level because of serious concerns on imprecision).

Comparison 5: Well-Fitting Masks Versus No Mask

A single study [45] of 80 patients was included under this comparison. The study evaluated OI and all-cause mortality as the primary outcomes and mortality caused by OI as the secondary outcome.

In the included study [45], aspergillosis infection was assessed. The outcome was divided into possible, probable, or proven aspergillosis.

- For possible OIs, there was certainly no clear difference among patients who used a well-fitted mask versus those without a well-fitted mask (RR 0.48, 95% CI 0.09 to 2.45; number of patients=80; analysis 5.1.1: quality of evidence was very low, which was decreased by 3 levels due to indirectness and severe concerns on imprecision).
- For probable OIs, there was no clear difference among patients who used a well-fitted mask versus those without a well-fitted mask (RR 1.90, 95% CI 0.37 to 9.81; number of patients=80; analysis 5.1.2: quality of evidence was very low, which was reduced by 3 levels due to indirectness and serious concerns on imprecision).
- For proven OIs, there was no significant difference among patients who received a well-fitted mask versus those without a well-fitted mask (RR 0.95, 95% CI 0.14 to 6.43; number of patients=80; analysis 5.1.3: quality of evidence was very low, which was decreased by 3 levels due to indirectness and severe serious concerns on imprecision).
- For combined possible, probable, and proven OIs, there was no clear difference among patients who received a well-fitted mask versus those without a well-fitted mask (RR 0.95; 95% CI 0.40 to 2.29; number of patients = 80; analysis 5.1.4: very low quality of evidence, which was reduced by 3 levels for indirectness and severe concerns on imprecision).

The all-cause mortality provided by a single study [45], which was assessed clinically, showed no significant difference among patients with a well-fitted mask versus those without a well-fitted mask (RR 1.00, 95% CI 0.14 to 6.93; number of patients=160; analysis 5.2: reduced by 2 levels with low-quality evidence obtained for indirectness and severe concerns on imprecision). This study also provided evidence on mortality due to OI assessed clinically, in which no clear difference was observed for patients with and without well-fitted masks (RR 1.00, 95% CI 0.06 to 15.71; number of patients=160; analysis 5.3: quality of evidence was low, which was reduced by 2 levels due to indirectness and serious concerns on imprecision).

Comparison 6: Amine Fluoride-Stannous Fluoride Versus Sodium Fluoride Mouthwash

All-cause mortality provided by a single study [44], which was assessed clinically, showed no significant difference among patients who used the amine fluoride-stannous mouthwash versus those who used the sodium fluoride mouthwash (RR 0.67, 95% CI 0.11 to 3.88; number of patients=152; analysis 6.1: quality of evidence was low, reduced by 3 levels for indirectness and severe concerns on imprecision). This study also clinically assessed the combined adverse effect of stinging, discomfort in the mouth, teeth staining, nausea, and bad taste for patients who used the amine fluoride-stannous fluoride and sodium fluoride mouthwash, which showed a higher incidence of adverse effects for patients who used the latter (RR 9.33, 95% CI 1.34 to 64.89; number of patients=45; analysis 6.2.1: the very low-quality evidence reduced by 3 levels due to indirectness and serious concerns on imprecision).

Comparison 7: Low-Bacterial Diet Versus Normal Diet

One study [46] showed no clear difference in the impact of diet consumed (low bacterial vs normal) on OI (RR 0.2, 95% CI 0.01 to 3.70; number of patients=20; analysis 7.1.1: evidence with a very low quality was observed, which was reduced by 3 levels for indirectness and severe concerns on imprecision). For the OI (candidemia) assessed clinically and by laboratory reports, there was no clear difference among patients who consumed a low-bacterial diet versus those who consumed a normal diet (RR 1.00, 95% CI 0.07 to 13.87; number of patients=20; analysis 7.2: quality of evidence was very low, downgraded with 3 levels due to indirectness and serious concerns on imprecision).

Comparison 8: Herbal Mouthwash Versus Placebo Mouthwash

A single study [47] with 60 patients was analyzed under this comparison. The study evaluated oral mucositis. There was no significant difference between patients who received herbal and placebo mouthwash (RR 0.81, 95% CI 0.64 to 1.04; number of patients=60; analysis 8.1: evidence quality of evidence was moderate, downgraded by 1 level due to serious concerns on risk of incomplete outcome data bias).

Discussion

Summary of the Principal Findings

A total of 519 participants were evaluated from 6 included studies. Although this review included a small number of studies, it represented the best existing evidence that addressed the use of nondrug intervention for OIs. This review identified 4b major types of nondrug interventions for hematological cancer: (1) mouthwash that contained either chlorhexidine, nystatin, saline, amine fluoride-stannous fluoride, sodium fluoride, or herbal substances; (2) CVCs that were coated with chlorhexidine silver-sulfadiazine or uncoated; (3) use of well-fitted masks; and (4) diet consumed (either a low-bacterial diet or a normal diet). However, each type of intervention was represented by 1 small study. Overall, the use of chlorhexidine mouthwash alone or in combination with nystatin or nystatin mouthwash alone showed no clear difference for reductions in mucositis based on mucositis score and ulcer size [43]. A study by Laine et al [44] assessed the use of amine fluoride-stannous fluoride or sodium fluoride mouthwash. The study, however, did not show any clear difference in all-cause mortality and adverse effects such as discomfort in the mouth, teeth staining, unpleasant taste, and nausea. Maschmeyer et al [45] assessed the use of a well-fitted mask with no clear difference in OIs for all-cause mortality. Another study by Tiel et al [46] evaluated the use of a low bacteria diet versus a standard diet with no clear difference in the reduction of possible OIs or OIs assessed clinically. A study by Ardakani et al [47] that assessed the use of herbal mouthwash in preventing OIs showed no clear difference when compared with the placebo mouthwash group.

Comparison With Prior Reviews or Studies

We found 2 published reviews that examined the effectiveness of nondrug interventions in preventing OIs among different populations. Helder et al [68] had assessed the effectiveness of 5 different nondrug interventions in preventing bloodstream infections among newborns admitted to a neonatal intensive care unit. The authors included 15 RCTs and found that proper CVC insertion and maintenance with a proper aseptic technique were the most effective interventions to prevent bloodstream infection in infants. The review differed from our review, as ours focused on evaluating the outcomes among adults with hematological malignancies. The other review by Wekesah et al [69] assessed the effectiveness of nondrug interventions in improving outcomes and quality of care among pregnant women in sub-Saharan Africa. The authors included 73 mixed design studies and identified many interventions for improving maternal health. The review differed in scope from our current review, as we only focused on the prevention of OIs among individuals with hematological cancers.

To the best of our knowledge, there is no systematic review that evaluated the use of nondrug interventions in preventing OIs among individuals with hematological cancer. The only available reviews that assessed interventions relevant to our review are those that evaluated antimicrobial-coated CVCs. One Cochrane systematic review and a related meta-analysis assessed the safety and effectiveness of antimicrobial catheters for patients in the intensive care unit, hematological and oncology unit (with all types of malignancies), and community settings [44,45], and reported that antimicrobial catheters in general reduced catheter colonization without clearly reducing catheter-related bloodstream infections, overall sepsis, and mortality rates. This is consistent with the finding of our single included study [42], which was also included in both reviews.

Strengths

The review also has strengths. First, this is the first systematic review that evaluated the use of nondrug interventions for preventing OIs among patients with hematological cancers. We confirmed the effectiveness through the synthesis of evidence, and the result has significant clinical implication for both oncologists and patients. Second, we established strict inclusion and exclusion criteria, which resulted in a uniform data to be evaluated. We also included our assessment on study quality, which allows readers to judge the strength of evidence.

Limitations

The review has a few limitations. First, despite our comprehensive search, we could find only very few studies related to OIs that are eligible to be included. This review also only focused on the effect of nondrug interventions for hematological cancers that limited the applicability of its results to other types of cancer or immunocompromised patients. Further, we might have missed relevant papers from smaller databases, especially those that are non-English. In addition, there might be publication bias that we were unable to rigorously evaluate as the number of included studies was too small. Besides, we only included RCTs in our review with a well-defined, relatively narrow set of patient population, and as such, serious or rare adverse events might not have been comprehensively captured.

Conclusions

Overall, the quality of evidence presented in this review was very low. We are uncertain on the efficacy and safety of various types of mouthwash, coated CVCs, use of well-fitted masks, and low bacterial diet in major clinical findings such as OIs and related outcomes. Insufficient evidence exists on the effect of the nondrug intervention for preventing OIs in people with hematological cancers, and in people who are immunocompromised. This lack of evidence should be kept in mind when balancing the beneficial effects of nondrug interventions against the cost and feasibility of implementation in specific settings and against the potential for the development of OIs, and thus no firm conclusion can be made to inform practice. Therefore, further evidence is needed regarding the effect of nondrug interventions in patients with cancers, or in those who are immunocompromised.

Acknowledgments

We thank the Director General of Health, Malaysia for permission to publish this review. We extend our appreciation to Professor Dr Jacqueline Ho from the Royal College of Surgeons in Ireland and University College Dublin (RCSI & UCD) Malaysia Campus (formerly Penang Medical College Malaysia) for her technical advice. This systematic review did not accept any specific grant from any funding agencies, or public, commercial, or not-for-profit sectors. .

Data Availability

The original contributions shown in the study are included in this systematic review or its multimedia appendices; further inquiries can be addressed to the corresponding author.

Authors' Contributions

NAM contributed to developing the project plan as well as the protocol, data analysis, and review writing. NHM aided in developing the project plan and rating the scheme evidence. NM led in entering the data, data analysis, and writing the full review. FNL and IAR supported in developing both inclusion and exclusion criteria as well as the measurement for the findings. NML contributed to writing the taxonomy review for the interventions as well as approving the final draft. TA contributed to writing the review and approved the final draft. MU analyzed the data and wrote the manuscript draft.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy for MEDLINE.
[DOCX File , 14 KB - ijmr_v12i1e43969_app1.docx]

Multimedia Appendix 2
Search strategy for CENTRAL (Cochrane Central Register of Controlled Trials).
[DOCX File , 13 KB - ijmr_v12i1e43969_app2.docx]

Multimedia Appendix 3
Risk of bias for 6 included studies.
[DOCX File , 36 KB - ijmr_v12i1e43969_app3.docx]

Multimedia Appendix 4
Summary of findings (SOF) tables.
[DOCX File , 60 KB - ijmr_v12i1e43969_app4.docx]

Multimedia Appendix 5
Characteristics of the 20 excluded studies and their basis for exclusion.
[DOCX File , 33 KB - ijmr_v12i1e43969_app5.docx]

Multimedia Appendix 6
Summary table of analysis for the 8 comparisons.
[DOCX File , 36 KB - ijmr_v12i1e43969_app6.docx]

References

1. Hoffbrand AV, Higgs DR, Keeling DM, Mehta AB, editors. WHO Classification: Tumours of the Haematopoietic and Lymphoid Tissues (2008). In: Postgraduate Haematology, Seventh Edition. Chichester, United Kingdom: John Wiley & Sons Ltd; Dec 2015:885-887.
2. Leukaemia incidence statistics. Cancer Research UK. 2014. URL: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/leukaemia/incidence [accessed 2022-04-05]
3. Centers for Disease Control and Prevention (CDC). Cancer: National Program of Cancer Registries. Division of Cancer Prevention and Control. Washington, DC: U.S. Department of Health & Human Services; 2010. URL: http://apps.nccd.cdc.gov/uscs [accessed 2022-04-06]
4. Chamilos G, Luna M, Lewis R, Bodey G, Chemaly R, Tarrand J, et al. Invasive fungal infections in patients with hematologic malignancies in a tertiary care cancer center: an autopsy study over a 15-year period (1989-2003). Haematologica 2006 Jul;91(7):986-989. [Medline: 16757415]
5. Martino R, Subirà M. Invasive fungal infections in hematology: new trends. Ann Hematol 2002 May;81(5):233-243. [doi: 10.1007/s00277-002-0466-3] [Medline: 12029531]
6. Nørgaard M. Risk of Infections in Adult Patients with Haematological Malignancies. TOIDJ 2012 Oct 02;6(1):46-51. [doi: 10.2174/1874279301206010046]
7. Styczynski J, Reusser P, Einsele H, de la Camara R, Cordonnier C, Ward KN, et al. Management of HSV, VZV and EBV infections in patients with hematological malignancies and after SCT: guidelines from the Second European Conference on Infections in Leukemia. In: Bone Marrow Transplant. 2009 May 01 Presented at: Second European Conference on Infections in Leukemia; Sept 2007; Juan-les-Pins, France p. 757-770. [doi: 10.1038/bmt.2008.386]
8. Flowers CR, Seidenfeld J, Bow EJ, Karten C, Gleason C, Hawley DK, et al. Antimicrobial prophylaxis and outpatient management of fever and neutropenia in adults treated for malignancy: American Society of Clinical Oncology clinical practice guideline. J Clin Oncol 2013 Feb 20;31(6):794-810. [doi: 10.1200/JCO.2012.45.8661] [Medline: 23319691]
9. Fife BL, Monahan PO, Abonour R, Wood LL, Stump TE. Adaptation of family caregivers during the acute phase of adult BMT. Bone Marrow Transplant 2009 Jun;43(12):959-966. [doi: 10.1038/bmt.2008.405] [Medline: 19104493]
10. Grulke N, Bailer H, Hertenstein B, Kächele H, Arnold R, Tschuschke V, et al. Coping and survival in patients with leukemia undergoing allogeneic bone marrow transplantation--long-term follow-up of a prospective study. J Psychosom Res 2005 Nov;59(5):337-346. [doi: 10.1016/j.jpsychores.2005.02.006] [Medline: 16253625]
11. Sasaki T, Akaho R, Sakamaki H, Akiyama H, Yoshino M, Hagiya K, et al. Mental disturbances during isolation in bone marrow transplant patients with leukemia. Bone Marrow Transplant 2000 Feb;25(3):315-318. [doi: 10.1038/sj.bmt.1702117] [Medline: 10673704]
12. Grulke N, Larbig W, Kächele H, Bailer H. Pre-transplant depression as risk factor for survival of patients undergoing allogeneic haematopoietic stem cell transplantation. Psychooncology 2008 May;17(5):480-487. [doi: 10.1002/pon.1261] [Medline: 17879971]
13. Hoodin F, Kalbfleisch KR, Thornton J, Ratanatharathorn V. Psychosocial influences on 305 adults' survival after bone marrow transplantation; depression, smoking, and behavioral self-regulation. J Psychosom Res 2004 Aug;57(2):145-154. [doi: 10.1016/S0022-3999(03)00599-3] [Medline: 15465068]
14. Pulgar, Garrido S, Alcalá A, Reyes del Paso GA. Psychosocial predictors of immune response following bone marrow transplantation. Behav Med 2012 Jan;38(1):12-18. [doi: 10.1080/08964289.2011.647118] [Medline: 22356598]
15. Martin LR, Williams SL, Haskard KB, Dimatteo MR. The challenge of patient adherence. Ther Clin Risk Manag 2005 Sep;1(3):189-199 [FREE Full text] [Medline: 18360559]
16. Oliveira Pinto SM, Alves Caldeira Berenguer SM, Martins JCA. Cancer, Health Literacy, and Happiness: Perspectives from Patients under Chemotherapy. Nurs Res Pract 2013;2013:291767 [FREE Full text] [doi: 10.1155/2013/291767] [Medline: 24089635]
17. Nirenberg A, Parry Bush A, Davis A, Friese CR, Wicklin Gillespie T, Rice RD. Neutropenia: state of the knowledge part II. Oncol Nurs Forum 2006 Nov 27;33(6):1202-1208. [doi: 10.1188/06.ONF.1202-1208] [Medline: 17149403]
18. Larson E. Hygiene of the skin: when is clean too clean? Emerg Infect Dis 2001;7(2):225-230 [FREE Full text] [doi: 10.3201/eid0702.010215] [Medline: 11294712]
19. Carter L. Bacterial translocation: nursing implications in the care of patients with neutropenia. Oncol Nurs Forum 1994 Jun;21(5):857-65; quiz 866. [Medline: 7937247]

20. DeMille D, Deming P, Lupinacci P, Jacobs LA. The effect of the neutropenic diet in the outpatient setting: a pilot study. Oncol Nurs Forum 2006 Nov 03;33(2):337-343. [doi: 10.1188/ONF.06.337-343] [Medline: 16518449]
21. Wilson BJ. Dietary recommendations for neutropenic patients. Semin Oncol Nurs 2002 Feb;18(1):44-49. [doi: 10.1053/sonu.2002.30044] [Medline: 11878039]
22. Duncan SL. APIC State-of-the-Art Report: the implications of service animals in health care settings. Am J Infect Control 2000 Apr;28(2):170-180. [Medline: 10760225]
23. Eckmanns T, Bessert J, Behnke M, Gastmeier P, Ruden H. Compliance with antiseptic hand rub use in intensive care units: the Hawthorne effect. Infect Control Hosp Epidemiol 2006 Sep;27(9):931-934. [doi: 10.1086/507294] [Medline: 16941318]
24. Garcia R, Raad I, Abi-Said D, Bodey G, Champlin R, Tarrand J, et al. Nosocomial respiratory syncytial virus infections: prevention and control in bone marrow transplant patients. Infect Control Hosp Epidemiol 1997 Jun;18(6):412-416. [doi: 10.1086/647640] [Medline: 9181397]
25. Kates S, McGinley K, Larson E, Leyden J. Indigenous multiresistant bacteria from flowers in hospital and nonhospital environments. Am J Infect Control 1991 Jun;19(3):156-161. [doi: 10.1016/0196-6553(91)90022-5] [Medline: 1907439]
26. Sehulster L, Chinn R, CDC, HICPAC. Guidelines for environmental infection control in health-care facilities. Recommendations of CDC and the Healthcare Infection Control Practices Advisory Committee (HICPAC). MMWR Recomm Rep 2003 Jun 06;52(RR-10):1-42 [FREE Full text] [Medline: 12836624]
27. Lam OL, McMillan AS, Samaranayake LP, Li LS, McGrath C. Effect of oral hygiene interventions on opportunistic pathogens in patients after stroke. Am J Infect Control 2013 Feb;41(2):149-154. [doi: 10.1016/j.ajic.2012.02.020] [Medline: 22818804]
28. Powers J, Peed J, Burns L, Ziemba-Davis M. Chlorhexidine bathing and microbial contamination in patients' bath basins. Am J Crit Care 2012 Sep;21(5):338-342. [doi: 10.4037/ajcc2012242] [Medline: 22941707]
29. O'Brien S, Blijlevens N, Mahfouz T, Anaissie E. Infections in patients with hematological cancer: recent developments. Hematology Am Soc Hematol Educ Program 2003:438-472. [doi: 10.1182/asheducation-2003.1.438] [Medline: 14633794]
30. Center for International BloodMarrow Transplant Research (CIBMTR), National Marrow Donor Program (NMDP), European BloodMarrow Transplant Group (EBMT), American Society of BloodMarrow Transplantation (ASBMT), Canadian BloodMarrow Transplant Group (CBMTG), Infectious Disease Society of America (IDSA), Society for Healthcare Epidemiology of America (SHEA), Association of Medical MicrobiologyInfectious Diseases Canada (AMMI), Centers for Disease Control Prevention (CDC). Guidelines for preventing infectious complications among hematopoietic cell transplant recipients: a global perspective. Bone Marrow Transplant 2009 Oct;44(8):453-558. [Medline: 20095071]
31. Johnson D, Druce J, Birch C, Grayson M. A quantitative assessment of the efficacy of surgical and N95 masks to filter influenza virus in patients with acute influenza infection. Clin Infect Dis 2009 Jul 15;49(2):275-277. [doi: 10.1086/600041] [Medline: 19522650]
32. Seto W, Tsang D, Yung R, Ching T, Ng T, Ho M, Advisors of Expert SARS group of Hospital Authority. Effectiveness of precautions against droplets and contact in prevention of nosocomial transmission of severe acute respiratory syndrome (SARS). Lancet 2003 May 03;361(9368):1519-1520 [FREE Full text] [doi: 10.1016/s0140-6736(03)13168-6] [Medline: 12737864]
33. Boyce JM, Pittet D, Healthcare Infection Control Practices Advisory Committee, HICPAC/SHEA/APIC/IDSA Hand Hygiene Task Force. Guideline for Hand Hygiene in Health-Care Settings. Recommendations of the Healthcare Infection Control Practices Advisory Committee and the HICPAC/SHEA/APIC/IDSA Hand Hygiene Task Force. Society for Healthcare Epidemiology of America/Association for Professionals in Infection Control/Infectious Diseases Society of America. MMWR Recomm Rep 2002 Oct 25;51(RR-16):1-45, quiz CE1 [FREE Full text] [Medline: 12418624]
34. Pittet D, Hugonnet S, Harbarth S, Mourouga P, Sauvan V, Touveneau S, et al. Effectiveness of a hospital-wide programme to improve compliance with hand hygiene. Infection Control Programme. Lancet 2000 Oct 14;356(9238):1307-1312. [doi: 10.1016/s0140-6736(00)02814-2] [Medline: 11073019]
35. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg 2010;8(5):336-341 [FREE Full text] [doi: 10.1016/j.ijsu.2010.02.007] [Medline: 20171303]
36. Page M, McKenzie J, Bossuyt P, Boutron I, Hoffmann T, Mulrow C, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
37. ISRCTN Registry. URL: https://www.isrctn.com/search?q= [accessed 2022-03-17]
38. International Clinical Trials Registry Platform (ICTRP). URL: https://trialsearch.who.int/ [accessed 2022-03-17]
39. ClinicalTrials.gov. URL: https://clinicaltrials.gov/ [accessed 2022-03-17]
40. Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. Chichester, UK: John Wiley & Sons, Ltd; Mar 2011.
41. Review Manager (RevMan) version 5.4. London, UK: The Cochrane Collaboration; 2020. URL: https://training.cochrane.org/online-learning/core-software/revman [accessed 2022-11-01]

42. Ostendorf T, Meinhold A, Harter C, Salwender H, Egerer G, Geiss HK, et al. Chlorhexidine and silver-sulfadiazine coated central venous catheters in haematological patients--a double-blind, randomised, prospective, controlled trial. Support Care Cancer 2005 Dec;13(12):993-1000. [doi: 10.1007/s00520-005-0812-9] [Medline: 15834740]
43. Epstein JB, Vickars L, Spinelli J, Reece D. Efficacy of chlorhexidine and nystatin rinses in prevention of oral complications in leukemia and bone marrow transplantation. Oral Surg Oral Med Oral Pathol 1992 Jun;73(6):682-689. [doi: 10.1016/0030-4220(92)90009-f] [Medline: 1437036]
44. Laine P, Meurman JH, Murtomaa H, Lindqvist C, Torkko H, Pyrhönen S, et al. One-year trial of the effect of rinsing with an amine fluoride-stannous-fluoride-containing mouthwash on gingival index scores and salivary microbial counts in lymphoma patients receiving cytostatic drugs. J Clin Periodontol 1993 Oct;20(9):628-634. [doi: 10.1111/j.1600-051x.1993.tb00707.x] [Medline: 8227449]
45. Maschmeyer G, Neuburger S, Fritz L, Böhme A, Penack O, Schwerdtfeger R, Infectious Diseases Working Party (AGIHO) of the German Society of HaematologyOncology. A prospective, randomised study on the use of well-fitting masks for prevention of invasive aspergillosis in high-risk patients. Ann Oncol 2009 Sep;20(9):1560-1564 [FREE Full text] [doi: 10.1093/annonc/mdp034] [Medline: 19451183]
46. van Tiel F, Harbers M, Terporten P, van Boxtel R, Kessels A, Voss G, et al. Normal hospital and low-bacterial diet in patients with cytopenia after intensive chemotherapy for hematological malignancy: a study of safety. Ann Oncol 2007 Jun;18(6):1080-1084 [FREE Full text] [doi: 10.1093/annonc/mdm082] [Medline: 17369599]
47. Tavakoli Ardakani M, Ghassemi S, Mehdizadeh M, Mojab F, Salamzadeh J, Ghassemi S, et al. Evaluating the effect of Matricaria recutita and Mentha piperita herbal mouthwash on management of oral mucositis in patients undergoing hematopoietic stem cell transplantation: A randomized, double blind, placebo controlled clinical trial. Complement Ther Med 2016 Dec;29:29-34. [doi: 10.1016/j.ctim.2016.08.001] [Medline: 27912953]
48. Bauer K, Kemen M, Senkal M, Wolfram G, Zumtobel V. Serum and tissue phospholipid fatty acids and leukotriene release from leukocytes after preoperative oral supplementation with an immunosupportive diet in cancer patients. Clinical Nutrition 1994 Jan;13:44. [doi: 10.1016/0261-5614(94)90244-5]
49. Bright-See E, McKeown-Eyssen G, Jacobson EA, Newmark HW, Mathews R, Morson L, et al. Dietary fiber and cancer: a supplement for intervention studies. Nutr Cancer 1985;7(4):211-220. [doi: 10.1080/01635588509513857] [Medline: 3010248]
50. De Beule F, Bercy P, Ferrant A. The effectiveness of a preventive regimen on the periodontal health of patients undergoing chemotherapy for leukemia and lymphoma. J Clin Periodontol 1991 May;18(5):346-347. [doi: 10.1111/j.1600-051x.1991.tb00440.x] [Medline: 2066451]
51. Enig B, Petersen HN, Smith DF, Larsen B. Food preferences, nutrient intake and nutritional status in cancer patients. Acta Oncol 1987;26(4):301-305. [doi: 10.3109/02841868709089979] [Medline: 3689583]
52. Franceschi S, Serraino D, Carbone A, Talamini R, La Vecchia C. Dietary factors and non-Hodgkin's lymphoma: a case-control study in the northeastern part of Italy. Nutr Cancer 1989;12(4):333-341. [doi: 10.1080/01635588909514034] [Medline: 2608538]
53. Gardner A, Mattiuzzi G, Faderl S, Borthakur G, Garcia-Manero G, Pierce S, et al. Randomized comparison of cooked and noncooked diets in patients undergoing remission induction therapy for acute myeloid leukemia. J Clin Oncol 2008 Dec 10;26(35):5684-5688 [FREE Full text] [doi: 10.1200/JCO.2008.16.4681] [Medline: 18955453]
54. Hulshof KF, Gooskens AC, Wedel M, Bruning PF. Food intake in three groups of cancer patients. A prospective study during cancer treatment. Hum Nutr Appl Nutr 1987 Feb;41(1):23-37. [Medline: 3558006]
55. Jehn U, Ruckdeschel G, Sauer H, Clemm C, Wilmanns. [Comparative study on the value of selective decontamination of the digestive tract in acute leukemia patients (author's transl)]. Klin Wochenschr 1981 Oct 01;59(19):1093-1099. [doi: 10.1007/BF01746196] [Medline: 6808225]
56. Kwiatkowski A. [Dietetic factors characterizing food habits of patients with acute leukemia]. Acta Haematol Pol 1990;21(2):207-218. [Medline: 2131720]
57. Levine A, Robinson RA, Hauser JM. Analysis of studies on protected environments and prophylactic antibiotics in adult acute leukemia. Eur J Cancer (1965) 1975 Aug;11suppl:57-66. [doi: 10.1016/b978-0-08-019964-1.50011-7] [Medline: 815093]
58. Levine A, Brennan M, Ramu A, Fisher R, Pizzo P, Glaubiger D. Controlled clinical trials of nutritional intervention as an adjunct to chemotherapy, with a comment on nutrition and drug resistance. Cancer Res 1982;42(2 Suppl):774s-781s. [Medline: 6799196]
59. Lindman A, Rasmussen HB, Andersen NF. Food caregivers influence on nutritional intake among admitted haematological cancer patients - a prospective study. Eur J Oncol Nurs 2013 Dec;17(6):827-834. [doi: 10.1016/j.ejon.2013.06.010] [Medline: 24012191]
60. McGaw W, Belch A. Oral complications of acute leukemia: prophylactic impact of a chlorhexidine mouth rinse regimen. Oral Surg Oral Med Oral Pathol 1985 Sep;60(3):275-280. [doi: 10.1016/0030-4220(85)90311-1] [Medline: 3862040]
61. McGrath P. Reflections on nutritional issues associated with cancer therapy. Cancer Pract 2002;10(2):94-101. [doi: 10.1046/j.1523-5394.2002.102007.x] [Medline: 11903274]

62. Oren I, Haddad N, Finkelstein R, Rowe JM. Invasive pulmonary aspergillosis in neutropenic patients during hospital construction: before and after chemoprophylaxis and institution of HEPA filters. Am J Hematol 2001 Apr;66(4):257-262 [FREE Full text] [doi: 10.1002/ajh.1054] [Medline: 11279636]
63. Rico H, Hernandez E, Diaz-Mediavilla J, Alvarez A, Martinez R, Espinos D. Treatment of multiple myeloma with nasal spray calcitonin: a histomorphometric and biochemical study. Bone Miner 1990 Mar;8(3):231-237. [doi: 10.1016/0169-6009(90)90108-r] [Medline: 2322695]
64. Sugahara H, Mizuki M, Matsumae S, Nabetani Y, Kikuchi M, Kanakura Y. Footwear exchange has no influence on the incidence of febrile neutropenia in patients undergoing chemotherapy for hematologic malignancies. Infect Control Hosp Epidemiol 2004 Jan;25(1):51-54. [doi: 10.1086/502292] [Medline: 14756220]
65. Trifilio S, Helenowski I, Giel M, Gobel B, Pi J, Greenberg D, et al. Questioning the role of a neutropenic diet following hematopoetic stem cell transplantation. Biol Blood Marrow Transplant 2012 Sep;18(9):1385-1390 [FREE Full text] [doi: 10.1016/j.bbmt.2012.02.015] [Medline: 22430084]
66. Wahlin YB. Effects of chlorhexidine mouthrinse on oral health in patients with acute leukemia. Oral Surg Oral Med Oral Pathol 1989 Sep;68(3):279-287. [doi: 10.1016/0030-4220(89)90212-0] [Medline: 2771374]
67. Withington S, Chambers S, Beard M, Inder A, Allen J, Ikram R, et al. Invasive aspergillosis in severely neutropenic patients over 18 years: impact of intranasal amphotericin B and HEPA filtration. J Hosp Infect 1998 Jan;38(1):11-18. [doi: 10.1016/s0195-6701(98)90170-0] [Medline: 9513064]
68. Helder O, van den Hoogen A, de Boer C, van Goudoever J, Verboon-Maciolek M, Kornelisse R. Effectiveness of non-pharmacological interventions for the prevention of bloodstream infections in infants admitted to a neonatal intensive care unit: A systematic review. Int J Nurs Stud 2013 Jun;50(6):819-831. [doi: 10.1016/j.ijnurstu.2012.02.009] [Medline: 22385913]
69. Wekesah FM, Mbada CE, Muula AS, Kabiru CW, Muthuri SK, Izugbara CO. Effective non-drug interventions for improving outcomes and quality of maternal health care in sub-Saharan Africa: a systematic review. Syst Rev 2016 Aug 15;5(1):137 [FREE Full text] [doi: 10.1186/s13643-016-0305-6] [Medline: 27526773]

Abbreviations

CENTRAL: Cochrane Central Register of Controlled Trials
CVC: central venous catheter
EHA: European Hematology Association
HEPA: high-efficiency particulate absorption
ICTRP: International Clinical Trials Registry Platform
MD: mean difference
MeSH: Medical Subject Heading
OI: opportunistic infection
PPE: personal protective equipment
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
RR: risk ratio
SOF: summary of findings
VAHO: Virginia Association of Hematologists and Oncologists
WHO: World Health Organization

Edited by A Mavragani; submitted 01.11.22; peer-reviewed by A Mohd Yacob, V Gosselin Boucher, KA Nguyen; comments to author 14.12.22; revised version received 19.01.23; accepted 20.01.23; published 31.03.23.

Please cite as:
Muhamad NA, Ma'amor NH, Mustapha N, Leman FN, Rosli IA, Umar M, Aris T, Lai NM
Nondrug Intervention for Opportunistic Infections in Individuals With Hematological Malignancy: Systematic Review
Interact J Med Res 2023;12:e43969
URL: https://www.i-jmr.org/2023/1/e43969
doi:10.2196/43969
PMID:37000482

Review

Heart Rate Variability and Pregnancy Complications: Systematic Review

Zahra Sharifi-Heris[1*], BSN, MSN, PhD; Amir M Rahmani[1*], PhD; Anna Axelin[2*], PhD; Mahkameh Rasouli[1*], BSc; Miriam Bender[1*], PhD

[1]University of California Irvine, Irvine, CA, United States

[2]University of Turku, Turku, Finland

[*]all authors contributed equally

Corresponding Author:
Zahra Sharifi-Heris, BSN, MSN, PhD
University of California Irvine
8420 Palo Verde
Irvine, CA, 92697
United States
Phone: 1 6506805432
Email: sharifiz@uci.edu

Abstract

Background: The autonomic nervous system (ANS) is known as a critical regulatory system for pregnancy-induced adaptations. If it fails to function, life-threatening pregnancy complications could occur. Hence, understanding and monitoring the underlying mechanism of action for these complications are necessary.

Objective: We aimed to systematically review the literature concerned with the associations between heart rate variability (HRV), as an ANS biomarker, and pregnancy complications.

Methods: We performed a comprehensive search in the PubMed, Medline Completion, CINAHL Completion, Web of Science Core Collection Classic, Cochrane Library, and SCOPUS databases in February 2022 with no time span limitation. We included studies concerned with the association between any pregnancy complications and HRV, with or without a control group. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guideline was used for the review of the studies, and Covidence software was used for the study selection process. For data synthesis, we used the guideline by Popay et al.

Results: Finally, 12 studies with 6656 participants were included. Despite the methodological divergency that hindered a comprehensive comparison, our findings suggest that ANS is linked with some common pregnancy complications including fetal growth. However, existing studies do not support an association between ANS and gestational diabetes mellitus. Studies that linked pulmonary and central nervous system disorders with ANS function did not provide enough evidence to draw conclusions.

Conclusions: This review highlights the importance of understanding and monitoring the underlying mechanism of ANS in pregnancy-induced adaptations and the need for further research with robust methodology in this area.

(Interact J Med Res 2023;12:e44430) doi:10.2196/44430

KEYWORDS

autonomic nervous system; heart rate variability; pregnancy complication; pregnancy; maternal; hypertensive disorder; fetal growth; global developmental delay; hypertension

Introduction

Background

Various physiological changes occur during pregnancy that contribute to optimal growth and development of the fetus and help protect the mother from pregnancy and delivery complications [1]. These changes are regulated through a nonlinear complex relationship between various vital systems in the body. The autonomic nervous system (ANS) is known as a critical regulatory system for pregnancy-induced adaptations [2]. The importance of the ANS lies, to a great extent, in the fact that every organ of the human body is innervated and thus regulated by the ANS [3,4]. This systemic innervation enables the ANS to restore relationships between the individual functional systems after a disturbance of the balance of the human body with the help of certain adjusting reactions [5]. Sympathetic nervous system (SNS) and parasympathetic nervous system (PNS) components, as 2 functional elements of ANS, have complementary roles to mediate the hemodynamic adaptation in the body. The SNS directs the body's rapid involuntary response to various internal or external demands [6] and mediates the vigilance, arousal, and activation of the bodily responses to adapt to increased metabolic needs in response to internal and external stimuli including pregnancy [7]. The PNS, on the other hand, is responsible for stimulation of “rest-and-digest” or “feed-and-breed” activities that occur when the body is at rest [8].

During pregnancy, hemodynamic adjustment is one of the main adaptations regulated by the ANS. This adaptation includes adjusting blood pressure, blood volume, and heart activity. Systemic vasodilation is the primary initial hemodynamic event that begins during the luteal phase of the menstrual cycle following the release of chemical mediators by the corpus luteum and is further amplified by adjunct factors including placental hormones and the vasodilatory sex steroid estrogen that are present during pregnancy [9-11]. The outcome of systemic vasodilation is a series of systemic accommodations regulated by the ANS that adapt to the pregnancy demands known as “hemostasis” but is dynamic and complex in function [12]. One of the initial hemodynamic accommodations is an increased volume of circulating blood (plasma, red cells), resulting in 40% to 45% higher volume than prior to pregnancy, in response to systemic vasodilation [13]. Due to the increased blood volume and decreased vascular resistance, heart rate and cardiac output increase, but maternal blood pressure is not elevated. These hemodynamic adaptations provide the required blood for the fetal need for growth [13] through increased sympathetic tone and parasympathetic withdrawal.

Heart rate variability (HRV) is a well-known, noninvasive variable that has been commonly used in the recent literature to detect various physical, and psychological disorders resulting from ANS dysfunction [14]. HRV, by definition, is the variation in the beat-to-beat (RR or NN) interval and measures oscillations in the interval between consecutive heart beats and reflects the variability in the intervals between R waves [15]. By function, HRV is representative of interrelated regulatory systems that indicate the person’s adaptation to internal and external stressors. Optimal variability indicates the responsiveness of the ANS and sympathetic-parasympathetic components to deal with the stressor [15]. Variability in the heart rate indicates the flexibility to cope with the uncertain and changing environment through the cardiovascular system. HRV is a surrogate parameter of the ANS reflecting the complex interaction between organ systems, and specifically the brain and cardiovascular system, to maintain hemostasis [16].

HRV is interpreted by various mathematical computations on the interbeat interval. These include time domain (eg, standard deviation of the normal-to-normal R-R intervals [SDNN], root mean square of successive differences between normal heartbeats [RMSSD]), frequency domain (eg, low frequency [LF], high frequency [HF], LF/HF, very low frequency [VLF]), and nonlinear (eg, entropy) metrics for which the SNS, PNS, or both may contribute. The PNS mainly contributes to HF, VLF, and RMSSD, and both the SNS and PNS play a role in LF, LF/HF, and SDNN (see Multimedia Appendix 1) [15].

Literature that has used HRV as an ANS biomarker has suggested that ANS activity is shifted toward higher sympathetic and lower vagal modulation in response to pregnancy-induced demands over the course of a pregnancy [17]. However, methodological gaps such as noncontinuous assessment of HRV has hindered the understanding of where, when, and how these ANS alterations occur during pregnancy and whether these changes predict good or bad outcomes.

A failure in ANS regulation has been described in multiple and diverse diseases, both those that directly affect the nervous system and those affecting other organs, where they indirectly trigger or enhance pathological symptoms in the body [18]. It has been suggested that there is absolutely no disorder nor ailment in which the ANS plays no role [5]. Likewise, in pregnant individuals, ANS dysfunction has been considered one of the main contributors to the development of some maternal or neonatal disorders. Understanding the relationship between the ANS and pregnancy complications may be a pathway to investigate mechanisms of action for these life-threatening complications. This is particularly important as the growing availability of technology enables us to continuously and cost-effectively assess ANS function and the potential associated complications.

Objectives

In this study, we aimed to systematically review the potential pregnancy complications associated with ANS function and reflected in HRV.

Methods

Design

We performed a systematic review using the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) standards [19] to guide the study.

Eligibility Criteria

The PECO (Population, Exposure, Comparison or Controls, and Outcome) framework was applied to develop the research question and inclusion criteria. Regarding population (P), all pregnant individuals were included. The exposure (E) component, which is required for all the studies, included any pregnancy-related complications during pregnancy. For the comparison group (C), studies with or without a control group were eligible to be included. Regarding outcome (O), HRV, assessed at least once during pregnancy, was considered the expected outcome for all the studies. Studies were included if they were in the English language. The exclusion criteria included articles that were a systematic review, protocol, conference, letter to the editor, unpublished or under review, or dissertation proposal.

Information Sources

The PubMed, Medline Completion, CINAHL Completion, Web of Science Core Collection Classic, Cochrane Library, and SCOPUS databases were searched initially in February 2022 (with no time span limitation). Although we used no limitation for the time span during the search, the time span varied depending on the history of each database. See Multimedia Appendix 2 for more details. To access further studies, the reference lists of the reviewed articles and Google Scholar were checked.

Search Strategy

Key words including "heart rate variability" and "pregnancy complications" were used for both simple and advanced searches in each database separately (see Multimedia Appendix 2 for all the terms).

Selection Process

Selected articles were peer reviewed in the online Covidence software by 2 independent reviewers. To assess relevancy, all the studies were screened by both reviewers, ZS and MR, based on titles, abstracts, and full texts in 2 steps. In the first step, the abstracts of all the articles that were gathered from the databases were screened in terms of their relevance to our study aim. Next, articles with relevant titles or abstracts that resulted from the first step underwent a full-text assessment. To resolve any disagreement, a third reviewer, MB, was consulted.

Data Items

We assessed results for HRV, as the main outcome collected; assessment tools used for HRV; HRV component(s); frequency and duration of the assessment; and gestational age at the assessment.

Effect Measures

Effect measures for the studies were included if they were reported, and the significance of the difference varied from study to study.

Quality and Risk of Bias Assessment

Two independent reviewers, ZS and MR, assessed the methodological quality of the selected studies using the National Heart, Lung, and Blood Institute (NHLBI) quality assessment scale for observational cohort and cross-sectional studies [20]. The NHLBI quality assessment tool, using 14 questions, assesses studies in terms of criteria such as study objectives, study population, sample size, exposures and outcome measures, and key potential confounding variables. Each study was assessed for risk of bias using "yes," "no," and "cannot determine/not applicable/not reported" answers for every single criterion. See Figure 1 for more details.

Figure 1. National Heart, Lung, and Blood Institute (NHLBI) quality assessment of the included studies: Question 1: Was the research question or objective in this paper clearly stated? Question 2: Was the study population clearly specified and defined? Question 3: Was the participation rate of eligible persons at least 50%? Question 4: Were all the subjects selected or recruited from the same or similar populations (including the same time period)? Were inclusion and exclusion criteria for being in the study prespecified and applied uniformly to all participants? Question 5: Was a sample size justification, power description, or variance and effect estimates provided? Question 6: For the analyses in this paper, were the exposure(s) of interest measured prior to the outcome(s) being measured? Question 7: Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if it existed? Question 8: For exposures that can vary in amount or level, did the study examine different levels of the exposure as related to the outcome (eg, categories of exposure, or exposure measured as continuous variable)? Question 9: Were the exposure measures (independent variables) clearly defined, valid, reliable, and implemented consistently across all study participants? Question 9: Question 10: Was the exposure(s) assessed more than once over time? Question 11: Were the outcome measures (dependent variables) clearly defined, valid, reliable, and implemented consistently across all study participants? Question 12: Were the outcome assessors blinded to the exposure status of participants? Question 13: Was loss to follow-up after baseline 20% or less? Question 14: Were key potential confounding variables measured and adjusted statistically for their impact on the relationship between exposure(s) and outcome(s)? Green +: "yes"; red x: "no"; yellow ?: "cannot determine, not applicable, or not reported.".

First author, publication year	NHLBI questions													
	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13	14
Weissman et al [23], 2006	+	+	?	+	+	+	x	+	+	x	+	?	+	x
Heiskanen et al [24], 2010	+	+	?	+	x	+	x	?	+	x	+	?	+	x
Pöyhönen-Alho et al [22], 2010	+	+	?	+	+	+	?	?	+	x	+	?	+	x
Maser et al [25], 2014	+	+	?	+	x	+	?	?	+	x	+	?	+	x
Žákovičová et al [26], 2014	+	+	?	+	x	+	+	?	+	x	+	?	+	x
Odendaal et al [29], 2019	+	+	?	+	x	+	+	+	+	+	+	?	+	+
Christian et al [27], 2021	+	+	?	+	x	+	+	x	+	x	+	?	+	x
Ecklund-Flores et al [30], 2016	+	+	?	+	x	+	+	x	+	x	+	?	+	+
Voss et al [28], 2006	+	+	?	+	x	+	+	+	+	+	+	?	+	x
Amador-Licona et al [31], 2009	+	+	?	+	x	+	+	+	+	+	+	?	+	+
Liao and Jaw [32], 2011	+	+	?	+	x	+	+	+	+	+	+	?	+	x
Faber et al [33], 2004	+	+	?	+	?	+	+	+	+	+	+	?	+	x

Data Extraction and Synthesis

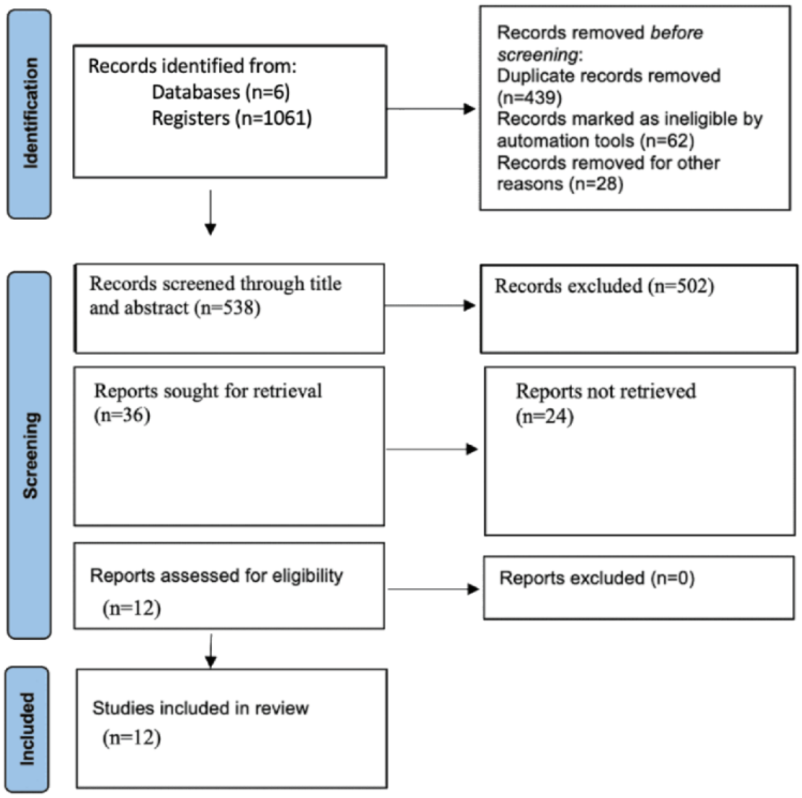
The reviewed studies were not homogenous in terms of the assessment time frame, component, and frequency; thus, a meta-analysis was not possible. A narrative synthesis was chosen to bring together the broad knowledge from a variety of approaches. This type of synthesis is not the same as a narrative description that accompanies many reviews. To synthesize the literature, we used a guideline from Popay et al [21]. The steps included (1) preliminary analysis, (2) exploration of relationships, and (3) assessment of the robustness of the synthesis. Theory development was not performed due to the exploratory nature of the research synthesized. The main synthesis consisted of extracting the descriptive characteristics of the included studies and presenting and producing a textual summary of the results. We categorized the studies based on the complication type in 4 groups. We then performed thematic analysis to extract main themes for each complication in all studies. The 4 themes developed in the results represent the main areas of knowledge available about HRV in pregnancy complications. These included study population (characteristics, sample size, exclusion criteria), design (study design, assessment duration and frequency, considered HRV metrics), measures (device used for HRV assessment), and findings.

Results

Study Selection

After removing duplicates, 538 papers were screened through review of title and abstract. Of these, 36 studies were screened by review of the full text, resulting in 12 papers that met the inclusion criteria: 5 for gestational diabetes mellitus (GDM), 4 for fetal growth, 2 for pulmonary function, and 1 for nervous system disorder. Reasons for exclusion at this stage were recorded and can be found in the flow diagram in Figure 2.

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta Analyses) diagram of studies identified via registers and databases.



Quality Assessment

The results of the NHLBI assessment are reported in Figure 1. The research objectives and questions were clear in all the studies. The study population was clearly specified for each study. The participation rate of the eligible individuals and whether the outcome assessors were blinded were not mentioned in any of the studies. Sample size justifications, power descriptions, or variance and effect estimates were not provided or were not clear in 10 of 12 (83%) studies.

Study Characteristics and Synthesis of the Results

Gestational Diabetes Mellitus (GDM)

In all studies, an oral glucose tolerance test (OGTT) was performed for either diagnosis or to confirm diagnosis of GDM. The control group included either analogous non-GDM pregnant women or nonpregnant women. In the majority of the studies (4/5, 80%), the pregnant control and GDM groups were analogous in terms of other health-related factors and sociodemographic characteristics, with GDM as the major difference between the 2 groups. Of the studies, 80% (4/5) suggested dietary and exercise or insulin therapy for the GDM group. All studies measured both frequency and time domains of the HRV assessment. In addition, 80% (4/5) of the studies specified the HRV assessment duration, and it varied between 6 minutes and 8 hours. An electrocardiogram (ECG) was used to assess HRV in 60% (3/5) of the studies. However, the type of spectrum analyzer was either not reported or varied from study to study. The gestational age was reported in 4 of the 5 (80%) studies in which the measurement was performed; the HRV measurement was conducted during pregnancy (all in the third trimester) with additional postpartum (within 3 months of childbirth) assessment in 2 of these studies.

There was no difference between GDM and non-GDM pregnant individuals in terms of HRV metrics in 80% (4/5) of the studies. In the remaining study, frequency domain components varied in terms of changes between control and GDM groups; LF (nu) was higher, and HF (nu) was lower in women with GDM than in pregnant controls [22]. For more information, see Table 1.

Table 1. Gestational diabetes mellitus (GDM) and heart rate variability (HRV; n=292).

Author, year, country	Cases	Controls	Study design	Measures	Findings
Weissman et al [23], 2006, Israel	12 pregnant with GDM at GA[a] of 24-28 weeks; mean age: 31.2 (SD 4.4) years; mean BMI: 26.4 (SD 2.2) kg/m^2; exclusion criteria: hypertension, thyroid disease, obesity, or a family history of diabetes; no smoking with no suggested diabetic management	15 pregnant without GDM (positive OGCT[b] and negative OGTT[c] tests); mean age: 30.8 (SD 3.7) years; mean BMI: 26.7 (SD 2.7) kg/m^2; analogous age, weight, BMI, and health status as the case group	Cross-sectional: outcome of HRV (VLF[d], LF[e], HF[f], TP[g], RMSSD[h], SDNN[i], SDANN[j]) was assessed for 10 min before the OGTT (phase 0) and was repeated at 60 min after glucose ingestion (phase 1); exposure: OGTT was measured once between 8 AM and 9 AM.	HRV: ECG[k] with a 12-bit analog-to-digital converter and autoregressive model for IBI[l]; GDM: Hitachi 747 method; OGCT for initial test and confirmation with OGTT (fasting, 1 hour, and 2 hours; at least two >normal)	No statistically significant changes in the TD[m] measures (TP, RMSSD, SDNN, SDANN) between the 2 groups in the different phases of the study. FD[n] metrics (LF and HF) in both groups decreased in phase 1; only HF changed significantly (without GDM: mean 104.5, SD 57.4; *P*<.01; with GDM: mean 78.7, SD 62.1; *P*<.01). LF (nu) increased (without GDM: mean 66.8, SD 9.9; with GDM: mean 70.1, SD 17.7; *P*<.05), and HF (nu) decreased (without GDM: mean 33.2, SD 9.9; with GDM: mean 29.9, SD 17.7; *P*<.05) in phase 1 with no significant difference between the 2 groups. LF/HF was higher in the GDM group in phase 0 (mean 3.0, SD 1.9) and increased in both groups in phase 1 (without GDM: mean 2.8, SD 2.1; with GDM: mean 3.9, SD 2.3; *P*<.04), with no statistical difference in changes between groups.
Heiskanen et al [24], 2010, Finland	51 pregnant with GDM; mean age: 31 (SD 1) years; dietary modification for diabetes suggested and insulin received, if necessary	28 pregnant without GDM; mean age: 32 (SD 1) years	Longitudinal: exposure of GDM diagnosed at the beginning of the 3rd trimester; outcomes of glucose, insulin, and HRV (VLF, LF, HF, TP) were measured for 10 min during the 3rd trimester of pregnancy and 3 months postpartum.	GDM: hexokinase method; OGTT (fasting, 1 hour, and 2 hours; at least one >normal) and confirmed with blood glucose profile (every 4 hours per 24 hours); continuous assessment was done (3 days/week); HRV: ECG using fast Fourier transform was used for power spectral estimates of HRV.	TP (pregnancy: mean 1183, SD 180; postpartum: mean 4036, SD 1219; *P*<.01), VLF (pregnancy: mean 417, SD 80; postpartum: mean 516, SD 84; *P*<.05), LF (pregnancy: mean 166, SD 24; postpartum: mean 374, SD 75 *P*<.001), HF (pregnancy: mean 480, SD 113; postpartum: mean 3002, SD 1113; *P*<.01), HF (nu; pregnancy: mean 0.64, SD 0.03; postpartum: mean 0.76, SD 0.03; *P*<.01) were lower during pregnancy than postpartum. LF (nu; pregnancy: mean 0.35, SD 0.03; postpartum: mean 0.22, SD 0.03; *P*<.01) and LF/HF (pregnancy: mean 91, SD 18; postpartum: mean 35, SD 6; *P*<.05) were higher during pregnancy than postpartum. There were no differences between GDM and control groups in HRV components during the third trimester and 3 months postpartum.

Author, year, country	Cases	Controls	Study design	Measures	Findings
Pöyhönen-Alho et al [22], 2010, Finland	41 Caucasian pregnant with GDM (18 with hypertension, 23 without hypertension); mean age: 34.0 (SD 5.6) years; mean BMI: 30.6 (SD 6.0) kg/m^2; exclusion criteria: smoking and any medication affecting glucose metabolism or the sympathetic nervous system; dietary modification for diabetes suggested and insulin received, if necessary	22 healthy pregnant controls (PC); mean age: 29.5 (SD 4.9) years; mean BMI: 26.9 (SD 3.0) kg/m^2; and 14 healthy nonpregnant controls (NPC); mean age: 30.4 (SD 6.9) years; mean BMI: 26.5 (SD 6.3) kg/m^2; analogous BMI and exclusion criteria as the case group	Cross-sectional: exposures of GDM diagnosed in pregnancy; outcome of HRV (SDNN, SDANN, LF, HF, VLF) was assessed once at night (11 PM through 8 AM for 8 hours)	GDM: OGTT (fasting, 1 hour, and 2 hour); HRV: Holter	No difference between 2 GDM and PC groups except at 4 AM: LF (nu) was higher (GDM: mean 60.0, SD 12.4; PC: mean 48.5, SD 12.8; *P*=.01), and HF (nu; GDM: mean 40.0, SD 12.8; PC: mean 51.5, SD 12.8; *P*=.01) and HF power (GDM: mean 842.7, SD 652.6; PC: mean 1053.4, SD 1024.4; *P*=.02) were lower. There were differences between GDM and NPC for the following: SDNN (GDM: mean 87.6, SD 23.8; NPC: mean 136.1, SD 26.9; *P*<.001), SDANN (GDM: mean 62.6, SD 18.9; NPC: mean 103.8, SD 23.7; *P*<.001), LF (GDM: mean 1178.5, SD 655.1; NPC: mean 3062.9, SD 1789.0; *P*<.001), HF (GDM: mean 842.7, SD 652.6; NPC: mean 1631.4, SD 1201.4; *P*=.02), and VLF (GDM: mean 3259, SD 1646; NPC: mean 5897, SD 1614; *P*<.001). In all 3 groups, changes in LF (nu), HF (nu), and LF/HF remained unchanged during the 8-hour time frame.
Maser et al [25], 2014, United States	31 pregnant with GDM; mean age: 32 (SD 4) years; mean BMI: 35.2 (SD 7.9) kg/m^2; exclusion criteria: type 1 or 2 diabetes mellitus, pregestational hypertension, preeclampsia during current pregnancy, preterm labor, cervical shortening, current multiple fetuses, cardiopulmonary diseases; dietary and exercise modification for diabetes suggested and insulin received, if needed	12 without GDM; mean age: 30 (SD 5) years; mean BMI: 30.9 (SD 6.4) kg/m^2; analogous with case group; exclusion criteria: the same as the case group in terms of age, blood pressure, BMI, and GA	Longitudinal: exposures of GDM diagnosed at GA of 30-35 weeks; outcome of overnight HRV (LF, HF, TP) was assessed for 6 min during the mid-third trimester (GA: 30-35 week) and 2-3 months postpartum	HRV: ANX 3.0 (ANSAR Medical Technologies Inc); GDM: OGCT for initial test and confirmation with OGTT (fasting, 1 hour, and 2 hours; at least two >normal); continuous glucose assessment was performed once a week.	There was no difference between GDM and control groups during late pregnancy and after delivery for any HRV metrics including normalized and nonnormalized metrics.

Author, year, country	Cases	Controls	Study design	Measures	Findings
Žákovičová et al [26], 2014, Czech Republic	35 pregnant with GDM; mean age: 32 (SD 4) years; mean BMI: 28.2 (SD 3.8) kg/m^2; exclusion criteria: history of hypertension, preeclampsia, and chronic diseases except for controlled hypothyroidism; dietary modification for diabetes suggested and insulin received, if necessary	31 pregnant without GDM; mean age: 30.3 (4.2 years); mean BMI: 27.1 (SD 4.1) kg/m^2; analogous in weight, height, BMI, and age with the case group	Cross-sectional: exposures of GDM diagnosed in the GA of 24-28 weeks; outcome of HRV (LF, HF, TP) was measured at GA of 36 weeks	GDM: OGCT was used to test for GDM; cases were confirmed with OGTT (fasting, 1 hour, and 2 hours; at least 1 >normal); OGTT was continued biweekly until childbirth; for HRV, ECG was used, and the VariaCardio TF4 device and Fourier transform algorithm were used for analysis.	TP, HF, and LF/HF did not differ between the 2 groups.

[a]GA: gestational age.
[b]OGCT: oral glucose challenge test.
[c]OGTT: oral glucose tolerance test.
[d]VLF: very low frequency.
[e]LF: low frequency.
[f]HF: high frequency.
[g]TP: total power.
[h]RMSSD: root mean square of successive differences between normal heartbeats.
[i]SDNN: standard deviation of the normal-to-normal R-R intervals.
[j]SDANN: standard deviation of the average normal-to-normal (NN) intervals.
[k]ECG: electrocardiogram.
[l]IBI: interbeat interval.
[m]TD: time domain.
[n]FD: frequency domain.

Fetal Growth

The study populations were mainly African (American or non-American), non-Hispanic White, Latino, and German. In 75% (3/4) of the studies, the population was healthy excluding the potential risk factors for ANS; of these, 1 study, however, included obese or overweight women as more than 50% of its population. In 1 of the 4 studies (25%), the population had hypertensive disorder along with fetal growth retardation. The considered HRV metrics, including the time domain (SDNN, RMSSD) and frequency domain (LF, HF, VLF), varied from study to study. HRV assessment was conducted in the second or third trimester; the assessment frequency varied from 1 time to 5 times among the studies, with 50% (2/4) of the studies conducting an assessment 1 time. The assessment duration varied from 10 minutes to 55 minutes in the studies. Fetal growth was assessed based on birth weight or *z* score for birth weight in 75% (3/4) of the studies. One study, however, used small for gestational age as the measurement of fetal growth. ECG was used for HRV assessment in all the studies, but the applied standards varied from study to study.

In the studies, among the considered HRV metrics, SDNN, RMSSD, and HF had a significant negative association with fetal growth. In 75% (3/4) of the studies, RMSSD, as the commonly assessed metric, had a significant negative association with fetal growth. In 1 of the studies that included White and African American women as the 2 study populations, HRV metrics did not differ with fetal growth in the African American group [27]. Also, in 1 of the studies [28], 2 case groups were included, both with impaired uterine perfusion; 1 group included poor outcome (eg, hypertensive disorder, impaired fetal growth), and the other included no poor outcomes. In the case group with poor outcomes, failing to perform a separate analysis on each poor outcome in terms of its independent associations with HRV metrics hindered the understanding of the actual association between fetal growth and autonomic function. This is problematic, as hypertensive disorder is known to be linked to autonomic dysfunction, which could mask the impact of fetal growth on ANS function. Thus, a pooled analysis without distinguishing the outcomes can threaten the specificity and therefore the validity of the findings. For more information, see Table 2.

Table 2. Fetal growth and heart rate variability (HRV; n=5988).

Author, year, country	Case	Control	Study design	Measures	Findings
Odendaal et al [29], 2019, South Africa	5655 pregnant in the first trimester (GA[a]: ≥6 weeks) between 16 years and 45 years old, singleton; prior history of heart disease, hyperthyroidism, diabetes, and placental abruption was found in <1%; hypertension was found in 12.1%; preeclampsia was found in 3.9%; mean age 24.5 (SD 6.0) years; mean BMI 25.5 (SD 5.7) kg/m^2; exclusion criteria: N/A[b]	N/A	Retrospective study using data from 2007-2015; exposure of HRV (SDNN, RMSSD) was assessed for 36-55 minutes at 3 times, at 20-24, 28-32, and ≥34 weeks; outcome of birth weight was assessed on a case report form.	HRV: ECG[c] with 5 electrodes was used and imported into MATLAB. Dawes-Redman guidelines were used to quantify IBI[d] features. Artifact management and sparsity-based epoch rejection were used for data preprocessing. Fetal growth: birth weight and *z* score	At both 20-24 weeks and ≥34 weeks, birth weight correlated positively with maternal heart rate but negatively with SDNN[e] (Spearman correlation=−0.0437; *P*<.02) and RMSSD[f] (Spearman correlation=−0.0627; *P*<.01).
Christian et al [27], 2021, United States	39 pregnant (19 African American; 20 White), at GA of 21-24 weeks; exclusion criteria: tobacco/drug use; chronic diseases such as endocrine, cardiovascular, cancer, diabetes, hypertensive disorder, anemia, medication use (psychotic, antibiotic), and excessive caffeine use; BMI ≥30 kg/m^2	N/A	Retrospective study; secondary data analysis between 2009 and 2011; exposure of HRV (HF[g]) was assessed for 10 min in the afternoon once in the second trimester; outcome of birth weight was assessed once right after childbirth using the medical record.	HRV: ECG (Task Force of the European Society of Cardiology) was used for signal capturing and artifact management; offline signal inspection with Mindware Technology's HRV 2.51 software; fetal growth: birth weight (grams) collected postdelivery and from the medical chart	White group had a negative relationship between HF and birth weight (correlation coefficient=–0.757, *P*=.002); no relationship was found in the African American group.
Ecklund-Flores et al [30], 2016, United States	227 pregnant; GA: 36 weeks; 50% obese or overweight; 54% primigravidae, singleton; 68% Latino; mean age 26.45 (SD 6.02) years; mean BMI 25.32 (SD 4.99) kg/m^2; exclusion criteria: no GDM, hypertension, or other related medical conditions, and no cigarette, alcohol, or drug use during pregnancy	N/A	Longitudinal observational study: outcome of birth weight adjusted for GA at birth and sex; exposure of adjusted HRV (RMSSD) was assessed for 5 min at baseline (GA: 36 weeks)	HRV: ECG digitized at 500 Hz (National Instruments 16XE50); fetal growth: birth weight was assessed after birth, and the data collection method was not reported.	RMSSD and birth weight had significant negative associations (Pearson correlation: –0.13; *P*<.001).
Voss et al [28], 2006, Germany	16 pregnant with abnormal uterine perfusion and normal outcome (AP-NO), age: 29 (range 28 to 33) years; 19 women with abnormal uterine perfusion and pathologic outcome (AP-PO; small for GA, gestational hypertension, preeclampsia), singleton, age: 26 (range 25 to 30) years	32 healthy pregnant with normal uterine perfusion (CON), age: 28 (24 to 31) years; analogous to the case group in maternal age, gravidity, and parity	Longitudinal observational study: exposure of HRV (normalized LF[h], VLF[i], mean NN[j], SDNN, RMSSD) for 30 min between 8 AM and 12 PM, 5 times during pregnancy (GA: 18-22, 23-26, 27-30, 31-34, 35-37weeks); outcome of small for GA was assessed (birth weight <10th percentile of an own-reference group).	Fetal growth: data collection record was not reported; HRV: ECG (1600 Hz) based on task force standards	RMSSD decreased (*P*=.009), and VLF (*P*<.001) and LF (n; *P*=.003) increased in CON during pregnancy. No HRV parameter changed significantly in the course of gestation in AP-NO and AP-PO. No intergroup differences in HRV were found between CON and AP-NO. The comparison of AP-PO with CON and AP-NO, however, revealed a significant increase in mean NN (*P*=.03) as well as RMSSD (CON vs AP-PO: *P*=.008; AP-NO vs AP-PO: *P*=.01). AP-PO group had a significantly increased SDNN compared with AP-NO (*P*=.03). Effect measure amount was not clear.

[a]GA: gestational age.

[b]N/A: not applicable.

[c]ECG: electrocardiogram.

[d]IBI: interbeat interval.

[e]SDNN: standard deviation of the normal-to-normal R-R intervals.

[f]RMSSD: root mean square of successive differences between normal heartbeats.

[g]HF: high frequency.

[h]LF: low frequency.

[i]VLF: very low frequency.

[j]NN: N-N interval.

Cardiovascular and Hemodynamic Variables

To avoid redundancy and provide a comprehensive overview of the literature, we incorporated the findings of a previously published systematic review that explored the association between hypertensive disorders and HRV. We did not include the review among our reviewed studies in the Results section but included a summary of the review's main results and conclusions in the Discussion section of our review and Multimedia Appendix 3 to provide transparency and facilitate replication.

Other

Other results were related to the pulmonary system and the central nervous system.

Regarding the pulmonary system, Amador-Licona et al [31] conducted a study to investigate cardiovascular autonomic and pulmonary function in obese pregnant women; 178 pregnant women with no chronic diseases were included. The study measures were assessed using spirometry, 10-minute oximetry, and 60-minute ECG monitoring twice during pregnancy. All assessments were performed between 8 AM and 10 AM to maintain consistency in the measurements. Their findings indicated that, in the obese group, the change in forced expiratory volume at 1 second to forced vital capacity (FEV1/FVC) during pregnancy was linked to the LF/HF metric in the third trimester after adjusting for confounding factors such as insulin, weight gain, and blood pressure (Table 3).

Liao and Jaw [32] also studied the potential of HRV analysis to assess progress in amniotic fluid embolism and disseminated intravascular coagulopathy. They compared 2 cases with 105 healthy pregnant women. Entropy (a nonlinear HRV metric) was used to assess HRV and significantly decreased in amniotic fluid embolism and disseminated intravascular coagulopathy and increased in the intensive care unit during recovery. They concluded that entropy analysis has the potential to be applied to monitor amniotic fluid embolism and the progress of disseminated intravascular coagulopathy in a patient (Table 3).

Regarding the central nervous system, a case was reported in Germany, suggesting the ANS is dysregulated right before and during the grand mal seizure [33]. The case was a 21-year-old gravida 1 para 0 who had epilepsy due to a frontotemporal arteriovenous malformation and was on anticonvulsant medication. The HRV was assessed twice, and the results indicated no changes in week 20; however, during the second monitoring session (week 24) when the patient developed a grand mal seizure, HRV was significantly altered. At the beginning of the monitoring (12 minutes prior to seizure), VLF increased, LF was delayed, and HF remained unchanged. HF, however, started to increase afterward.

Table 3. Heart rate variability (HRV) and other complications such as pulmonary disease or seizures (n=376).

Author, year, country	Case	Control	Study design	Measures	Findings
Pulmonary system (n=375)					
Amador-Licona et al [31], 2009, Mexico	178 pregnant women (88 obese, BMI>27 kg/m^2); age: 28.2 years; exclusion criteria: no chronic diseases (eg, hypertension, diabetes)	90 nonobese (BMI >18.5 kg/m^2 and <24.9 kg/m^2); age: 26.6 years; exclusion criteria: no chronic diseases (eg, hypertension, diabetes)	Longitudinal: exposures of spirometry, 10-min oximetry; outcome of HRV by 60-minute electrocardiograph monitoring; both exposure and outcome were assessed twice between 8 AM and 10 AM during pregnancy, at GA[a] of 24-28 weeks and 36-37 weeks	Spirometry (using EasyOne 2001-2 spirometer, NDD) and oximetry (using the Onix 2001 oximeter; HRV using a 3-channel Holter recorder (model GBI-3S, Galix Biomedical Instrumentation Inc)	Change in FEV1/FVC during pregnancy was linked to the LF[c]/HF[d] metric in the third trimester after adjusting for confounding factors such as insulin, weight gain, and blood pressure (β=−0.42; *P*<.001).
Liao and Jaw [32], 2011, Taiwan	n=2; case 1 (sudden acute dyspnea with cyanosis): 36-year-old primigravid woman was admitted in active labor with no history of epilepsy, cardiopulmonary, or renal disease; caes 2 (acute dyspnea with cyanosis): 35-year-old, gravid 2, para 1 woman was admitted at 37 weeks of gestation due to labor pain and ruptured membranes.	105 healthy pregnant women with no complication during labor	Longitudinal: exposures of amniotic fluid embolism and disseminated intravascular coagulopathy; outcome of HRV (entropy) was assessed continuously after admission to labor until recovery form the complication.	Amniotic fluid embolism and disseminated intravascular coagulopathy: specialist's diagnosis; HRV: ECG[e]	HRV significantly decreased in amniotic fluid embolism and disseminated intravascular coagulopathy and increased in ICU[f] during recovery from these complications.
Central nervous system (n=1)					
Faber et al [33], 2004, Netherland	n=1; 21-year-old gravid 1, para 0 woman had epilepsy due to a frontotemporal arteriovenous malformation in the 24th week of gestation	N/A[g]	Longitudinal: exposure of HRV (frequency domain: VLF[h], LF, HF, LF/HF) was assessed starting from 20 weeks of GA through 24 weeks of GA for 30 minutes every 4 weeks and for 12 minutes until the onset of seizure; outcome of grand mal seizure	HRV: N/A; grand mal seizure: tonic-clonic convulsions diagnosed by a specialist	HRV showed significant alterations at 12 minutes prior to the seizure: VLF increased; LF was delayed, and HF remained unchanged. HF, however, started to increase afterward.

[a]GA: gestational age.
[b]FEV1/FVC: forced expiratory volume at 1 second to forced vital capacity.
[c]LF: low frequency.
[d]HF: high frequency.
[e]ECG: electrocardiogram.
[f]ICU: intensive care unit.
[g]N/A: not applicable.
[h]VLF: very low frequency.

Difference in HRV Changes in Pregnancy Complications

The significance of changes in various HRV components varied from complication to complication. However, impaired fetal growth and hypertensive disorder reached common significance in an HRV component (RMSSD) based on the majority of the studies but in the opposite direction. This reverse association supports studies that linked a high birth weight to maternal hypertension [34-37]. See more information in Table 4.

Table 4. Heart rate variability (HRV) components in pregnancy complications, with the difference in HRV based on the majority of studies.

Pregnancy complications	Linear HRV component in the frequency domain (FD)							Linear HRV component in the time domain (TD)				Nonlinear HRV component
	HF[a]	LF[b] (nu)	LF	HF (nu)	LF/HF	LF/HF (ln)	VLF[c]	TP[d]	SDNN[e]	RMSSD[f]	SDANN (ln)	Entropy
GDM[g] (4 of 5 studies)	Not significant[h]	Not significant[h]	Not significant[h]	Not significant[h]	Not significant[h]	Not significant[h]	Not significant[h]	Not significant[h]	Not significant[h]	Not significant[h]	Not significant[h]	Not significant[h]
Impaired fetal growth (4 of 4 studies)	—[i]	—	—	—	—	—	—	—	—	Increased	—	—
Hypertensive disorders (n=1) that were reviewed in 24 studies												
GH[j] (n≥12)	Decreased	Increased	—	—	—	Increased	—	Decreased	Decreased	Decreased	—	—
Preeclampsia (n≥12)	Decreased	Increased	—	Decreased	Increased	Increased	—	—	Decreased	—	Increased	—
HPD[k] (n≥12)	Not significant[h]	Not significant[h]	Not significant[h]	Not significant[h]	Not significant[h]	Not significant[h]	Not significant[h]	Not significant[h]	Not significant[h]	Not significant[h]	Not significant[h]	Not significant[h]
Pulmonary function (n=2)												
Low FEV1/FVC[l] (n=1)	—	—	—	—	Increased	—	—	—	—	—	—	—
Amniotic fluid embolism (n=1)	—	—	—	—	—	—	—	—	—	—	—	Decreased
Central nervous system (n=1)												
Grand mal seizure	—	—	Decreased	—	—	—	Increased	—	—	—	—	—

[a]HF: high frequency.
[b]LF: low frequency.
[c]VLF: very low frequency.
[d]TP: total power.
[e]SDNN: standard deviation of the normal-to-normal R-R intervals.
[f]RMSSD: root mean square of successive differences between normal heartbeats.
[g]GDM: gestational diabetes mellitus.
[h]The common biomarkers in the majority of the studies showed no significant difference from the controls.
[i]Not assessed by the majority of the studies.
[j]GH: gestational hypertension
[k]HPD: hypertensive disorders
[l]FEV1/FVC: forced expiratory volume at 1 second to forced vital capacity.

Discussion

Principal Findings and Comparison With Existing Literature

The findings of this study support potential ANS dysregulation in some pregnancy complications including hypertensive disorder and fetal growth.

Cardiovascular and Hemodynamic

These adaptations in the cardiovascular system include adjusting blood pressure, blood volume, and heart activity. Systemic vasodilation is the primary initial hemodynamic event that begins during the luteal phase of the menstrual cycle following the release of chemical mediators by the corpus luteum [9-11]. These hemodynamic adaptations provide the required blood for fetal needs of growth [13] through increased sympathetic tone and parasympathetic withdrawal.

Moors et al [38] conducted a systematic review on the difference in autonomic function using HRV between hypertensive and normotensive pregnancies. Higher LF/HF and lower HF and RMSSD were found in the hypertensive group, compared with the normotensive pregnant controls. This can be explained by overactivation of sympathetic over parasympathetic tone resulting from functional or structural damage of the vascular or nervous system. As we discussed earlier, ANS physiologically regulates hemodynamics (eg, blood volume, cardiac output) in response to general vasodilation and consecutive hypotension. This requires sympathetic activation and, thus, sympathetic dominance and parasympathetic withdrawal. Failure of the ANS to function in response to decreased blood pressure following pregnancy-induced systemic vasodilation may increase the risk of hypertensive disorders.

GDM

Pregnancy affects both ANS function and metabolic regulations [1,39,40]. In addition to ANS alterations during pregnancy, a diabetogenic effect on metabolism has been indicated in pregnant women. Placental-derived hormones induce this impact by reprogramming maternal physiology to achieve an insulin-resistant state, by reducing insulin sensitivity [41]. This, in turn, increases the risk of developing new diabetes (GDM) or worsening existing diabetes during pregnancy.

Due to the destructive impact of diabetes on the ANS, intensified diabetic neuropathy in pregnant women with GDM is expected compared with non-GDM pregnant and nonpregnant individuals. In this review study, however, an association between GDM and maternal ANS during pregnancy and postpartum was not supported. Part of this can be due to the short assessment duration. According to the literature about HRV assessment protocols, any duration less than 24 hours (circadian rhythm) lacks the ability to capture reliable HRV metrics and, thus, ANS function [15]. Longer recording epochs better represent slower fluctuations affected by circadian rhythms as well as the cardiovascular response to a wider range of environmental stimuli. Interestingly, in this review, studies that assessed HRV for a shorter period (minutes) showed no difference between pregnant women with GDM and without GDM, whereas those with a longer assessment time (hours) indicated a difference between the 2 groups. For example, Pöyhönen-Alho et al [22] assessed HRV for 8 hours and indicated significant differences between the 2 groups in terms of frequency domain metrics.

Additionally, most of the studies often assessed GDM and autonomic function in a cross-sectional or longitudinal method with a short interval between the 2 assessments (GDM diagnosis and HRV assessment), failing to consider a period for the autonomic system to be affected. This consideration is important, as GDM can be asymptomatic for the long term and may not indicate a significant influence on the ANS in the short term. One of the potential solutions may be the use of glycated hemoglobin to represent glycemic control in the long term. The impact of long-term diabetes control on HRV is more valuable and clinically useful to study than diagnosed diabetes, as it focuses on control rather than treatment for diabetes, which is an often chronic and noncurable condition and relies on symptom management. In the reviewed studies, 80% suggested dietary modification for diabetes, insulin therapy, or exercise for people with diabetes. This manipulation may have impacted the findings of the studies, resulting in no difference in ANS function in people with diabetes. This manipulation leads to a failure in representing actual glycemic control following the autonomy of the population in real life.

Fetal Growth

Our study indicates that there is a negative relationship between fetal development and vagal tone during pregnancy. This association can be partially explained by the fetal-maternal physiology for developmental adaptations during pregnancy. The uterine complex, including the placenta, is a multifunctional organ, representing the vital interface between the mother and fetus, and placental blood circulation holds the link between the mother and fetus. The required supply of oxygen and nutrients for fetal development is maintained via uteroplacental perfusion. On the other hand, circulation and transportation of the supply via blood flow are regulated by the ANS by adjusting vascular resistance. Sympathetic and parasympathetic tone determine the vascular tone in response to hemodynamic dysregulations following pregnancy-induced demands. Any changes in this tone can affect uterine perfusion by increasing or decreasing uteroplacental vascular resistance. Impaired perfusion, if it lasts long or occurs frequently, in turn can cause less supply transfer and, thus, lower birth weight [42,43]. It is worth mentioning that this uteroplacental autonomic regulation is originally for fetal protection against harmful maternal products (eg, cortisol and insulin). This protective mechanism, however, can result in poor fetal outcomes by reducing the essential blood supply to the fetus. Our findings, however, challenge studies that have linked higher sympathetic tone to low birth weight. Our review indicates that increased vagal tone is associated with low birth weight, which is in conflict with the theory that suggests the potential of increased sympathetic tone for explaining low birth weight by disturbing uteroplacental perfusion.

Interestingly, maternal autonomic and fetal autonomic tone have also been linked in a recent study [44]. This association can be explained by vascular resistance of the uteroplacental pathway following maternal ANS regulation in response to fetal developmental demands. This mechanism may lead to fetal autonomic regulation in response to the changed perfusion. This finding, if supported by more relevant literature, may indicate the potential of maternal ANS function to predict fetal autonomic regulation.

Complication Co-occurrence

It is evident that pregnancy complications are often correlated and tend to co-occur. For example, individuals with GDM may increase the chance of developing hypertensive disorder by 30% [45-48], which is explainable by diabetic neuropathy and vascular damage. Indeed, vascular impairment resulting from neuropathy may lead to hemodynamic dysregulation and, thus, hypertensive disorder. These two conditions, GDM and hypertensive disorders, can also impact other pregnancy complications such as fetal growth. For example, 15% to 45%

of babies born to mothers with diabetes have macrosomia [47], which also occurs in babies born to hypertensive pregnant women [34-37]. This correlation explains how fetal development can be affected by both GDM and hypertensive disorders. To explain the uncertainty in the responsiveness of the ANS to GDM in this review, the stage of diabetic neuropathy may be the determinative factor. To understand this critical stage when the hemodynamic dysregulation starts to occur, a continuous ANS assessment is required to indicate when and how GDM may lead to neuropathy and vascular damage and thus ANS tone alteration.

Strengths and Limitations

This study is a PRISMA-guided systematic review that sheds light on the utilization of HRV in representing ANS dysregulation and its possible link with pregnancy complications. Although this study provides insight in understanding the potential pathway for pregnancy complications, a meta-analysis was not conducted due to the divergent time periods used for the measurement. Another limitation of this study is that it was not registered prior to beginning the review.

In terms of using HRV, HRV has been suggested as a health index for various concepts such as mental distress, physical activity, and meditation. Although these suggestions can be accepted, considering the aforementioned impact on ANS regulation, it may affect the specificity of HRV for reflecting the ANS. Careful reconsideration is needed to define biomarkers for these concepts. Additionally, in the evidence, there is a concern that HRV may only reflect the cardiac vagal tone and not necessarily sympathetic activity. Although most of the HRV metrics represent parasympathetic activity, there are still metrics that reflect the sympathetic/parasympathetic balance. In addition, due to the dynamic balance between sympathetic and parasympathetic activity, it is expected that the activity of each branch can predict the other, acting like the 2 sides of a homeostasis seesaw.

One of the advantages of using HRV to assess the ANS is the ability to continuously assess using wearable smart technologies. This continuity in the assessment may increase the reliability of the assessments in reflecting pregnancy and the ANS. This is due to the dynamic and ever-changing nature of both pregnancy and the ANS that cannot be represented by episodic and short-term assessments. HRV also can be assessed cost-effectively and noninvasively, reflecting the real-life function of the ANS.

Conclusion and Implications

Due to the divergent HRV bands considered for assessment in the different studies, it was not practical to compare the studies comprehensively. However, our findings, which are based on the majority of studies for each complication (within-group), suggest that ANS function has been associated with some common pregnancy complications including hypertensive disorder and fetal growth. However, existing studies do not support an association between the ANS and GDM. Studies that have linked pulmonary and central nervous system disorders to ANS function did not provide sufficient evidence to draw conclusions. More studies are needed to understand how the ANS, through HRV, is associated with other systems during pregnancy. Future studies are suggested to cover the methodological gaps in HRV assessment (eg, short assessment duration, noncontinuous assessment, low-quality standards) to represent more reliable findings.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Heart rate variability (HRV) components and metrics.
[DOCX File , 14 KB - ijmr_v12i1e44430_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOCX File , 14 KB - ijmr_v12i1e44430_app2.docx]

Multimedia Appendix 3
Hypertensive disorders and heart rate variability.
[DOCX File , 15 KB - ijmr_v12i1e44430_app3.docx]

References

1. Cerritelli F, Frasch MG, Antonelli MC, Viglione C, Vecchi S, Chiera M, et al. A review on the vagus nerve and autonomic nervous system during fetal development: searching for critical windows. Front Neurosci 2021 Sep 20;15:721605 [FREE Full text] [doi: 10.3389/fnins.2021.721605] [Medline: 34616274]
2. Waxenbaum A, Reddy V, Varacallo M. Anatomy, Autonomic Nervous System. In: StatPearls. Treasure Island, FL: StatPearls Publishing; 2021.
3. Karemaker JM. An introduction into autonomic nervous function. Physiol Meas 2017 May;38(5):R89-R118. [doi: 10.1088/1361-6579/aa6782] [Medline: 28304283]

4. Wehrwein EA, Orer HS, Barman SM. Overview of the anatomy, physiology, and pharmacology of the autonomic nervous system. Compr Physiol 2016 Jun 13;6(3):1239-1278. [doi: 10.1002/cphy.c150037] [Medline: 27347892]
5. Ziemssen T, Siepmann T. The investigation of the cardiovascular and sudomotor autonomic nervous system-a review. Front Neurol 2019 Feb 12;10:53 [FREE Full text] [doi: 10.3389/fneur.2019.00053] [Medline: 30809183]
6. Bruno RM, Ghiadoni L, Seravalle G, Dell'oro R, Taddei S, Grassi G. Sympathetic regulation of vascular function in health and disease. Front Physiol 2012;3:284 [FREE Full text] [doi: 10.3389/fphys.2012.00284] [Medline: 22934037]
7. Braeken M. Psychological Functioning and the Autonomic Nervous System During Pregnancy: Impact on Mother and Child. Tilburg University. 2014. URL: https://pure.uvt.nl/ws/portalfiles/portal/10985153/braeken_psychological_11_04_2014_emb_tot_08_09_2016_1.pdf [accessed 2023-04-23]
8. Johnson JO, Hemmings HC, Talmage DE. Autonomic nervous system physiology. In: Pharmacology and Physiology for Anesthesia. Philadelphia, PA: Saunders; 2013:208-217.
9. Conrad KP. Maternal vasodilation in pregnancy: the emerging role of relaxin. Am J Physiol Regul Integr Comp Physiol 2011 Aug;301(2):R267-R275 [FREE Full text] [doi: 10.1152/ajpregu.00156.2011] [Medline: 21613576]
10. Leo CH, Jelinic M, Ng HH, Marshall SA, Novak J, Tare M, et al. Vascular actions of relaxin: nitric oxide and beyond. Br J Pharmacol 2017 May 30;174(10):1002-1014 [FREE Full text] [doi: 10.1111/bph.13614] [Medline: 27590257]
11. Tkachenko O, Shchekochikhin D, Schrier RW. Hormones and hemodynamics in pregnancy. Int J Endocrinol Metab 2014 Apr 01;12(2):e14098 [FREE Full text] [doi: 10.5812/ijem.14098] [Medline: 24803942]
12. Balajewicz-Nowak M, Furgala A, Pitynski K, Thor P, Huras H, Rytlewski K. The dynamics of autonomic nervous system activity and hemodynamic changes in pregnant women. Neuro Endocrinol Lett 2016;37(1):70-77. [Medline: 26994389]
13. Ishida J, Matsuoka T, Saito-Fujita T, Inaba S, Kunita S, Sugiyama F, et al. Pregnancy-associated homeostasis and dysregulation: lessons from genetically modified animal models. J Biochem 2011 Jul 25;150(1):5-14. [doi: 10.1093/jb/mvr069] [Medline: 21613291]
14. Duong HTH, Tadesse GA, Nhat PTH, Hao NV, Prince J, Duong TD, et al. Heart rate variability as an indicator of autonomic nervous system disturbance in tetanus. Am J Trop Med Hyg 2020 Feb;102(2):403-407 [FREE Full text] [doi: 10.4269/ajtmh.19-0720] [Medline: 31833471]
15. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. Front Public Health 2017;5:258 [FREE Full text] [doi: 10.3389/fpubh.2017.00258] [Medline: 29034226]
16. Ernst G. Heart-rate variability-more than heart beats? Front Public Health 2017 Sep 11;5:240 [FREE Full text] [doi: 10.3389/fpubh.2017.00240] [Medline: 28955705]
17. Sharifiheris Z, Rahmani A, Onwuka J, Bender M. The utilization of heart rate variability for autonomic nervous system assessment in healthy pregnant women: systematic review. JMIR Bioinform Biotech 2022;3(1):e36791 [FREE Full text] [doi: 10.2196/36791]
18. Zygmunt A, Stanczyk J. Methods of evaluation of autonomic nervous system function. Arch Med Sci 2010 Mar 01;6(1):11-18 [FREE Full text] [doi: 10.5114/aoms.2010.13500] [Medline: 22371714]
19. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan 01;4(1):1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
20. Study Quality Assessment Tools. National Heart, Lung, and Blood Institute. URL: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools [accessed 2023-05-22]
21. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the Conduct of Narrative Synthesis in Systematic Reviews. Lancaster University. 2006 Apr. URL: https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/fhm/dhr/chir/NSsynthesisguidanceVersion1-April2006.pdf [accessed 2023-04-23]
22. Pöyhönen-Alho M, Viitasalo M, Nicholls MG, Lindström BM, Väänänen H, Kaaja R. Imbalance of the autonomic nervous system at night in women with gestational diabetes. Diabet Med 2010 Sep;27(9):988-994. [doi: 10.1111/j.1464-5491.2010.03062.x] [Medline: 20722671]
23. Weissman A, Lowenstein L, Peleg A, Thaler I, Zimmer EZ. Power spectral analysis of heart rate variability during the 100-g oral glucose tolerance test in pregnant women. Diabetes Care 2006 Mar;29(3):571-574. [doi: 10.2337/diacare.29.03.06.dc05-2009] [Medline: 16505508]
24. Heiskanen N, Saarelainen H, Kärkkäinen H, Valtonen P, Lyyra-Laitinen T, Laitinen T, et al. Gestational diabetic patients with adequate management have normal cardiovascular autonomic regulation during the third trimester of pregnancy and 3 months after delivery. J Diabetes Complications 2010 Jul;24(4):234-241. [doi: 10.1016/j.jdiacomp.2008.12.006] [Medline: 19282202]
25. Maser RE, Lenhard MJ, Kolm P. Autonomic modulation in gestational diabetes mellitus. J Diabetes Complications 2014 Sep;28(5):684-688. [doi: 10.1016/j.jdiacomp.2014.05.005] [Medline: 24972765]
26. Žákovičová E, Kittnar O, Slavíček J, Medová E, Šváb P, Charvát J. ECG body surface mapping in patients with gestational diabetes mellitus and optimal metabolic compensation. Physiol Res 2014 Dec 16;63(Suppl 4):S479-S487 [FREE Full text] [doi: 10.33549/physiolres.932921] [Medline: 25669679]

27. Christian LM, Koenig J, Williams DP, Kapuku G, Thayer JF. Impaired vasodilation in pregnant African Americans: Preliminary evidence of potential antecedents and consequences. Psychophysiology 2021 Jan 11;58(1):e13699. [doi: 10.1111/psyp.13699] [Medline: 33040402]
28. Voss A, Baumert M, Baier V, Stepan H, Walther T, Faber R. Autonomic cardiovascular control in pregnancies with abnormal uterine perfusion. Am J Hypertens 2006 Mar;19(3):306-312. [doi: 10.1016/j.amjhyper.2005.08.008] [Medline: 16500519]
29. Odendaal H, Kieser E, Nel D, Brink L, du Plessis C, Groenewald C, et al. Effects of low maternal heart rate on fetal growth and birthweight. Int J Gynaecol Obstet 2019 Aug 03;146(2):250-256 [FREE Full text] [doi: 10.1002/ijgo.12873] [Medline: 31131885]
30. Ecklund-Flores L, Myers MM, Monk C, Perez A, Odendaal HJ, Fifer WP. Maternal depression during pregnancy is associated with increased birth weight in term infants. Dev Psychobiol 2017 Apr;59(3):314-323 [FREE Full text] [doi: 10.1002/dev.21496] [Medline: 28323349]
31. Amador-Licona N, Guízar-Mendoza JM, Juárez M, Linares-Segovia B. Heart sympathetic activity and pulmonary function in obese pregnant women. Acta Obstet Gynecol Scand 2009 Jan;88(3):314-319 [FREE Full text] [doi: 10.1080/00016340802702201] [Medline: 19241226]
32. Liao W, Jaw F. A noninvasive evaluation analysis of amniotic fluid embolism and disseminated intravascular coagulopathy. J Matern Fetal Neonatal Med 2011 Nov 19;24(11):1411-1415 [FREE Full text] [doi: 10.3109/14767058.2010.549975] [Medline: 21247233]
33. Faber R, Stepan H, Baumert M, Voss A, Walther T. Changes of blood pressure and heart rate variability precede a grand mal seizure in a pregnant woman. J Perinat Med 2004;32(6):538-540. [doi: 10.1515/JPM.2004.124] [Medline: 15576279]
34. Kiy AM, Rugolo LM, Luca AKCD, Corrente JE. Growth of preterm low birth weight infants until 24 months corrected age: effect of maternal hypertension. J Pediatr (Rio J) 2015 May;91(3):256-262 [FREE Full text] [doi: 10.1016/j.jped.2014.07.008] [Medline: 25431856]
35. Sato N, Fudono A, Imai C, Takimoto H, Tarui I, Aoyama T, et al. Placenta mediates the effect of maternal hypertension polygenic score on offspring birth weight: a study of birth cohort with fetal growth velocity data. BMC Med 2021 Nov 04;19(1):260 [FREE Full text] [doi: 10.1186/s12916-021-02131-0] [Medline: 34732167]
36. Xiong X, Demianczuk NN, Buekens P, Saunders LD. Association of preeclampsia with high birth weight for age. Am J Obstet Gynecol 2000 Jul;183(1):148-155. [doi: 10.1067/mob.2000.105735] [Medline: 10920323]
37. Xiong X, Fraser WD. Impact of pregnancy-induced hypertension on birthweight by gestational age. Paediatr Perinat Epidemiol 2004 May;18(3):186-191. [doi: 10.1111/j.1365-3016.2004.00553.x] [Medline: 15130157]
38. Moors S, Staaks K, Westerhuis M, Dekker L, Verdurmen K, Oei S, et al. Heart rate variability in hypertensive pregnancy disorders: A systematic review. Pregnancy Hypertens 2020 Apr;20:56-68. [doi: 10.1016/j.preghy.2020.03.003] [Medline: 32179490]
39. Fu Q, Levine B. Autonomic circulatory control during pregnancy in humans. Semin Reprod Med 2009 Jul 15;27(4):330-337 [FREE Full text] [doi: 10.1055/s-0029-1225261] [Medline: 19530067]
40. Parrettini S, Caroli A, Torlone E. Nutrition and metabolic adaptations in physiological and complicated pregnancy: focus on obesity and gestational diabetes. Front Endocrinol (Lausanne) 2020 Nov 30;11:611929 [FREE Full text] [doi: 10.3389/fendo.2020.611929] [Medline: 33424775]
41. McCurdy CE, Bishop JM, Williams SM, Grayson BE, Smith MS, Friedman JE, et al. Maternal high-fat diet triggers lipotoxicity in the fetal livers of nonhuman primates. J. Clin. Invest 2009 Jan 19:1. [doi: 10.1172/jci32661]
42. Rakow A, Katz-Salamon M, Ericson M, Edner A, Vanpée M. Decreased heart rate variability in children born with low birth weight. Pediatr Res 2013 Sep 14;74(3):339-343. [doi: 10.1038/pr.2013.97] [Medline: 23770921]
43. Johansson K, Linné Y, Rössner S, Neovius M. Maternal predictors of birthweight: The importance of weight gain during pregnancy. Obes Res Clin Pract 2007 Dec;1(4):223-290. [doi: 10.1016/j.orcp.2007.09.001] [Medline: 24351583]
44. Zöllkau J, Dölker EM, Schmidt A, Schneider U, Hoyer D. Dependencies between maternal and fetal autonomic tone. J Perinat Med 2019 Apr 24;47(3):323-330. [doi: 10.1515/jpm-2018-0221] [Medline: 30676005]
45. Sullivan SD, Umans JG, Ratner R. Hypertension complicating diabetic pregnancies: pathophysiology, management, and controversies. J Clin Hypertens (Greenwich) 2011 Apr;13(4):275-284 [FREE Full text] [doi: 10.1111/j.1751-7176.2011.00440.x] [Medline: 21466626]
46. Wang L, Leng J, Liu H, Zhang S, Wang J, Li W, et al. Association between hypertensive disorders of pregnancy and the risk of postpartum hypertension: a cohort study in women with gestational diabetes. J Hum Hypertens 2017 Nov 29;31(11):725-730 [FREE Full text] [doi: 10.1038/jhh.2017.46] [Medline: 28660887]
47. Kc K, Shakya S, Zhang H. Gestational diabetes mellitus and macrosomia: a literature review. Ann Nutr Metab 2015;66 Suppl 2:14-20 [FREE Full text] [doi: 10.1159/000371628] [Medline: 26045324]
48. Schneider S, Freerksen N, Röhrig S, Hoeft B, Maul H. Gestational diabetes and preeclampsia--similar risk factor profiles? Early Hum Dev 2012 Mar;88(3):179-184. [doi: 10.1016/j.earlhumdev.2011.08.004] [Medline: 21890288]

Abbreviations

ANS: autonomic nervous system

ECG: electrocardiogram
FEV1/FVC: forced expiratory volume at 1 second to forced vital capacity
GDM: gestational diabetes mellitus
HF: high frequency
HRV: heart rate variability
LF: low frequency
NHLBI: National Heart, Lung, and Blood Institute
PECO: Population, Exposure, Comparison or Controls, and Outcome
PNS: parasympathetic nervous system
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RMSSD: root mean square of successive differences between normal heartbeats
SDNN: standard deviation of the normal-to-normal R-R intervals
SNS: sympathetic nervous system
VLF: very low frequency

Edited by A Mavragani; submitted 18.11.22; peer-reviewed by B Schwartz, H Liu, M Shvigulam; comments to author 23.01.23; revised version received 21.03.23; accepted 28.03.23; published 05.06.23.

Please cite as:
Sharifi-Heris Z, Rahmani AM, Axelin A, Rasouli M, Bender M
Heart Rate Variability and Pregnancy Complications: Systematic Review
Interact J Med Res 2023;12:e44430
URL: https://www.i-jmr.org/2023/1/e44430
doi:10.2196/44430
PMID:37276013

Review

Possibilities and Challenges of Delivering Health-Related Small Group Interventions Online: Scoping Review

Katharina Preuhs[1], MSc; Mariska Klein Velderman[1], PhD; Pepijn van Empelen[1], PhD

Netherlands Organization for Applied Scientific Research, Expertise Group Child Health, Leiden, Netherlands

Corresponding Author:
Katharina Preuhs, MSc
Netherlands Organization for Applied Scientific Research
Expertise Group Child Health
PO Box 2215
Leiden, 2301 CE
Netherlands
Phone: 31 615597505
Email: katharina.preuhs@tno.nl

Abstract

Background: The outbreak of the COVID-19 pandemic required the transition of health-related face-to-face group interventions to an online setting. While it seems that group outcomes can be realized in an online setting, less is known about resulting potential challenges (and advantages) and how these can be overcome.

Objective: The aim of this article is to explore what challenges and advantages may arise when providing health-related small group interventions in an online setting and how to overcome these challenges.

Methods: Scopus and Google Scholar databases were searched for relevant literature. Effect studies, meta-analyses, literature reviews, theoretical frameworks, and research reports relating to synchronous, face-to-face, health-related small group interventions, online group interventions, and video teleconferencing group interventions were identified and screened. Findings relating to potential challenges and corresponding strategies are described. In addition, potential advantages of online group settings were explored. Relevant insights were gathered until saturation of results relating to the research questions was reached.

Results: The literature indicated several aspects that require extra attention and preparation in the online group setting. These include the delivery of nonverbal communication and affect regulation, as well as the build-up of group cohesion and therapeutic alliance, which seem more challenging online. Yet there are strategies to overcome these challenges, such as metacommunication, collecting participant feedback, and providing guidance concerning technical accessibility. In addition, the online setting provides opportunities to reinforce group identity, such as by allowing independence and the ability to create homogeneous groups.

Conclusions: While online, health-related small group interventions offer a considerable number of possibilities and benefits compared to face-to-face groups, there are also potential drawbacks to consider, which, if anticipated, can be to a great extent overcome.

(Interact J Med Res 2023;12:e43783) doi:10.2196/43783

KEYWORDS

small groups; online groups; group intervention; group cohesion; therapeutic alliance; scoping review

Introduction

Background

Small group interventions are often used to promote health behavioral changes, psychological well-being, and treatment of mood disorders. These interventions may include group support or therapy in the context of addiction, HIV prevention, and lifestyle support (eg, among people at risk for chronic diseases or in psychotherapy or family therapy) [1-4]. There is good evidence of the effects of small group treatment in the context of mental problems, patient self-management, health promotion, and risk reduction behaviors (eg, [1,5]). Comparison with individual support shows that group interventions yield better results on some occasions, but generally have equivalent effects [1]. Compared to individual interventions, small groups may be more time- and cost-efficient [6]. In addition, in group interventions, members can benefit from the advantages of group processes, such as group comparison, modeling and identification, exchange between group members, practicing new skills, and social support [1,5-11]. As a result, favorable personal changes are found, such as positive psychological states, increased self-efficacy, and improved self-management [7,9,12], as well as unique group outcomes, such as group bond [9,10,13], social support [5,7,10,11,14], collective efficacy [15], and decreased feelings of isolation [7,8]. Previous research has examined which group characteristics and group processes facilitate the effectiveness of health-related small group interventions (eg, [5,11]). For optimal functioning, small group interventions generally seem to have some requirements that need to be met in order to be effective. These concern (1) group processes, referring to the way in which the group operates and exchanges information, and (2) characteristics of the group, referring to the features of the group and group members.

Group Process

For effective group processes, the literature shows the importance of group cohesion and therapeutic alliance. Group cohesion has been found to be an essential condition that contributes to the effectiveness of group outcomes [11,16-18]. Group cohesion refers to group alliance, climate, and the relationships between group members and between group members and the group moderator (ie, therapeutic alliance; see below). A cohesive group contributes to feelings of belongingness and identification [5,7,13,19], trust [10,20,21], personal empowerment, and perceived social support [12]. Correspondingly, programs that establish trust have been shown to have higher retention rates than programs that do not [20,22].

In addition, processes such as self-disclosure of group members and feedback may facilitate changes and behavioral changes. Self-disclosure of group members may encourage the provision and reception of valuable feedback while requesting social validation [5] and strengthening cooperation [11]. Feedback, which can be defined as a reaction to a certain behavior to alter the future execution of that behavior, is often deployed as an intervention itself [23]. Giving and receiving feedback may therefore stimulate personal change through interpersonal influence, as it may enable various psychological effects, such as reinforcement, self-disclosure, reassurance, and affirmation [5,23].

The alliance between the group moderator and group members has also been shown to positively affect the outcomes of small group interventions, although the relationship is more minor than in the context of individual interventions [16]. Therapeutic alliance or working alliance refers to the mutual agreement between group members and the moderator regarding goals, tasks, and the extent to which there is an emotional bond between the moderator and group members. Therapeutic alliance has been shown to have an effect on therapy outcomes, regardless of type of intervention or therapeutic approach, and like-group cohesion has an independent effect on the outcome [24].

The group moderator plays an important role in facilitating such positive group processes and hence outcomes [11,13,14,25,26]. Often, this moderator is a trained psychologist, therapist, or other professional with relevant skills and knowledge to facilitate the group. Facilitation methods that moderators can use include role modeling, psychological education, setting rules of communication by appointing turns, framing, supporting, and initiating themes and activities, with the goals of creating psychological safety and respectful interaction while enabling participants to feel free to share ideas and concerns [27]. Group moderators can emphasize member interaction, create a positive group climate, and handle conflicts immediately upon occurrence to help develop and maintain group cohesion and therapeutic alliance with group members [11,16,17].

Group Characteristics

For group characteristics, the literature points to the relevance of homogeneity, a certain group size, and setting and environment. To start with, homogeneity refers to similarities between group members, which can include age, cultural identity, or, for instance, similar health problems. Group homogeneity has been found to enhance both group cohesion and group identification by establishing a sense of being equal [5,10,28,29] and may reduce experiences of social stigma and lower the threshold to share sensitive information [5,10]. A meta-analytic review by Burlingame and colleagues [30] revealed that groups that were more homogeneous improved more compared to their heterogeneous counterparts.

Next, although there is no particular evidence on the most effective group size, it is generally recommended to hold groups with between 8 to 12 people [6,13,31]. The rationale for this is that while too-small groups may hamper interaction and exchange, too-large groups may undermine the interaction between group members [6].

Furthermore, setting and environment play an essential role in group interventions in general [25]. According to Weinberg [25], managing the setting in which group therapy takes place is an essential element to consider. “Creating a holding environment” may involve a certain choice of furniture, seating order, and placement of a box of tissues for participants. Additionally, calming music in certain areas, such as waiting rooms, may create the impression that the therapist or moderator is taking care of the participants’ needs.

Generally, small health-related group sessions take place in a face-to-face fashion, where people interact in a group in a particular setting. However, since the COVID-19 outbreak, face-to-face group meetings have no longer been able to proceed in their original form. Whereas in many cases these group interventions were postponed, as it was believed that the essential requirements for small group processes could not be met, people also started experimenting with online group meetings, shifting to a digital environment [25,32]. It is likely that these kinds of online interventions remain.

As was shown in a meta-analysis of studies of the effectiveness of internet-based interventions for therapy by Barak and colleagues [7], outcomes of online and face-to-face groups are comparable in terms of effectiveness (eg, [10,25,32,33]). These findings are related to group intervention studies in the context of physical activity–related behavior [14], treatment of anxiety disorders [34,35] or depression [35,36], and the promotion of personal empowerment in online support groups for patients with dental anxiety [7]. These results lead to the conclusion that online and face-to-face groups have comparable effectiveness in different health care contexts and in domains varying from psychosocial effects to treatment outcomes (ie, lowering depressive symptoms) [37]. Yet looking back at crucial group elements, such as homogeneity, group setting and environment, and the establishment of supportive relationships between group members (ie, group cohesion) and with the moderator (ie, therapeutic alliance), shows that the online setting may provide both challenges and opportunities with regard to these group characteristics and group processes that are less well-known.

In this paper, we therefore aim to explore (1) potential challenges that may arise when providing health-related interventions to groups in an online setting, (2) how these challenges can be overcome or avoided, and (3) what possible advantages arise from the online format. We reviewed the literature to address these questions, with a focus on synchronous groups, that is, those in which individuals come together online with the aim to participate in any sort of group activity at the same time, such as to learn, share experiences, change health-related behaviors, and support one another via screen teleconferencing [26]. These synchronous, online, group-based health-related programs are often led by a trained peer or a professional (eg, a psychologist, therapist, or other relevant professional).

Methods

Overview

In this study, we conducted a scoping review; that is, an exploration of a topic that is less well-established in the literature to provide a first overview and potential requirements. In our review, we aimed for a general up-to-date overview of various publications to allow for a comprehensive outline instead of answering a more narrow or specific research question. Where applicable, PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines were followed.

Eligibility Criteria

We included published reports and studies that (1) concerned adult populations taking part in a face-to-face or synchronous online group or support group (this included group therapy and group interventions, thereby excluding educational small group meetings in organizational team-building contexts and online forums or chat groups where individuals do not necessarily come together to engage at the same time), (2) were conducted in a psychological, clinical, or health promotional setting and were led by a moderator (often a psychologist or other professional with relevant expertise), (3) explored effectiveness or provided definitions of group interventions, and (4) reported behavioral outcomes relevant to healthy eating, physical activity, mental health (eg, quality of life), smoking cessation, or health status in any way. Studies published from 2000 to 2021 were eligible. We chose to include studies published within the last 2 decades to ensure relevance to current social, health, and health care climates.

Information Sources

The search took place from February 2021 to June 2021 and was carried out on Scopus and Google Scholar.

Search Strategy

Search terms referred to relevant constructs and included *small groups*, *group intervention*, *support groups*, *online group therapy*, *videoconferencing*, *online*, *face-to-face*, and *web-based*. Furthermore, we included essential core elements in our search, such as *group cohesion*, *therapeutic alliance*, and *essential elements of groups*. Additional articles were further snowballed via relevant articles until saturation of results relating to the research questions was reached.

Results

The search resulted in a large number of relevant publications that could answer the 3 research questions central to this study, including meta-analyses (eg, [7,17]), systematic reviews (eg, [8,20,32,38]), effect studies (eg, [3,35]), publications providing theoretical frameworks and informing the design of online groups (eg, [1,2,10,39,40]), and articles elaborating on past experiences of relevant stakeholders in the context of online groups (eg, [9,25,41]). Relevant research reports were in the contexts of health promotion and disease prevention relating to group support or therapy (eg, [7,25,30]), group processes and dynamics (eg, [1,5]), and specific characteristics of group interventions (eg, [6,10]). In our review, we extracted relevant information from these articles that seemed to partly overlap in outcomes. We recognize that this approach is not reproducible and bears the potential of incompleteness.

Challenges

The identified publications pointed to several challenges that need to be considered when providing small group interventions in an online setting. To start with, in light of the importance of building up group cohesion, one major challenge seems to be the lack of nonverbal communication in an online setting. The absence of body-to-body interaction, with the absence of eye contact being especially relevant and fundamental for group therapy, can be seen as one of the main obstacles in an online setting [9,25,42]. This is due to the fact that an online setting makes it difficult to read and react to both body language and nonverbal signs. Additionally, an online setting hinders affect regulation by the moderator and may make it difficult to express desired messages clearly [41]. The conveyance of nonverbal cues during therapeutic communication is one important part of displaying empathy and affect, reassuring and boosting disclosure, evolving alliance, authenticating care, and ensuring authenticity, among other functions [9]. Recent research implies that online group members do not feel as connected to other group members as in-person group members, which is indicative of lower group cohesion [43].

Another challenge to be tackled is the establishment of a therapeutic alliance [44]. In their review, Gentry et al [32] examined the extent to which therapeutic alliance can be maintained in online group-based treatment. They found that the online setting may result in small decreases in therapeutic alliance.

Some challenges pertain to the setting of the group. In face-to-face groups, the moderator normally controls the setting in which the sessions take place. This includes the arrangement of seats and the environment, the placement of boxes of tissues or plants, and even the choice of music to promote a calm and welcoming experience. These functions, also referred to as “dynamic administration” [45], include the overall setup of the group and handling of the time and space of the meeting, as well as matters concerning boundaries. Since the environment in the online setting is partly dependent on the participants themselves, the moderator requires the participants to prepare a “holding” environment for themselves, such as a quiet place where they feel free to open up and speak.

The online setting also needs to be accounted for, as it offers potential distractions that may not play a role in a face-to-face environment [25]. These include background noises and individuals outside the group setting who are in or enter the same space as the participant and distractions due to the chosen platform [25].

Aside from the elements essential for face-to-face group interventions, online group interventions need to handle technical concerns. More specifically, technical problems such as consistency and speed of the internet connection and overall technical infrastructure, including audio or visual difficulties, delays, dropout, background noise, and poor lighting, may be a limitation [9,25,42]. While offering increased access and recruitment for certain groups, these technical issues can potentially lead to the exclusion of participants who do not have access to a computer, technology, or the internet [9,25,39]. Older participants who are less tech-savvy may therefore be especially at risk of exclusion or be less able to easily access the group [25,39]. The same holds true for individuals who due to their condition may not be able to sit behind a screen for a long period of time [10,39]. Additionally, online groups may be less suitable for participants who are prone to or are currently in acute distress or easily deregulated (such as severely depressed participants or those with suicidal ideation) [9,25]. Reaching out to the aforementioned participants when intervention is needed can be difficult or even impossible, as doing so usually requires more time and attention than the group can provide, especially when it is conducted online [9].

How to Overcome Challenges

Some reports in the literature suggest that participants in online groups experience less group cohesion than face-to-face group participants ([43], see above). Other reports [9,46] suggest that to build up group cohesion and reduce the effect of the absence of bodily interaction and nonverbal cues, moderators can stimulate the presence and input of all participants by directing them to provide input, verbalize what they take from others’ contributions, check in on how they feel, and actively identify mutual understanding (eg, “I see many of you nodding, so it seems like you agree with what Jennifer said”). In order to guarantee adequate and effective metacommunication, the literature [25,29,46] recommends that group moderators consider receiving skills training beforehand. Metacommunication can be defined as “communication about communication” [47], which in the case of the online setting relates to participants verbalizing thoughts and feelings that are evoked by what others say; other participants might otherwise miss these due to the lack of eye contact and body language.

Furthermore, we found that moderators should ensure that time is given to every participant by distributing turns, and that they should acknowledge feelings by verbalizing observations [9,46]. Although group cohesion among online groups might be less strong than in face-to-face groups, the overall convenience of online group sessions seems to outweigh the negative factors of the online setting [43].

Establishment of therapeutic alliance was found to be a second challenge. Based on a literature review and practical experiences, Kneeland et al [46] described five strategies to facilitate therapeutic alliance in a group-based videoconferencing setting: (1) explicitly express gratitude to group members, (2) start the group with an introduction exercise with all attendees and use “ice breaker” questions, (3) self-disclose the group moderator to humanize the face on the screen and build rapport, (4) provide validation, which is the recognition of someone’s feelings and thoughts to underscore that listening is nonjudgmental, and (5) promote rapport between group members and the moderator [48].

As with face-to-face groups, online groups can establish ground rules. These may relate to respectful communication concerning the online setting, such as the use of the camera and microphone during sessions, as well as how to transparently deal with events such as other people entering a participant’s home environment [9,25,29]. To facilitate an online session, a moderator may coordinate activities such as breakout rooms. To ensure the provision of an overall positive experience, Lalande et al [29] recommend that group moderators obtain feedback from group participants during, as well as after, the session.

The literature suggests that is the responsibility of the moderator to consider a digital format that is easily accessible, safe, and convenient for participants to use [25,29]. This comprises the use of an online consent form and making sure the online platform is indeed accessible to users [29].

To overcome technical challenges and ensure that the chosen platform is safe, moderators should run a pilot in order to test whether the session can go through as planned [29,46]. According to Lalande et al [29], participants may benefit from instructions on how to log in and navigate through the digital platform, including tips in case they encounter any problems. These instructions should be offered to participants prior to the start of the session. Furthermore, in order to successfully start the session, the literature advises moderators to make sure that all participants can log in before the session starts and to invite feedback on technological aspects throughout the session [29].

As moderators in online settings are often unable to intervene with participants who require an intervention during imminent emergency situations, Stephen et al [9] recommend that participants be asked to provide contact details (eg, their whereabouts or address and the phone number of an emergency contact) in case of emergency.

Opportunities

While there are certain challenges to overcome in online group settings, there are also opportunities. One of the opportunities of online groups as opposed to face-to-face groups is group size. Commonly used online platforms, such as Zoom or Microsoft

Teams, allow for groups of up to roughly 50 participants to be seen on the screen at once. As it remains to be discussed if such a high number of participants is desirable, limiting online group intervention sessions to 15 participants (as suggested in the literature) therefore seems preferable. However, communication in an online setting may require more effort and moderating strategies, which will be discussed later on.

Furthermore, the online setting provides an opportunity to manipulate group composition due to increased accessibility for individuals who face challenges meeting in person [10,29]. For example, online settings are accessible to people with rare diseases or disabilities, in certain sociodemographic groups, or who are otherwise excluded for reasons such as transportation difficulties, distance, mobility problems, or caregiving responsibilities [7,9,10,39,42,43]. The online setting offers individuals the chance to connect even across the globe, including individuals living in rural areas, thereby enriching demographic diversity while promoting homogeneity of the group [29,43]. Moreover, the possibility to access group interventions online can be time-saving, cost-effective, and convenient due to decreased travel costs and time demands on participants, as they can participate from the comfort of their home [25,39].

While shared characteristics of participants can promote a sense of safety due to decreased stigma and felt recognition, the screen barrier separating participants from each other may be seen as another opportunity, as it may stimulate that feeling even more. The anonymity that can more easily be realized in online groups seems to reduce stigmatization [7,10,25,39,49], power differentials (through neutralizing of status) [9,39], and, consequently, potential inhibition of participants who may otherwise not dare to speak up [9]. As an additional benefit, taboos can be discussed more freely and participants can be encouraged to self-disclose [7,39]. Participants may thus perceive less rejection, which in return promotes honest discussions of feelings and otherwise avoided topics. This seems especially true for male individuals, who have been found to participate more freely in online settings, notably when sensitive topics are discussed, such as suicide and depression [39]. At the same time, group interventions can at times be emotionally overwhelming for participants, which is why the screen barrier may lead to less negative mental impact and defensiveness, as participants may feel sheltered behind their screens [7,9,25]. Socially anxious participants and participants with dissociative symptoms may especially gain from this approach, as they may experience less anxiety and lower their dissociative defenses more easily due to reductions in immediacy and a sense of self-consciousness [25]. Participants with a borderline personality disorder diagnosis may also benefit from the screen barrier due to a greater distance from the therapist, leading to the perception of online groups as being safer [25].

Discussion

Principal Results

In this paper, we elaborate on potential challenges when executing small, synchronous group interventions in an online setting, how these challenges can be overcome or prevented, and what possible opportunities arise from the online format. Essential factors related to small groups include group processes, such as group cohesion, therapeutic alliance, self-disclosure, and feedback, as well as factors relating to the characteristics of the group, such as the size and composition of the group and its setting and environment. From our review of the literature, we conclude that while comparable group outcomes and group processes can be realized in online settings and in offline, face-to-face group settings, both may come with specific benefits and challenges that need to be addressed. On the one hand, challenges include the lack of nonverbal communication, which impacts the establishment of a therapeutic alliance and group cohesion; potential technical concerns; and a lack of suitability for certain participant groups, such as those in acute crisis. Yet the literature suggests measures and strategies to avoid or overcome these pitfalls. Some can be overcome by moderators improving their communication skills (eg, by practicing metacommunication, such as disclosing their own feelings, distributing turns, and recognizing participants' feelings), technical measures (eg, choosing a secure platform, running a pilot test, and providing participants with instructions), and setting the environment (eg, establishing ground rules). While most of the effort to make online group sessions work falls on the moderator, participants themselves can play an active role by ensuring they take the time and make the effort to prepare a holding environment for themselves and the other participants, give feedback to the moderator, and provide the moderator with relevant contact details and their whereabouts in case of an emergency situation.

On the other hand, we encountered some advantages of providing group interventions in an online setting. The advantages include convenience (eg, saving time and being cost-effective), accessibility, and inclusion; the online setting enables individuals to connect with each other who may not otherwise have come together in a face-to-face setting. Furthermore, the screen barrier and higher perception of anonymity may promote participants' sense of safety, potentially leading to a decrease in stigmatization of topics and self-inhibition, thereby encouraging self-disclosure.

Previously, scholars [50,51] have stressed that essential conditions of behavior change methods (eg, stimulating group cohesion) need to be met when translating these methods into practical applications. We hope we have provided some guidelines to intervention designers and practitioners on how the essential conditions of small group interventions can be created in an online context.

Strengths and Limitations

This review was not carried out in a systematic way, nor does it offer a quantitative overview of the effectiveness of online groups. While this means that the reproducibility of the review is low and its completeness cannot be guaranteed, it can offer an up-to-date scoping overview of current knowledge and relevant considerations when transferring or organizing small group interventions in an online setting. Given contemporary developments in overall digitization and changing regulations concerning, among other topics, the COVID-19 pandemic, this discussion seems to be eminently relevant.

Conclusions

The COVID-19 pandemic has emphasized the need to continue group interventions while switching to an online setting [25]. Even though face-to-face groups are starting to return, online groups seem to be the "new normal" in many cases, implying the possibility that more and more group interventions may transfer to an online setting. While online groups offer a considerable number of new possibilities and benefits compared to face-to-face groups (eg, accessibility, the screen barrier, and time effectiveness), there are pitfalls to consider and avoid when setting up an online group (eg, technical concerns and ensuring that emergency contact details are available for participants), as executing online group interventions demands thorough preparation and, in some cases, even extra training in order to maximize effective group outcomes. This includes actions executed before, during, and after the group sessions that relate to group characteristics in terms of the frame and overall setting, technical aspects of the sessions, and group moderation, as well as attention to group processes and participant care. In conclusion, online groups may be a very suitable way to support individuals in groups, not only when face-to-face meetings are difficult or impossible, but even under normal circumstances, given the numerous benefits and possibilities of the online setting.

Acknowledgments

KP was involved with writing the original draft, review and editing, and investigation. MKV was involved with writing the original draft, review and editing, conceptualization, and supervision. PvE was involved with writing the original draft, reviewing and editing, supervision, project administration, funding acquisition, conceptualization, and methodology. We would also like to thank both anonymous reviewers for their constructive comments on the first draft.

Conflicts of Interest

None declared.

References

1. Burlingame GM, Strauss B, Joyce A. Change mechanisms and effectiveness of small group treatments. In: Lambert MJ, editor. Bergin and Garfield's Handbook of Psychotherapy and Behavior Change. Hoboken, NJ: John Wiley & Sons; 2013:640-689.
2. van Empelen P, Kok G, van Kesteren NM, van den Borne B, Bos AE, Schaalma HP. Effective methods to change sex-risk among drug users: a review of psychosocial interventions. Soc Sci Med 2003 Nov;57(9):1593-1608. [doi: 10.1016/s0277-9536(02)00557-9] [Medline: 12948569]
3. Klein Velderman M, Pannebakker F, van Vliet W, Reijneveld S. Prevention of divorce-related problems in Dutch 4- to 8-Year-Olds. Res Soc Work Pract 2016 May 03;28(4):415-427. [doi: 10.1177/1049731516644504]
4. Rijnders M, Jans S, Aalhuizen I, Detmar S, Crone M. Women-centered care: Implementation of CenteringPregnancy® in The Netherlands. Birth 2019 Sep;46(3):450-460. [doi: 10.1111/birt.12413] [Medline: 30592082]
5. Borek AJ, Abraham C. How do small groups promote behaviour change? An integrative conceptual review of explanatory mechanisms. Appl Psychol Health Well Being 2018 Mar;10(1):30-61. [doi: 10.1111/aphw.12120] [Medline: 29446250]
6. Stewart L, Usher A, Allenby K. A review of optimal group size and modularisation or continuous entry format for program delivery. Correctional Service Canada. 2010. URL: https://www.csc-scc.gc.ca/research/005008-0215-01-eng.shtml [accessed 2023-04-25]
7. Barak A, Hen L, Boniel-Nissim M, Shapira N. A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. J Technol Hum Serv 2008 Jul 03;26(2-4):109-160. [doi: 10.1080/15228830802094429]
8. Hoey LM, Ieropoli SC, White VM, Jefford M. Systematic review of peer-support programs for people with cancer. Patient Educ Couns 2008 Mar;70(3):315-337. [doi: 10.1016/j.pec.2007.11.016] [Medline: 18191527]
9. Stephen JE, Christie G, Flood K, Golant M, Rahn M, Rennie H, et al. Facilitating online support groups for cancer patients: the learning experience of psycho-oncology clinicians. Psychooncology 2011 Aug;20(8):832-840. [doi: 10.1002/pon.1791] [Medline: 20677179]
10. Van Uden-Kraan CF, Drossaert CH, Taal E, Smit WM, Bernelot Moens HJ, Van de Laar MA. Determinants of engagement in face-to-face and online patient support groups. J Med Internet Res 2011 Dec 07;13(4):e106 [FREE Full text] [doi: 10.2196/jmir.1718] [Medline: 22155649]
11. Borek AJ, Abraham C, Greaves CJ, Gillison F, Tarrant M, Morgan-Trimmer S, et al. Identifying change processes in group-based health behaviour-change interventions: development of the mechanisms of action in group-based interventions (MAGI) framework. Health Psychol Rev 2019 Sep;13(3):227-247. [doi: 10.1080/17437199.2019.1625282] [Medline: 31190619]
12. Buchanan H, Coulson NS. Accessing dental anxiety online support groups: an exploratory qualitative study of motives and experiences. Patient Educ Couns 2007 Jun;66(3):263-269. [doi: 10.1016/j.pec.2006.12.011] [Medline: 17320336]

13. Ezhumalai S, Muralidhar D, Dhanasekarapandian R, Nikketha B. Group interventions. Indian J Psychiatry 2018 Feb;60(Suppl 4):S514-S521 [FREE Full text] [doi: 10.4103/psychiatry.IndianJPsychiatry_42_18] [Medline: 29540924]
14. Strand M, Eng LS, Gammon D. Combining online and offline peer support groups in community mental health care settings: a qualitative study of service users' experiences. Int J Ment Health Syst 2020;14:39 [FREE Full text] [doi: 10.1186/s13033-020-00370-x] [Medline: 32514303]
15. Bandura A. Exercise of human agency through collective efficacy. Curr Dir Psychol Sci 2016 Jun 22;9(3):75-78. [doi: 10.1111/1467-8721.00064]
16. Alldredge CT, Burlingame GM, Yang C, Rosendahl J. Alliance in group therapy: A meta-analysis. Group Dyn Theory Res Pract 2021 Mar;25(1):13-28. [doi: 10.1037/gdn0000135]
17. Burlingame GM, McClendon DT, Yang C. Cohesion in group therapy: A meta-analysis. Psychotherapy (Chic) 2018 Dec;55(4):384-398. [doi: 10.1037/pst0000173] [Medline: 30335452]
18. Cota AA, Longman RS, Evans CR, Dion KL, Kilik L. Using and misusing factor analysis to explore group cohesion. J Clin Psychol 1995 Mar;51(2):308-316. [doi: 10.1002/1097-4679(199503)51:2<308::aid-jclp2270510223>3.0.co;2-h] [Medline: 7797657]
19. Huber J, Muck T, Maatz P, Keck B, Enders P, Maatouk I, et al. Face-to-face vs. online peer support groups for prostate cancer: A cross-sectional comparison study. J Cancer Surviv 2018 Feb;12(1):1-9. [doi: 10.1007/s11764-017-0633-0] [Medline: 28861845]
20. Sokol R, Fisher E. Peer support for the hardly reached: A systematic review. Am J Public Health 2016 Jul;106(7):e1-e8. [doi: 10.2105/AJPH.2016.303180] [Medline: 27196645]
21. Leimeister JM, Ebner W, Krcmar H. Design, implementation, and evaluation of trust-supporting components in virtual communities for patients. J Manag Inf Syst 2014 Dec 08;21(4):101-131. [doi: 10.1080/07421222.2005.11045825]
22. Bender JL, Katz J, Ferris LE, Jadad AR. What is the role of online support from the perspective of facilitators of face-to-face support groups? A multi-method study of the use of breast cancer online communities. Patient Educ Couns 2013 Dec;93(3):472-479. [doi: 10.1016/j.pec.2013.07.009] [Medline: 23928354]
23. Claiborn CD, Goodyear RK. Feedback in psychotherapy. J Clin Psychol 2005 Feb;61(2):209-217. [doi: 10.1002/jclp.20112] [Medline: 15609354]
24. Khabir L, Mohamadi N, Rahimi C, Dastgheib S. Therapeutic alliance and group cohesion in group therapy based on mentalization and dialectical behavior in borderline personality disorder: A randomized controlled clinical trial. J Health Sci Surveillance Sys 2018;6(4):181-189 [FREE Full text]
25. Weinberg H. Online group psychotherapy: Challenges and possibilities during COVID-19—A practice review. Group Dyn Theory Res Pract 2020 Sep;24(3):201-211 [FREE Full text] [doi: 10.1037/gdn0000140]
26. Gidron Y. Group therapy/intervention. In: Gellman MD, Turner JR, editors. Encyclopedia of Behavioral Medicine. New York, NY: Springer; 2013:880-881.
27. Edmondson AC, Lei Z. Psychological safety: The history, renaissance, and future of an interpersonal construct. Annu Rev Organ Psychol Organ Behav 2014 Mar 21;1(1):23-43 [FREE Full text] [doi: 10.1146/annurev-orgpsych-031413-091305]
28. Cohen S, Gottlieb B, Underwood L. Social relationships and health. In: Cohen S, Underwood L, Gottlieb B, editors. Social Support Measurement And Intervention: A Guide for Health and Social Scientists. New York, NY: Oxford University Press; 2000.
29. Lalande K, Greenman PS, Bouchard K, Johnson SM, Tulloch H. The healing hearts together randomized controlled trial and the COVID-19 pandemic: a tutorial for transitioning from an in-person to a web-based intervention. J Med Internet Res 2021 Apr 06;23(4):e25502 [FREE Full text] [doi: 10.2196/25502] [Medline: 33729984]
30. Burlingame GM, Fuhriman A, Mosier J. The differential effectiveness of group psychotherapy: A meta-analytic perspective. Group Dyn Theory Res Pract 2003 Mar;7(1):3-12. [doi: 10.1037/1089-2699.7.1.3]
31. Biggs K, Hind D, Gossage-Worrall R, Sprange K, White D, Wright J, et al. Challenges in the design, planning and implementation of trials evaluating group interventions. Trials 2020 Jan 29;21(1):116 [FREE Full text] [doi: 10.1186/s13063-019-3807-4] [Medline: 31996259]
32. Gentry MT, Lapid MI, Clark MM, Rummans TA. Evidence for telehealth group-based treatment: A systematic review. J Telemed Telecare 2019 Jul;25(6):327-342. [doi: 10.1177/1357633X18775855] [Medline: 29788807]
33. Arrow K, Yap K, Chester A. Group climate in online group cognitive behaviour therapy predicts treatment outcomes. Clin Psychol 2021 Mar 03;25(2):153-163. [doi: 10.1080/13284207.2020.1829944]
34. Berryhill MB, Culmer N, Williams N, Halli-Tierney A, Betancourt A, Roberts H, et al. Videoconferencing psychotherapy and depression: A systematic review. Telemed J E Health 2019 Jun;25(6):435-446. [doi: 10.1089/tmj.2018.0058] [Medline: 30048211]
35. Douma M, Maurice-Stam H, Gorter B, Krol Y, Verkleij M, Wiltink L, et al. Online psychosocial group intervention for parents: Positive effects on anxiety and depression. J Pediatr Psychol 2021 Feb 19;46(2):123-134 [FREE Full text] [doi: 10.1093/jpepsy/jsaa102] [Medline: 33230541]
36. Berryhill MB, Halli-Tierney A, Culmer N, Williams N, Betancourt A, King M, et al. Videoconferencing psychological therapy and anxiety: a systematic review. Fam Pract 2019 Jan 25;36(1):53-63. [doi: 10.1093/fampra/cmy072] [Medline: 30188992]

37. Khatri N, Marziali E, Tchernikov I, Shepherd N. Comparing telehealth-based and clinic-based group cognitive behavioral therapy for adults with depression and anxiety: a pilot study. Clin Interv Aging 2014;9:765-770 [FREE Full text] [doi: 10.2147/CIA.S57832] [Medline: 24855345]
38. Nguyen M, Bin YS, Campbell A. Comparing online and offline self-disclosure: a systematic review. Cyberpsychol Behav Soc Netw 2012 Feb;15(2):103-111. [doi: 10.1089/cyber.2011.0277] [Medline: 22032794]
39. White M, Dorman SM. Receiving social support online: implications for health education. Health Educ Res 2001 Dec;16(6):693-707. [doi: 10.1093/her/16.6.693] [Medline: 11780708]
40. Introne J, Erickson I, Semaan B, Goggins S. Designing sustainable online support: Examining the effects of design change in 49 online health support communities. J Assoc Inf Sci Tech 2019 May 17;71(4):379-394. [doi: 10.1002/asi.24250]
41. Feijt M, de Kort Y, Bongers I, Bierbooms J, Westerink J, IJsselsteijn W. Mental health care goes online: Practitioners' experiences of providing mental health care during the covid-19 pandemic. Cyberpsychol Behav Soc Netw 2020 Dec;23(12):860-864. [doi: 10.1089/cyber.2020.0370] [Medline: 32815742]
42. Pandey V, Mukherjee S, Jha A. Online patient support groups during COVID-19. Ann Romanian Soc Cell Biol 2021 Aug 23:2021-3328 [FREE Full text]
43. Lopez A, Rothberg B, Reaser E, Schwenk S, Griffin R. Therapeutic groups via video teleconferencing and the impact on group cohesion. Mhealth 2020;6:13 [FREE Full text] [doi: 10.21037/mhealth.2019.11.04] [Medline: 32270005]
44. Bordin ES. The generalizability of the psychoanalytic concept of the working alliance. Psychother Theory Res Pract 1979;16(3):252-260. [doi: 10.1037/h0085885]
45. Foulkes SH. Therapeutic Group Analysis. London, UK: Routledge; 1964.
46. Kneeland ET, Hilton BT, Fitzgerald HE, Castro-Ramirez F, Tester RD, Demers C, et al. Providing cognitive behavioral group therapy via videoconferencing: Lessons learned from a rapid scale-up of telehealth services. Pract Innov 2021 Dec;6(4):221-235 [FREE Full text] [doi: 10.1037/pri0000154]
47. Craig RT. Metacommunication. In: Jensen KB, Craig RT, editors. The International Encyclopedia Of Communication Theory and Philosophy. Hoboken, NJ: John Wiley & Sons; 2016:1-8.
48. Kocabas E, Üstündağ-Budak M. Validation skills in counselling and psychotherapy. Int J Sci Study 2017;5(8):319-322 [FREE Full text]
49. Berger M, Wagner TH, Baker LC. Internet use and stigmatized illness. Soc Sci Med 2005 Oct;61(8):1821-1827. [doi: 10.1016/j.socscimed.2005.03.025] [Medline: 16029778]
50. Kok G, Schaalma H, Ruiter RAC, van Empelen P, Brug J. Intervention mapping: protocol for applying health psychology theory to prevention programmes. J Health Psychol 2004 Jan;9(1):85-98. [doi: 10.1177/1359105304038379] [Medline: 14683571]
51. Kok G, Gottlieb NH, Peters GY, Mullen PD, Parcel GS, Ruiter RA, et al. A taxonomy of behaviour change methods: an Intervention Mapping approach. Health Psychol Rev 2016 Sep;10(3):297-312 [FREE Full text] [doi: 10.1080/17437199.2015.1077155] [Medline: 26262912]

Abbreviations

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews

Edited by A Mavragani; submitted 24.10.22; peer-reviewed by O Borek, LL Gao; comments to author 28.12.22; revised version received 07.03.23; accepted 28.03.23; published 20.06.23.

Please cite as:
Preuhs K, Klein Velderman M, van Empelen P
Possibilities and Challenges of Delivering Health-Related Small Group Interventions Online: Scoping Review
Interact J Med Res 2023;12:e43783
URL: https://www.i-jmr.org/2023/1/e43783
doi:10.2196/43783
PMID:37338971

Review

The Role of Artificial Intelligence Model Documentation in Translational Science: Scoping Review

Tracey A Brereton[1], MSc; Momin M Malik[1], PhD; Mark Lifson[1], PhD; Jason D Greenwood[2], MSc, MD; Kevin J Peterson[1], PhD; Shauna M Overgaard[1], PhD

[1]Center for Digital Health, Mayo Clinic, Rochester, MN, United States

[2]Department of Family Medicine, Mayo Clinic, Rochester, MN, United States

Corresponding Author:
Tracey A Brereton, MSc
Center for Digital Health
Mayo Clinic
200 1st St SW
Rochester, MN, 55905
United States
Phone: 1 (507) 284 2511
Email: brereton.tracey@mayo.edu

Abstract

Background: Despite the touted potential of artificial intelligence (AI) and machine learning (ML) to revolutionize health care, clinical decision support tools, herein referred to as medical modeling software (MMS), have yet to realize the anticipated benefits. One proposed obstacle is the acknowledged gaps in AI translation. These gaps stem partly from the fragmentation of processes and resources to support MMS transparent documentation. Consequently, the absence of transparent reporting hinders the provision of evidence to support the implementation of MMS in clinical practice, thereby serving as a substantial barrier to the successful translation of software from research settings to clinical practice.

Objective: This study aimed to scope the current landscape of AI- and ML-based MMS documentation practices and elucidate the function of documentation in facilitating the translation of ethical and explainable MMS into clinical workflows.

Methods: A scoping review was conducted in accordance with PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines. PubMed was searched using Medical Subject Headings key concepts of *AI*, *ML*, *ethical considerations*, and *explainability* to identify publications detailing AI- and ML-based MMS documentation, in addition to snowball sampling of selected reference lists. To include the possibility of implicit documentation practices not explicitly labeled as such, we did not use *documentation* as a key concept but as an inclusion criterion. A 2-stage screening process (title and abstract screening and full-text review) was conducted by 1 author. A data extraction template was used to record publication-related information; barriers to developing ethical and explainable MMS; available standards, regulations, frameworks, or governance strategies related to documentation; and recommendations for documentation for papers that met the inclusion criteria.

Results: Of the 115 papers retrieved, 21 (18.3%) papers met the requirements for inclusion. Ethics and explainability were investigated in the context of AI- and ML-based MMS documentation and translation. Data detailing the current state and challenges and recommendations for future studies were synthesized. Notable themes defining the current state and challenges that required thorough review included bias, accountability, governance, and explainability. Recommendations identified in the literature to address present barriers call for a proactive evaluation of MMS, multidisciplinary collaboration, adherence to investigation and validation protocols, transparency and traceability requirements, and guiding standards and frameworks that enhance documentation efforts and support the translation of AI- and ML-based MMS.

Conclusions: Resolving barriers to translation is critical for MMS to deliver on expectations, including those barriers identified in this scoping review related to bias, accountability, governance, and explainability. Our findings suggest that transparent strategic documentation, aligning translational science and regulatory science, will support the translation of MMS by coordinating communication and reporting and reducing translational barriers, thereby furthering the adoption of MMS.

(Interact J Med Res 2023;12:e45903) doi:10.2196/45903

KEYWORDS

health; informatics; artificial intelligence; machine learning; documentation; explainability; ethics; translational science; scoping review; medical modeling software; clinical decision support; decision support intervention

Introduction

Background

Artificial intelligence (AI)- and machine learning (ML)-based tools have been hailed as having the potential to revolutionize health care with innovative, efficient, and intuitive approaches to care [1-4]. The successful integration of such tools into clinical settings necessitates a meticulous evaluation conducted by interdisciplinary teams throughout the AI life cycle, ensuring favorable outcomes; however, the promised value of delivering scalable and sustained value for patients has yet to be realized, as the field has recognized a gap in implementation [1,3-7]. *Clinical decision support* (CDS) tools range from computerized alerts and reminders, clinical guidelines, order sets, patient data reports and summaries, documentation templates, diagnostic support, among others, and aim to provide clinicians, staff, and patients with knowledge and person-specific information to support and enhance decision-making in the clinical workflow [8]. The Office of the National Coordinator for Health Information Technology has proposed revising and renaming the CDS criterion in the 2023 Health Data, Technology, and Interoperability: Certification Program Updates, Algorithm Transparency, and Information Sharing Proposed Rule to reflect the array of contemporary and emerging functionalities, data elements, and software applications that aid decision-making in health care and introducing *decision support interventions* [9]. Decision support intervention encompasses "technology that is intended to support decision-making based on algorithms or models that derive relationships from training or example data and then are used to produce an output or outputs related to, but not limited to, prediction, classification, recommendation, evaluation, or analysis" [9]. In this paper, we refer to such tools broadly as *AI- and ML-based medical modeling software* (MMS), as our core concern is with models in medical care based on the AI and ML approach of algorithmic modeling [5] that finds optimal (often noncausal) correlations, rather than traditional statistical theory–based models that try to capture causal processes and underlying mechanisms. This style of modeling introduces novel questions around validation, methodology, communication, coordination, and ethics. We introduce a new term to focus on these specific issues because, although AI- and ML-based MMS may be used for CDS or be implemented in software as a medical device (SaMD), CDS includes systems that are not AI- and ML-based and the software in SaMD may not involve any AI or ML modeling component.

Although progress has been made in AI and ML innovation and many solutions are being developed with high-performance metrics, most software remains within the realm of research rather than real-world settings, and even the most technology-literate academic institutions are not routinely using AI and ML in clinical workflows [1-3,6]. Seneviratne et al [2] asks, "If model performance is so promising, why is there such a chasm between development and deployment?" To recognize the importance of accounting for the complexities of health care delivery throughout the life cycle of MMS production, it is essential to understand what barriers exist and then work to close the implementation gap.

Documentation of MMS may reduce these barriers, but it must first go beyond the assessment of technical performance and involve a holistic, interdisciplinary evaluation process that complements and works in tandem with the software life cycle [1,3,4,7]. Currently, the available documentation frameworks for MMS are fragmented, and there needs to be more guidance spanning all disciplines and stages of development [1-3,7]. Li et al [1] call on the need for a "delivery science" that encompasses a broad set of tools to encourage iterative design thinking among data scientists and clinical informaticists and to promote implementation science techniques across health care operations, ethics, and so on that can be transparently documented. Similarly, the International Telecommunications Union and World Health Organization (WHO) Focus Group on Artificial Intelligence for Health [10] calls for the alignment of 4 pillars—ethics, regulations, technology, and clinical evaluation and use cases—to appropriately evaluate and guide development and ensure the feasibility of a solution to generate sufficient knowledge and evidence to support implementation.

The lack of available, professionally accepted, and ubiquitous references describing appropriate documentation makes it challenging to create evidence supporting the safe and effective translation of MMS from research into clinical practice. To help close the gap, comprehensive and practical documentation processes must be in place to capture critical information about software, incorporating all phases of the software life cycle.

Previous Studies

An initial literature review was conducted to understand the current state of documentation and its impact on the translation of AI- and ML-based MMS into clinical practice. Papers obtained through keyword searches in PubMed were analyzed and synthesized. The search focused on characterizing the extent of existing materials rather than exploring any potential issues that may have been overlooked. We found no consensus on "best practices" for documentation around what we identified as AI- and ML-based MMS. However, we recorded the relevant reporting guidelines offered by government and oversight bodies, ethical principles, and theoretical guiding frameworks. Overlapping principles prioritized explainability, transparency, accountability, and trustworthiness, but descriptions were highly variable throughout the field [11-13]. Despite its potential, the adoption of AI- and ML-based MMS remained fragmented, and there were reports about bias after deployment that put patient safety at risk, providing inaccurate or skewed outcomes and recommendations, propagating inequalities, and introducing group harm [14-17]. The findings were presented to multidisciplinary stakeholders across the MMS life cycle. This workshop highlighted the relevance and urgency for continuing studies regarding the current state and documentation challenges to support the development and translation of AI- and ML-based MMS.

Objectives

The findings from the initial literature review and internal research motivated this scoping review to further evaluate the current state and direction of AI- and ML-based MMS documentation. This study aimed to scope AI- and ML-based MMS documentation practices and define the role of documentation in facilitating the translation of ethical and explainable MMS into clinical workflows.

Methods

Study Design

Covidence (Veritas Health Innovation) [18], a web-based collaboration software platform that streamlines the production of systematic and other literature reviews and developed in accordance with PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) guidelines [19], was leveraged to ensure compliance with scoping review standards and facilitate a systematic process to define eligibility criteria, search the literature, screen results, select evidence for inclusion, and conduct data extraction.

Define the Eligibility Criteria

Key concepts were identified as *AI*, *ML*, *ethical considerations*, and *explainability* to broadly search the literature for evidence of recommendations to support AI- and ML-based MMS documentation practices, scope the development of ethical and explainable software, and define the role of documentation for safe and ethical translation. The use of broad concepts aimed to account for the range of AI- and ML-based MMS and the available recommendations for documentation practices. To include the possibility of implicit documentation practices not explicitly labeled as such, we did not use *documentation* as a key concept but as an inclusion criterion. Keywords and Medical Subject Heading terms were used to support and generate the search query and search constraints including literature found only in the PubMed database; publications dated after 2015; and journals identified to be relevant to the study objectives as defined by the "find journals" functionality of the website, Jane [20]—a website that mines documents in PubMed to find the best matching journals, authors, or papers. The inclusion and exclusion criteria used to assess the publications retrieved are provided in Textbox 1. The search query used for this scoping review, combined using the AND query function, is available in Multimedia Appendix 1.

Textbox 1. Inclusion and exclusion criteria that were used to assess the publications retrieved in the scoping review.

Inclusion criteria

- Paper type—all study designs and publication types
- Language—English
- Setting—artificial intelligence (AI) and machine learning (ML) in health care
- Topic—documentation practices involving AI, ML, ethics, and explainability

Exclusion criteria

- Paper type—none
- Language—non-English
- Setting—non–health care
- Topic—topics other than documentation practices involving AI, ML, ethics, and explainability

Searching and Screening the Literature

Publications were retrieved and imported into Covidence to conduct a 2-stage screening process, including title and abstract screening, followed by a full-text review completed by 1 author. Because this was a scoping review with a relatively small sample size, it was determined that additional coders were not needed. Papers that did not meet the predefined eligibility criteria and those that were not health care related or did not evaluate the current state and direction of AI- and ML-based MMS documentation and translation were excluded.

Data Extraction

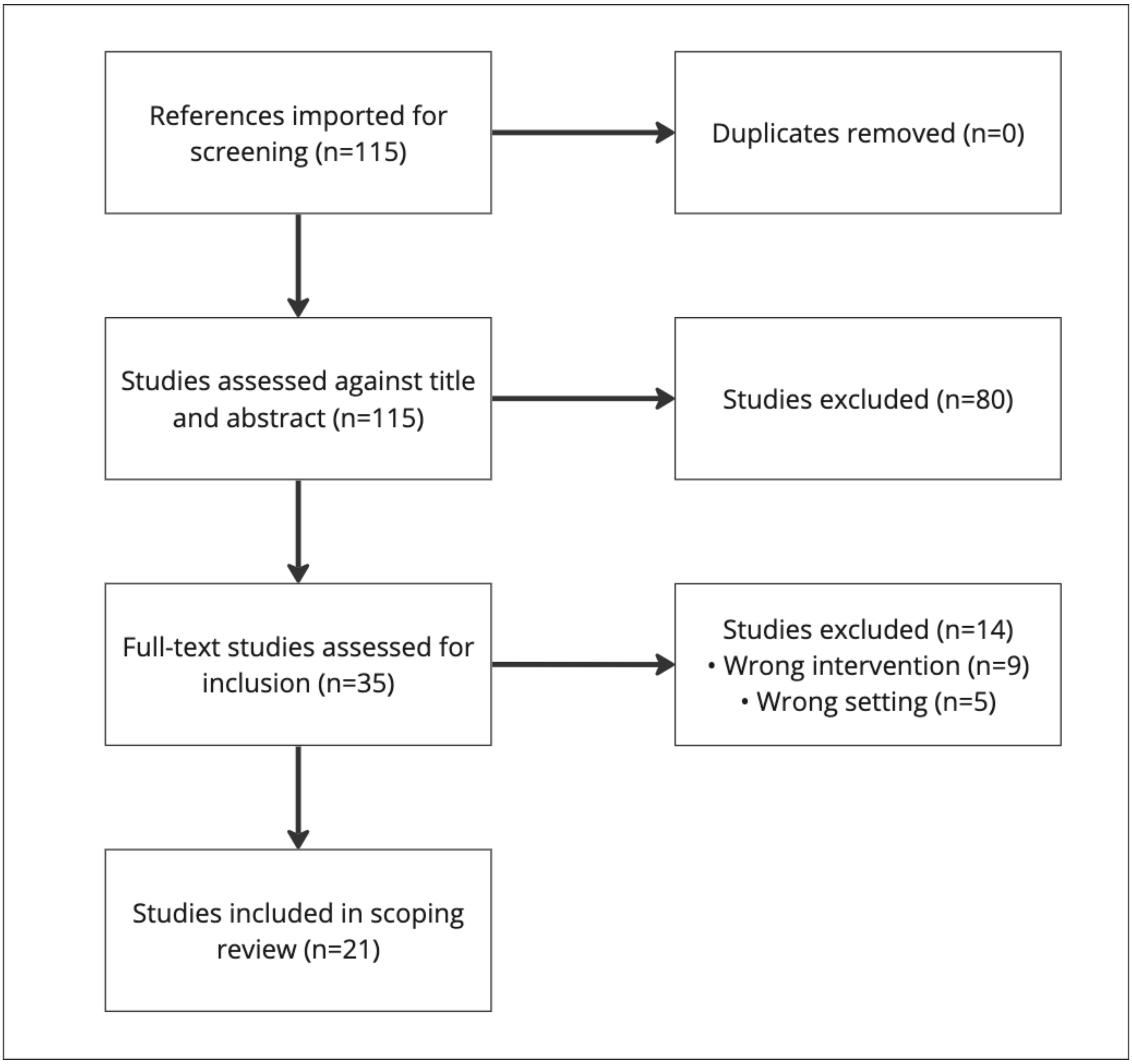
Included papers were analyzed to consolidate evidence of current documentation practices, scope the development of ethical and explainable software, and define the role of documentation in facilitating translation. A data extraction template (Multimedia Appendix 2) was developed and used by 1 author within Covidence to determine the extraction criteria and synthesize the data of the included papers in a consistent format. To effectively consolidate the evidence of the current state and challenges to support the objectives of this study, literature was synthesized by *bias*, *accountability*, *governance*, *explainability*, and *detailed communicated recommendations.*

Results

Characteristics of Included Literature

The initial search in PubMed retrieved 115 papers, which were imported into Covidence for screening and extraction, as shown in the PRISMA-ScR guidelines flow diagram (Figure 1). The PRISMA-ScR checklist is available in Multimedia Appendix 3. Title and abstract screening were conducted, and 29.6% (34/115) of the papers met the eligibility criteria for inclusion. After title and abstract screening, papers were subjected to full-text reviews, where 62% (21/34) of them met the eligibility criteria for inclusion. Of the 21 papers, all (n=21, 100%) included *AI*, *ML*, or *CDS* as keywords, 15 (71%) included *ethics* or *bias* as keywords, and 12 (57%) had *explainability*, *interpretability*, *translation*, *governance*, or *policymaking* as keywords. Relevance was of importance because the current state was being evaluated. Therefore, an analysis of publication year was conducted; of the 21 papers, 3 (14%) were published in 2019, a total of 9 (43%) in 2020, a total of 7 (33%) in 2021, and 2 (10%) in 2022. Additional characteristics of the publications included are available in Multimedia Appendix 4 [13,21-40].

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) flow diagram summarizing the screening process used to conduct the scoping review.



Snowball sampling [41] was used to collect materials such as documents from governmental and regulatory agencies not indexed in PubMed and those referenced in the retrieved documents but not explained with sufficient detail or explained with minimal context. This allowed a more extensive summary of existing frameworks, principles, and standards for ethics (Multimedia Appendix 5 [22-26,28,29,39,40]) and explainability (Multimedia Appendix 6 [22-24,27,28,31,32,34,36-39]) on which existing literature relies on but falls outside of peer-reviewed, indexed literature.

Current State and Challenges

Overview

Existing studies focus on how the translation of AI- and ML-based MMS into patient care settings requires collaboration throughout the AI life cycle and across interdisciplinary specialties of a product team. Furthermore, effective communication and documentation strategies enable collaborative teams to (1) report about developments, (2) inform system progression, and (3) ensure the cross-functional evaluation and documentation of functions from system functionality to ethical considerations. Product teams face challenges involving insufficient collaboration, isolated and fragmented development and documentation processes that fail to prioritize knowledge continuity, timing of interventions, and deficient or unavailable reporting resources needed to progress production, thus disrupting translation into clinical practice [21-23].

The absence of practical tools and best practices to guide translational approaches throughout the AI life cycle contributed to the previously discussed implementation gap. Guidance for ethics and explainability is expected to meet the needs of a broad range of stakeholders, but what constitutes AI ethics and explainability best practices, who determines them, how principles are applied and regulated in practice, and how it is documented still needs to be defined [13,21,22,24,25,42]. Reddy et al [25] noted that “there is little dialogue or recommendations as to how to practically address these concerns in healthcare,” suggesting the need for standardized or mandated regulations to guide processes of collaboration, documentation, and compliance and enhance translational processes with sufficient evidence to support production [22].

Ethics and explainability were investigated in the context of AI- and ML-based MMS documentation and translation. Current challenges involving bias, accountability, governance, and explainability were notable themes requiring thorough review as they pose present barriers to AI- and ML-based MMS translation.

Bias

Differential model performance is a limiting factor in the applicability and deployment of AI- and ML-based MMS owing to the direct implications that biased models have on equity, feasibility of use cases, patient safety, performance, and time and cost of corrections [22,23,26-28]. There is a lack of resources to guide what to do when models perform systematically less well on identifiable patient subgroups and specifically marginalized subgroups, which are often found to be undetected at deployment and identified in ex post reviews [22,23,26,27]. Reddy et al [25] used the adage “biases in, biases out” to describe the influence of bias and how algorithmic decisions and output directly reflect input [25,29]. When efforts are not taken to increase data diversity and quality, then the existing digital divide, social inequalities, and health care disparities are widened; underrepresented and at-risk populations are penalized; vulnerabilities are exacerbated; and harm is increased [28,30,31]. Fragmented processes from which MMS are derived, often from single institutions with single-institution data for training, testing, and validation, may incorrectly assume that the target population mimics the training data population (demographically, socioeconomically, medically, etc). This, in turn, limits external reproducibility owing to generalizability concerns across fields [24]. If created from single-institution data and not externally validated, risks extend beyond appropriate use cases, further jeopardizing patient safety and providing insufficient outputs [24,25,28,31,32]. Kerasidou [33] pointed to the regularity with which inequitable models are deployed, emphasizing the need for evidence regarding prospective evaluation and representativeness relative to training and deployment environments to minimize consequences of patient harm and increase model performance, accuracy, fairness, and trust [30].

Accountability

The introduction of AI solutions into clinical workflows has raised questions about accountability. Does a clinician’s responsibility extend to decisions made from following model outputs, including any harmful or poor outcomes [22,28,33,34]? Must clinicians disclose the level of autonomy or obtain consent from patients regarding the use of MMS in their care [21,25,28,29]? If a clinician’s decision contradicts the AI output, how must a clinician proceed, and who is liable [23,25,32,34]? Navigating accountability about the intent of MMS requires the constructs of safety, responsibility, autonomy, consent, and trust to be deeply considered throughout the AI life cycle [23].

With the application and expectation of MMS aiming to enhance clinician decisions and patient outcomes, proactive solutions to navigate responsibility in response to incorrectness or harm are also imperative. Although the level of autonomy upon which MMS operates varies, when systems fail, there is an expectation of accountability [25,28]. Contradicting opinions between clinicians and tools present unique problems. There exist different approaches to defensive medicine within AI. For instance, when a clinician and model have opposing diagnostic decisions, some clinicians are compelled to follow their instinct and practical experience. In contrast, others are obliged to defer to a model’s logic. Regardless of clinical outcome, patients often remain uninformed as to whether the final decision heeds or disregards an AI recommendation [21].

Responsibility and liability were associated with the usability and trust of MMS within patient-clinician relationships and public support for the integration of the MMS [23,25,30,33]. Concerns surrounding the dehumanization of health care are often present when MMS is deployed, such as the risk of clinician overreliance, disruptions in patient-clinician relationships, and management of false societal expectations of AI [21,28,33]. Overreliance reduces or eliminates patient encounters with clinicians, changing the landscape of patient involvement and potentially missing data that would otherwise be collected outside electronic health records [21,25]. Grote and Berens [21] stated that “facilitating successful collaboration between ML algorithms and clinicians is proving to be a recalcitrant problem that may exacerbate ethical issues in clinical medicine.”

Governance

A gap between ethics and regulation is said to exist with the range of AI- and ML-based MMS classifications [25,30]. For the comparable category of medical device software, there are many international standards and guidance documents for building high-quality, safe, and effective medical device software, but there is no explicit definition of ethical criteria; however, the impact of the regulatory process accounts for many of the identified ethical considerations. Various ideologies, principles, and frameworks try to define what ethical AI is, including the AI Ethics Principles, Ethically Aligned Design principles developed by the Institute of Electrical and Electronics Engineers (IEEE) and Trustworthy AI principles defined by the High-level Expert Group on Artificial Intelligence, but best practices are found to be scattered [26,32,38]. Software developers desire increased guidance, resources, and reporting guidelines to develop and deploy ethically designed MMS fields [22,34]. Clinicians, one of the primary end users of MMS, communicate the need for documentation resources to serve as training material for the intended use and a means of communicating the objective, functionality, and limitations of MMS to ensure the appropriate and actionable translation of software into practice [21,22]. In addition to ethical considerations, although adequate information or recommendations, privacy protection, data integrity, and regulatory frameworks may be facilitated by government oversight, evidence suggests that AI- and ML-based MMS are not deployed with sufficient levels of documentation to support explainability [30]. With such leadership and resources allocated to MMS, institutions can avoid inconclusive, inscrutable, and misguided evidence; unfair outcomes; and the lack of traceability and transformative effectiveness [32,34]. Subjecting researchers to self-monitoring of ethical conduct and determining the level of explainability to accompany MMS is of substantial concern but may be combated with increased reporting rigor and guardrails for evaluation and reporting of evidence [13,22,25,26,30].

Although regulators guide how to deploy safe and effective medical device software, including requiring robust clinical validations and postdeployment monitoring and surveillance programs, these processes may be out of the scope of requirements for some AI- and ML-based MMS, including those with limited or no regulation requirements by the Food and Drug Administration (FDA). Questions such as "How much accuracy [or other relevant performance metric] is sufficient for deployment?" "What level of transparency is required?" and "Do we understand when the model outputs are likely to be unreliable and therefore should not be trusted?" have made defining expectations and requirements for governance difficult, especially for varying levels of system autonomy and risk [13,21,24,25,35]. Although frameworks and principles have been developed by various organizations with goals of increasing explainability (FDA; European Union General Data Protection Regulation; WHO; IEEE; and Findability, Accessibility, Interoperability, and Reusability principles), no recognition of established best practices were identified, only recommended principles to abide by, such as those defined within the AI Ethics Principles [23,24,31,32,35,36]. Reported consequences of fragmented documentation resources and reporting requirements include the lack of explainability, accountability, validation, transparency, and trust [22,25,28,29].

Explainability

Although the field points to explainability to help facilitate successful documentation and translation, the literature acknowledges difficulties with such resources and best practices that are expected to contribute to achieving explainable AI- and ML-based MMS [13,22,36,37]. Although some have argued for distinguishing interpretability (inherently interpretable models) from explainability (post hoc, simplified summaries of model functionality that supplement the model) [43], within the scoped literature, the terms interpretation or interpretability and explanation or explainability were used interchangeably. Since the overall use is closer to the proposed meaning of *explainability*, we exclusively use this term. Explainability, then, defined as understandable (post hoc and supplementary) explanations of ML model outcomes, accounts for how users should be able to understand the logic of ML modeling to implement it appropriately within clinical workflows [13,27,38]. Establishing explainability is functionally complex without supplemental documentation, creating challenges in understanding decision logic, deploying transparent tools, and defining accountability and responsibility (especially with requirements varying with different risk classifications) and threatening patient safety and trust [13,21,25,27,37].

The complex nature of AI- and ML-based MMS, comprising multifaceted computations that drive decision-aiding output, complicates explainability and initiates debate regarding the prioritization of performance versus explainability in system development [13,25]. "Black boxes" have been found to make the clinical application and decision procedures "notoriously hard to interpret and explain in detail," limiting the ability to identify and document technical and logical justifications for decisions and conflicting with core values of patient consent and awareness about the role AI in their care [13,25,32,36,38]. Kerasidou [33] stated that with black box systems, "the 'thinking process' by which outcomes are produced is not obvious to those who use the AI or even to those who develop it," raising explainability, transparency, and justification concerns from developers to clinicians and from clinicians to patients [13,32,44]. Amann et al [13] questioned whether, owing to their complexity, black boxes are even documentable. Without a way to document and report about explanations of AI, it is "hard to determine if differences in diagnoses reflect diagnostically relevant differences between patients or if they are instances of bias or diagnostic errors and over-/underdiagnosis," further emphasizing the need to mitigate and address biases before deployment [36]. Once clinicians can no longer comprehend decisions fully, they cannot explain to the patient how specific outcomes or recommendations were derived, thus affecting patient safety, trust, and care plans [13,27]. The literature describes the necessity to interpret MMS logic because omitting it has been found to "pose a threat to core ethical values in medicine and may have detrimental consequences for individual and public health," including evidence of disregard for ethical and regulatory practices, unsuitable clinical application, and making it impossible to investigate and rectify causes of errors; however, Yoon et al [27] argued that the extent to which explainability is required still needs to be determined [13,27,33,38].

Beyond an understanding of system complexity, explanations of system functionality were found to be critical for developers, stakeholders, clinicians, and patients to understand a system about clinical applicability and intended use, despite varying levels of familiarity with AI [38,39]. The context of appropriate clinical application is critical to disclose, owing to the impact of training data on system performance and deployment environment, but developing explainable models that satisfy requirements of providing supporting information for clinical decision-making proves to be challenging, given that clinical decisions are made based on different modalities and reasoning strategies [21,22,35]. When explanations detect, analyze, and assess artifacts of MMS throughout design, development, and implementation, the field may anticipate more informed and trustworthy adoption [21,25,37]. Therefore, the complex nature of clinical decisions is said to call for the transparent traceability of the logic leading to output and how clinicians interact with what they understand from explanations, thus requiring guidance as to what constitutes satisfactory explanations of decisions and how such resources should be documented [21,22,26,32]. In the United States, for products that do not fall under the scope of FDA regulation, reporting guidelines (TRIPOD [Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis], CONSORT [Consolidated Standards of Reporting Trials], IEEE, or STROBE [Strengthening the Reporting of Observational Studies in Epidemiology]) that support the transparent documentation of software development and validation exist but are inconsistently adopted and do not show evidence of driving analysis and documentation throughout all areas of product development, evaluation, deployment, and postdeployment [25,28,29,39].

Discussion

Principal Findings

With an understanding of the current state and challenges of AI- and ML-based documentation, addressing present barriers, including those identified in this scoping review as involving bias, accountability, governance, and explainability, is required to enhance documentation efforts and promote the translation of MMS. Recommendations identified in the literature call for a proactive evaluation of MMS, multidisciplinary collaboration, adherence to investigation and validation protocols, transparency and traceability requirements, and guiding standards and frameworks that enable innovation across translational aspects and support MMS throughout the software life cycle.

Documentation serving to proactively outline and guide translational processes and encourage multidisciplinary collaboration is recommended to support system development throughout the AI life cycle to promote patient safety, support appropriate clinical use cases, and ensure that essential testing and validation processes are completed before deployment [22-24,28]. Proactively accounting for translation reflects ex ante (as opposed to ex post) regulation that is "pre-emptive of foreseeable risks and has a more open and participatory character" [25,32]. To help facilitate the proactive evaluation of MMS, researchers such as Wiens et al [35] and Allen et al [24] suggest the need for a road map for deploying AI- and ML-based tools that consist of a stepwise framework that introduces methods and tools for development, testing, validation, and monitoring from the beginning of production (problem identification and idea formulation) to the end (widespread deployment and maintenance).

Such a road map is said to be enhanced by multidisciplinary collaboration, promoting robust partnerships among stakeholders and project teams through iterative and joint approaches such as participatory design [25,35]. Within many of the proposed and available road maps, frameworks, and recommendations reviewed in the scoping review, the engagement of stakeholders is encouraged to occur early in the process to ensure the development of an optimal solution; this includes helping to determine clinical relevance, identifying appropriate data and collaborators, considering ethical implications and engaging with ethicists, rigorous evaluations and reporting of predictions and model code, organizing clinical trials and safety monitoring of strategies, and market deployment approaches, all of which ought to be documented and transparent to the interdisciplinary team throughout the product's life cycle [24,35,40].

Communicating evidence of adherence to investigation and validation protocols was said to support MMS implementation, as it provides an objective measure of testing efforts and is especially important for instances of uncertainty and conflict between AI and clinicians to bolster traceability for liability and legal obligations and mitigate ethical concerns [13,21,25,26]. Providers can use the documentation of adherence to such protocols in practice, as provider explanations to end users are said to mitigate concerns and threats of overreliance on MMS through the ability to communicate and document the training for decision logic, system output, and role definition in the final decisions made for patients [13,21]. Leveraging the evidence of investigation and validation is critical for ensuring representativeness and appropriate clinical application to minimize the consequence of patient harm and inequitable model performance. In addition, by using transparent documentation, Ploug and Holm [36] also propose "contestability by design," an approach that focuses on system development and optimization and provides "design principles for algorithmic systems that will enable professionals and expert users to challenge the reasoning of these systems in an ongoing process" in a way that does not come at the cost of system performance and supports the introduction of a minimal set of criteria to serve as a practical guideline.

Documenting such efforts fulfills the recommendations for promoting transparency and traceability requirements suggested to support the explainability of MMS. Explanations, referred to as a "cardinal responsibility of medical practitioners," function as an additional safeguard to ensure the reliability of a system's reasoning process, whereas "trust" in clinical decisions correlates with the construct of explainability at all levels of expertise [21,27]. The complexity of reasoning underlying explanations highlights the need to address the varying definitions of explainability (and interpretability) and the extent to which they are required to effectively support MMS and manage the inconsistency and vagueness surrounding bias and the implications of equity and patient safety [21,27]. These requirements span beyond the scope of 1 team within a project, making explainability a necessity to be adopted and prioritized from a multidisciplinary perspective in a way that promotes knowledge continuity across disciplines; provides relevant explanations of MMS logic for both technical and translational purposes; and provides evidence of MMS traceability [13,21,22,24,27,39].

Investigation and validation protocols and transparency and traceability requirements were also recommended to be aided by guiding standards and frameworks, from reporting guidelines developed by research groups to regulatory protocols published by governance bodies [13,21,23,24,29,37]. Although a consensus for universal best practices was not recognized in the literature, the importance of leveraging guiding standards and frameworks was consistently identified, from proposed reporting guidelines published by research teams to legislation defined by governing bodies [22,23]. Available and proposed standards, frameworks, and governance structures to promote ethical and explainable AI- and ML-based MMS documentation and translation mentioned in this scoping review's literature were recorded. They are available in Multimedia Appendix 5 (ethics) and Multimedia Appendix 6 (explainability).

Integrating recommendations related to proactive evaluation; multidisciplinary collaboration; and adherence to investigation and validation protocols, transparency and traceability requirements, and guiding standards and frameworks are expected to enhance documentation efforts by increasing the transparent reporting of scientific evidence, promoting knowledge continuity across disciplines, providing guidance throughout the AI life cycle, and improving usability in clinical practice while also addressing the implementation gap.

Limitations

Limitations of this study include the restriction of the search to only 1 database, PubMed, to conduct the scoping review. This limitation likely restricted the available and proposed governance standards, guidelines, and frameworks. Notably, additional resources found in our previous studies were not explicitly mentioned in the literature, for example, Google's Model Card for model reporting [45], Model Facts Labels [46], DECIDE-AI (Developmental and Exploratory Clinical Investigations of Decision support systems driven by Artificial Intelligence) [47], and AI Factsheets [48]. Therefore, continuing this investigation as other resources are available (eg, Coalition for Health AI guardrails [49] and National Institute of Standards and Technology recommendations [50]) is critical. In addition, owing to the exploratory nature of scoping reviews, this study aimed to assess the available academic literature, organize findings within themes, and highlight present gaps regarding AI- and ML-based MMS documentation and translation practices rather than explore a more defined research question. This study also is limited to scoping the concerns in existing literature; the extent to which these concerns may be incomplete, or misguided, is not within our scope, for example around the question of when AI and ML may be not just inappropriate but illegitimate [51], or the danger that practitioners mistakenly think that explainable models reflect causality in the world [52-55] and therefore base their trust and decision-making on a fundamental error.

Implications for Practice and Future Development

This study identifies a need for proactive evaluation, standards, frameworks, and transparency and traceability requirements to enhance documentation efforts and promote the translation of MMS. However, how might practitioners achieve these goals? What steps can be taken to improve documentation efforts and promote transparency and traceability, and how might they be implemented into practice?

As indicated by the findings in this scoping review, although there are existing standards, guidelines, frameworks, and governance structures to guide the documentation of AI- and ML-based MMS, there is acknowledgment that these are insufficient and that additional resources are needed to provide appropriate guidance related to ethics and explainability, promote safety and efficacy, provide support throughout the AI life cycle, and reduce present barriers to translation. The available but fragmented and phase-specific resources create an opportunity either to streamline and merge complementary standards, guidelines, frameworks, and governance structures or to encourage the development of new resources. Recommendations highlighted from the literature may help inform the development of such new resources to operationalize the theoretical into actionable tools that answer questions such as, How do I know if the implementation of an MMS tool is the right solution? What agreed-upon principles can be leveraged to analyze data, build tools, and guide implementation? What level of evidence and types of study designs will be expected to assess model effectiveness in terms of validity, acceptability, fairness, equity, transparency, and health impact? How can external validation of models be facilitated? What processes are required to effectively document and navigate the regulatory and quality assurance pathway? and What constitutes effective postdeployment monitoring?

The proactive and standardized evaluation and documentation of MMS through a multidisciplinary collaboration of accountable contributors may contribute to the operationalization of guidelines. By promoting enterprise adherence to investigation and validation protocols, clinician judgment (eg, model output thresholding), and traceability requirements, such an approach aligns evidence-based best practices. Conceptually, the central testing, documentation, and multistakeholder coordination applied to patient's longitudinal electronic health record is applied to the AI model throughout its life cycle.

Such a unified model document would represent a framework for easy discovery of the critical scientific, ethical, and regulatory requirements of an AI- and ML-based MMS while allowing for future expandability or customization to fit the needs of each specialty area and possible future requirements. This unified model document could be shared across all levels of model development, from the ideate phase to postdeployment monitoring. Ideally, the unified model document would capture the fundamental features of any AI- and ML-based MMS while guiding accountable stakeholders on how to address issues. For example, if an MMS developer answers a question about explainability as a "black box," it could provide some basic recommendations for expanding explainability. The document could be created and translated to a standard markup language such as XML or JSON. This would allow it to have a formal structure for the documentation but allow for consumption by others for implementation (eg, Epic and Cerner). Moreover, large institutions that routinely use these tools in clinical practice (eg, health care centers) could adopt a risk-based approach, creating a centralized team of experts to standardize the categorization of all AI- and ML-based MMS tools and determine the type of control measures, evidence, and overall rigor required for their safe and effective development, deployment, and monitoring.

In addition to the unified model document, a possible direction of the industry is to follow others who develop tools and architectures that follow a DevOps framework for ML called Machine Learning Operations. The first step could be at the national level, through regulatory bodies (the Office of the National Coordinator for Health Information Technology, FDA, IEEE, etc), medical academies (National Academy of Medicine, American Medical Informatics Association, American Medical Association, American College of Obstetricians and Gynecologists, American Academy of Family Physicians, American College of Surgeons, etc), or industry and academic collaboration (Google, Microsoft, Amazon, Epic, etc) to develop a set of practice standards, similar to other regulations such as SaMD, which are adopted across the AI- and ML-based MMS industry and require specific documentation and disclosure of MMS technologies that create consistency throughout development to the release to the public. In addition, organizations could create enterprise AI- and ML-based MMS translational boards comprising various experts in the fields of AI and ML, including data science or engineering, IT, ethics, electronic health records, nursing or clinical or translational or pharmacy informatics, and clinical expertise. Similar to an institutional review board found at any research institution, the AI and ML translational board would help address gaps in ethics or bias, explainability, and efficacy and reduce the translational barriers. Finally, institutions and medical colleges could require training and regular refresher courses for clinicians about AI- and ML-based MMS tools that would follow up-to-date standards of practice with many other medical devices (Clinical Laboratory Improvement Amendments, Long-Acting Reversible Contraception training, Basic Life Support and Advanced Cardiac Life Support, etc).

As health centers are venturing into the development of SaMD's internal deployment and commercialization, an enterprise-wide system to enable guided development without stifling innovation may prove to be valuable. To achieve this goal, the transparency and traceability of documentation should align with requirements for FDA submission and Centers for Medicare and Medicaid Services and Joint Commission compliance while leveraging the best practices of technical stakeholders and internal oversight. Such a coordinated effort could be implemented in practice by deploying an enterprise quality management system for AI development, translation, and continuous monitoring, such that internal and external evaluation reports become standardized artifacts for inclusion in MMS documentation. Stakeholders across the AI life cycle would contribute to this work, each considered accountable for subject matter expertise and their action in the quality management system. Examples of stakeholders may include data scientists, informaticists, software engineers, translational and implementation scientists, educators, and providers. Each stakeholder can enhance transparency by increasing their awareness about risk in their domain, and dependencies associated with their action and documentation. For example, a data scientist's selection of clinical features will have substantial implications for the application of the model. This information, if transparently documented, will serve clinical stakeholders, informaticians, and translational scientists when discussing model performance, tuning, and deployment. Further down the AI translation process, the implementation scientist's workflow evaluation and documentation of model insertion points will be leveraged to facilitate discussions between data scientists and clinicians to appropriately threshold and present model output information based on the documented considerations of model risk and performance. Promoting traceability and transparency for national and international reportable standards is in process. For example, there are ongoing efforts to draw consensus around current best practices and recommendations for standardized AI evaluation and governance in health care; examples include the Coalition for Health AI [49], Health AI Partnership [56], and WHO and International Telecommunications Union Focus Group on Artificial Intelligence for Health [10]. Meanwhile, the National Academy of Medicine is developing an AI Code of Conduct [57], which may be an opportunity to define each stakeholder category and clarify the determination of distributed accountability and responsibility. Academic health centers have begun forming governance bodies that enforce evaluation and checkpoints, with accompanying AI- and ML-based MMS tool artifacts. If aggregated across the AI life cycle, such artifacts could be the foundation for MMS documentation [58] and evolve with regulations. In their development, oversight bodies may leverage published AI translation evaluation frameworks and join or learn from ongoing studies in this area that is actively developing detailed guidelines.

Conclusions

To know whether promises about how the adoption of AI- and ML-based MMS tools has the power to revolutionize health care are valid, and to then achieve such benefits, requires strategic translation that prioritizes ethical considerations, the ability to provide explanations that transparently communicates how a decision is reached, and disclosures on the intended use that are accessible to multidisciplinary perspectives across experts and nonexperts. The ability of MMS to deliver on expectations depends on resolving translation barriers, including those related to bias, accountability, governance, and explainability. Our findings suggest that aligning translational and regulatory science through strategic documentation developed to promote proactive evaluation, multidisciplinary collaboration, investigation and validation protocols, transparency and traceability requirements, and guiding standards and frameworks will support the translation and adoption of MMS. Further, we propose that leveraging such transparent documentation processes through a quality management system may support enterprise coordination toward the development of an SaMD regulatory framework.

Data Availability

All data generated or analyzed during this study are included in this published paper and the Multimedia Appendices 1-6.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Key concepts and search limitations used to generate the scoping review query, including the keywords and Medical Subject Headings (MeSH) terms and date and journal constraints.

[PDF File (Adobe PDF File), 133 KB - ijmr_v12i1e45903_app1.pdf]

Multimedia Appendix 2

Data extraction template developed on Covidence to standardize information extracted from studies to satisfy scoping review objectives.

[PDF File (Adobe PDF File), 11 KB - ijmr_v12i1e45903_app2.pdf]

Multimedia Appendix 3

PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.

[PDF File (Adobe PDF File), 97 KB - ijmr_v12i1e45903_app3.pdf]

Multimedia Appendix 4

Overview of publications included in the scoping review (N=21).

[PDF File (Adobe PDF File), 93 KB - ijmr_v12i1e45903_app4.pdf]

Multimedia Appendix 5

Available and proposed standards, guidelines, frameworks, and governance structures to promote ethical artificial intelligence– and machine learning–based medical modeling software documentation and translation mentioned throughout literature.

[PDF File (Adobe PDF File), 240 KB - ijmr_v12i1e45903_app5.pdf]

Multimedia Appendix 6

Available and proposed standards, legislation, organizations, guidelines, frameworks, and governance bodies to promote explainable artificial intelligence– and machine learning–based medical modeling software tool documentation and translation mentioned throughout literature.

[PDF File (Adobe PDF File), 192 KB - ijmr_v12i1e45903_app6.pdf]

References

1. Li RC, Asch SM, Shah NH. Developing a delivery science for artificial intelligence in healthcare. NPJ Digit Med 2020 Aug 21;3:107 [FREE Full text] [doi: 10.1038/s41746-020-00318-y] [Medline: 32885053]
2. Seneviratne MG, Shah NH, Chu L. Bridging the implementation gap of machine learning in healthcare. BMJ Innov 2019 Dec 20;6(2):45-47 [FREE Full text] [doi: 10.1136/bmjinnov-2019-000359]
3. Gama F, Tyskbo D, Nygren J, Barlow J, Reed J, Svedberg P. Implementation frameworks for artificial intelligence translation into health care practice: scoping review. J Med Internet Res 2022 Jan 27;24(1):e32215 [FREE Full text] [doi: 10.2196/32215] [Medline: 35084349]
4. Lovis C. Unlocking the power of artificial intelligence and big data in medicine. J Med Internet Res 2019 Nov 08;21(11):e16607 [FREE Full text] [doi: 10.2196/16607] [Medline: 31702565]
5. Breiman L. Statistical modeling: the two cultures. Statist Sci 2001 Aug;16(3):199-231 [FREE Full text] [doi: 10.1214/ss/1009213726]
6. Kulkarni V, Gawali M, Kharat A. Key technology considerations in developing and deploying machine learning models in clinical radiology practice. JMIR Med Inform 2021 Sep 09;9(9):e28776 [FREE Full text] [doi: 10.2196/28776] [Medline: 34499049]
7. Reddy S, Rogers W, Makinen VP, Coiera E, Brown P, Wenzel M, et al. Evaluation framework to guide implementation of AI systems into healthcare settings. BMJ Health Care Inform 2021 Oct;28(1):e100444 [FREE Full text] [doi: 10.1136/bmjhci-2021-100444] [Medline: 34642177]
8. Clinical decision support. The Office of the National Coordinator for Health Information Technology. URL: https://www.healthit.gov/topic/safety/clinical-decision-support [accessed 2023-06-01]

9. Marchesini K, Smith J, Everson J. Increasing the transparency and trustworthiness of AI in health care. HealthITBuzz. 2023 Apr 13. URL: https://www.healthit.gov/buzz-blog/health-innovation/transparent-and-trustworthy-ai-in-health-care [accessed 2023-06-28]
10. FG-AI4H deliverables overview. International Telecommunications Union (ITU). URL: https://www.itu.int/en/ITU-T/focusgroups/ai4h/Pages/deliverables.aspx [accessed 2023-05-01]
11. Pesapane F, Volonté C, Codari M, Sardanelli F. Artificial intelligence as a medical device in radiology: ethical and regulatory issues in Europe and the United States. Insights Imaging 2018 Oct;9(5):745-753 [FREE Full text] [doi: 10.1007/s13244-018-0645-y] [Medline: 30112675]
12. Rudin C, Chen C, Chen Z, Huang H, Semenova L, Zhong C. Interpretable machine learning: fundamental principles and 10 grand challenges. Statist Surv 2022;16:1-85 [FREE Full text] [doi: 10.1214/21-ss133]
13. Amann J, Blasimme A, Vayena E, Frey D, Madai VI, Precise4Q consortium. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. BMC Med Inform Decis Mak 2020 Nov 30;20(1):310 [FREE Full text] [doi: 10.1186/s12911-020-01332-6] [Medline: 33256715]
14. Mullainathan S, Obermeyer Z. Does machine learning automate moral hazard and error? Am Econ Rev 2017 May;107(5):476-480 [FREE Full text] [doi: 10.1257/aer.p20171084] [Medline: 28781376]
15. Jayatilake SM, Ganegoda GU. Involvement of machine learning tools in healthcare decision making. J Healthc Eng 2021 Jan 27;2021:6679512 [FREE Full text] [doi: 10.1155/2021/6679512] [Medline: 33575021]
16. DeCamp M, Lindvall C. Latent bias and the implementation of artificial intelligence in medicine. J Am Med Inform Assoc 2020 Dec 09;27(12):2020-2023 [FREE Full text] [doi: 10.1093/jamia/ocaa094] [Medline: 32574353]
17. Lysaght T, Lim HY, Xafis V, Ngiam KY. AI-assisted decision-making in healthcare: the application of an ethics framework for big data in health and research. Asian Bioeth Rev 2019 Sep 12;11(3):299-314 [FREE Full text] [doi: 10.1007/s41649-019-00096-0] [Medline: 33717318]
18. Mellor L. The difference between a systematic review and scoping review. Covidence. URL: https://www.covidence.org/blog/the-difference-between-a-systematic-review-and-a-scoping-review/ [accessed 2022-03-01]
19. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
20. Schuemie M. Welcome to Jane. Journal/Author Name Editor. 2007. URL: https://jane.biosemantics.org/ [accessed 2022-03-01]
21. Grote T, Berens P. How competitors become collaborators-bridging the gap(s) between machine learning algorithms and clinicians. Bioethics 2022 Feb;36(2):134-142. [doi: 10.1111/bioe.12957] [Medline: 34599834]
22. Morley J, Floridi L, Kinsey L, Elhalal A. From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices. Sci Eng Ethics 2020 Aug;26(4):2141-2168 [FREE Full text] [doi: 10.1007/s11948-019-00165-5] [Medline: 31828533]
23. Liaw ST, Liyanage H, Kuziemsky C, Terry AL, Schreiber R, Jonnagaddala J, et al. Ethical use of electronic health record data and artificial intelligence: recommendations of the primary care informatics working group of the international medical informatics association. Yearb Med Inform 2020 Aug;29(1):51-57 [FREE Full text] [doi: 10.1055/s-0040-1701980] [Medline: 32303098]
24. Allen Jr B, Seltzer SE, Langlotz CP, Dreyer KP, Summers RM, Petrick N, et al. A road map for translational research on artificial intelligence in medical imaging: from the 2018 national institutes of Health/RSNA/ACR/the academy workshop. J Am Coll Radiol 2019 Sep;16(9 Pt A):1179-1189. [doi: 10.1016/j.jacr.2019.04.014] [Medline: 31151893]
25. Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. J Am Med Inform Assoc 2020 Mar 01;27(3):491-497 [FREE Full text] [doi: 10.1093/jamia/ocz192] [Medline: 31682262]
26. Samuel G, Derrick G. Defining ethical standards for the application of digital tools to population health research. Bull World Health Organ 2020 Apr 01;98(4):239-244 [FREE Full text] [doi: 10.2471/BLT.19.237370] [Medline: 32284646]
27. Yoon CH, Torrance R, Scheinerman N. Machine learning in medicine: should the pursuit of enhanced interpretability be abandoned? J Med Ethics 2022 Sep;48(9):581-585 [FREE Full text] [doi: 10.1136/medethics-2020-107102] [Medline: 34006600]
28. Mökander J, Morley J, Taddeo M, Floridi L. Ethics-based auditing of automated decision-making systems: nature, scope, and limitations. Sci Eng Ethics 2021 Jul 06;27(4):44 [FREE Full text] [doi: 10.1007/s11948-021-00319-4] [Medline: 34231029]
29. Landau AY, Ferrarello S, Blanchard A, Cato K, Atkins N, Salazar S, et al. Developing machine learning-based models to help identify child abuse and neglect: key ethical challenges and recommended solutions. J Am Med Inform Assoc 2022 Jan 29;29(3):576-580 [FREE Full text] [doi: 10.1093/jamia/ocab286] [Medline: 35024859]
30. Zhao IY, Ma YX, Yu MW, Liu J, Dong WN, Pang Q, et al. Ethics, integrity, and retributions of digital detection surveillance systems for infectious diseases: systematic literature review. J Med Internet Res 2021 Oct 20;23(10):e32328 [FREE Full text] [doi: 10.2196/32328] [Medline: 34543228]
31. Ho CW, Ali J, Caals K. Ensuring trustworthy use of artificial intelligence and big data analytics in health insurance. Bull World Health Organ 2020 Apr 01;98(4):263-269 [FREE Full text] [doi: 10.2471/BLT.19.234732] [Medline: 32284650]

32. Ho CW, Soon D, Caals K, Kapur J. Governance of automated image analysis and artificial intelligence analytics in healthcare. Clin Radiol 2019 May;74(5):329-337 [FREE Full text] [doi: 10.1016/j.crad.2019.02.005] [Medline: 30898383]
33. Kerasidou A. Ethics of artificial intelligence in global health: explainability, algorithmic bias and trust. J Oral Biol Craniofac Res 2021 Oct;11(4):612-614 [FREE Full text] [doi: 10.1016/j.jobcr.2021.09.004] [Medline: 34567966]
34. Baig MA, Almuhaizea MA, Alshehri J, Bazarbashi MS, Al-Shagathrh F. Urgent need for developing a framework for the governance of AI in healthcare. Stud Health Technol Inform 2020 Jun 26;272:253-256. [doi: 10.3233/SHTI200542] [Medline: 32604649]
35. Wiens J, Saria S, Sendak M, Ghassemi M, Liu VX, Doshi-Velez F, et al. Do no harm: a roadmap for responsible machine learning for health care. Nat Med 2019 Sep;25(9):1337-1340. [doi: 10.1038/s41591-019-0548-6] [Medline: 31427808]
36. Ploug T, Holm S. The four dimensions of contestable AI diagnostics - a patient-centric approach to explainable AI. Artif Intell Med 2020 Jul;107:101901 [FREE Full text] [doi: 10.1016/j.artmed.2020.101901] [Medline: 32828448]
37. Loi M, Ferrario A, Viganò E. Transparency as design publicity: explaining and justifying inscrutable algorithms. Ethics Inf Technol 2021;23(3):253-263 [FREE Full text] [doi: 10.1007/s10676-020-09564-w] [Medline: 34867077]
38. Makridis C, Hurley S, Klote M, Alterovitz G. Ethical applications of artificial intelligence: evidence from health research on veterans. JMIR Med Inform 2021 Jun 02;9(6):e28921 [FREE Full text] [doi: 10.2196/28921] [Medline: 34076584]
39. Roski J, Maier EJ, Vigilante K, Kane EA, Matheny ME. Enhancing trust in AI through industry self-governance. J Am Med Inform Assoc 2021 Jul 14;28(7):1582-1590 [FREE Full text] [doi: 10.1093/jamia/ocab065] [Medline: 33895824]
40. Stix C. Actionable principles for artificial intelligence policy: three pathways. Sci Eng Ethics 2021 Feb 19;27(1):15 [FREE Full text] [doi: 10.1007/s11948-020-00277-3] [Medline: 33608756]
41. Greenhalgh T, Peacock R. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. BMJ 2005 Nov 05;331(7524):1064-1065 [FREE Full text] [doi: 10.1136/bmj.38636.593461.68] [Medline: 16230312]
42. Alabi RO, Youssef O, Pirinen M, Elmusrati M, Mäkitie AA, Leivo I, et al. Machine learning in oral squamous cell carcinoma: current status, clinical concerns and prospects for future-a systematic review. Artif Intell Med 2021 May;115:102060. [doi: 10.1016/j.artmed.2021.102060] [Medline: 34001326]
43. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 2019 May;1(5):206-215 [FREE Full text] [doi: 10.1038/s42256-019-0048-x] [Medline: 35603010]
44. Abimanyi-Ochom J, Bohingamu Mudiyanselage S, Catchpool M, Firipis M, Wanni Arachchige Dona S, Watts JJ. Strategies to reduce diagnostic errors: a systematic review. BMC Med Inform Decis Mak 2019 Aug 30;19(1):174 [FREE Full text] [doi: 10.1186/s12911-019-0901-1] [Medline: 31470839]
45. Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, Hutchinson B, et al. Model cards for model reporting. In: Proceedings of the Conference on Fairness, Accountability, and Transparency. 2019 Presented at: FAT* '19; January 29-31, 2019; Atlanta, GA, USA p. 220-229 URL: https://dl.acm.org/doi/10.1145/3287560.3287596 [doi: 10.1145/3287560.3287596]
46. Sendak MP, Gao M, Brajer N, Balu S. Presenting machine learning model information to clinical end users with model facts labels. NPJ Digit Med 2020 Mar 23;3(1):41 [FREE Full text] [doi: 10.1038/s41746-020-0253-3] [Medline: 32219182]
47. DECIDE-AI Steering Group. DECIDE-AI: new reporting guidelines to bridge the development-to-implementation gap in clinical artificial intelligence. Nat Med 2021 Feb;27(2):186-187. [doi: 10.1038/s41591-021-01229-5] [Medline: 33526932]
48. Arnold M, Bellamy RK, Hind M, Houde S, Mehta S, Mojsilović A, et al. FactSheets: increasing trust in AI services through supplier's declarations of conformity. IBM J Res Dev 2019 Jul;63(4/5):6:1-613 [FREE Full text] [doi: 10.1147/JRD.2019.2942288]
49. Providing guidelines for the responsible use of AI in healthcare. Coalition for Health AI. URL: https://coalitionforhealthai.org/ [accessed 2022-07-01]
50. Artificial intelligence risk management framework (AI RMF 1.0). National Institute of Standards and Technology. 2023. URL: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf [accessed 2023-06-28]
51. Wang A, Kapoor S, Barocas S, Narayanan A. Against predictive optimization: on the legitimacy of decision-making algorithms that optimize predictive accuracy. SSRN. Preprint posted online on October 18, 2022 [FREE Full text]
52. Malik MM. A hierarchy of limitations in machine learning. arXiv. Preprint posted online on February 29, 2020 [FREE Full text] [doi: 10.48550/arXiv.2002.05193]
53. Lipton ZC. The myth of model interpretability. KDnuggets. 2015 Apr 27. URL: https://www.kdnuggets.com/2015/04/model-interpretability-neural-networks-deep-learning.html [accessed 2023-06-28]
54. Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. arXiv. Preprint posted online on March 2, 2017 [FREE Full text] [doi: 10.48550/arXiv.1702.08608]
55. Caruana RA, Lou Y, Gehrke JE, Koch P, Sturm M, Elhadad N. Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2015 Presented at: KDD '15; August 10-13, 2015; Sydney, Australia p. 1721-1730 URL: https://dl.acm.org/doi/10.1145/2783258.2788613
56. Health AI learning network. health AI learning artnership. URL: https://healthaipartnership.org/ [accessed 2023-03-01]
57. Health care artificial intelligence code of conduct. National Academy of Medicine. URL: https://nam.edu/programs/value-science-driven-health-care/health-care-artificial-intelligence-code-of-conduct/ [accessed 2023-06-19]

58. Bedoya AD, Economou-Zavlanos NJ, Goldstein BA, Young A, Jelovsek JE, O'Brien C, et al. A framework for the oversight and local deployment of safe and high-quality prediction models. J Am Med Inform Assoc 2022 Aug 16;29(9):1631-1636 [FREE Full text] [doi: 10.1093/jamia/ocac078] [Medline: 35641123]

Abbreviations

AI: artificial intelligence
CDS: clinical decision support
CONSORT: Consolidated Standards of Reporting Trials
DECIDE-AI: Developmental and Exploratory Clinical Investigations of Decision support systems driven by Artificial Intelligence
FDA: Food and Drug Administration
IEEE: Institute of Electrical and Electronics Engineers
ML: machine learning
MMS: medical modeling software
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
SaMD: software as a medical device
STROBE: Strengthening the Reporting of Observational Studies in Epidemiology
TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis
WHO: World Health Organization

Edited by A Mavragani; submitted 25.01.23; peer-reviewed by S Reddy, R Marshall; comments to author 19.04.23; revised version received 10.05.23; accepted 11.05.23; published 14.07.23.

Please cite as:
Brereton TA, Malik MM, Lifson M, Greenwood JD, Peterson KJ, Overgaard SM
The Role of Artificial Intelligence Model Documentation in Translational Science: Scoping Review
Interact J Med Res 2023;12:e45903
URL: https://www.i-jmr.org/2023/1/e45903
doi:10.2196/45903
PMID:37450330

Review

The Global Prevalence of Nonalcoholic Fatty Liver Disease and its Association With Cancers: Systematic Review and Meta-Analysis

Nor Asiah Muhamad[1*], MPH, MBBS; Nur Hasnah Maamor[1*], BSc, MSc, PhD; Fatin Norhasny Leman[1], BSc, MSc; Zuraifah Asrah Mohamad[2], BSc, MSc, PhD; Sophia Karen Bakon[2], BSc, MSc; Mohd Hatta Abdul Mutalip[3], BSc, Msc; Izzah Athirah Rosli[1], BSc; Tahir Aris[2], MBBS, MPH; Nai Ming Lai[4], MBBS; Muhammad Radzi Abu Hassan[5], MBBS

[1]Sector for Evidence-based Healthcare, National Institutes of Health, Ministry of Health, Setia Alam, Malaysia
[2]Institute for Medical Research, National Institutes of Health, Ministry of Health, Shah Alam, Malaysia
[3]Institute for Public Health, National Institutes of Health, Ministry of Health, Shah Alam, Malaysia
[4]School of Medicine, Taylor's University, Subang Jaya, Selangor, Malaysia
[5]Office of the Director General of Health, Ministry of Health, Putrajaya, Malaysia
[*]these authors contributed equally

Corresponding Author:
Nor Asiah Muhamad, MPH, MBBS
Sector for Evidence-based Healthcare
National Institutes of Health
Ministry of Health
Jalan Setia Murni U13/52, Seksyen U13
Setia Alam, 40170
Malaysia
Phone: 60 33628000 ext 8705
Email: norasiahdr@gmail.com

Abstract

Background: Nonalcoholic fatty liver disease (NAFLD) is one of the common causes of chronic liver disease globally. Obesity, metabolic diseases, and exposure to some environmental agents contribute to NAFLD. NAFLD is commonly considered a precursor for some types of cancers. Since the leading causes of death in people with NAFLD are cardiovascular disease and extrahepatic cancers, it is important to understand the mechanisms of the progression of NAFLD to control its progression and identify its association with extrahepatic cancers. Thus, this review aims to estimate the global prevalence of NAFLD in association with the risk of extrahepatic cancers.

Objective: We aimed to determine the prevalence of various cancers in NAFLD patients and the association between NAFLD and cancer.

Methods: We searched PubMed, ProQuest, Scopus, and Web of Science from database inception to March 2022 to identify eligible studies reporting the prevalence of NAFLD and the risk of incident cancers among adult individuals (aged ≥18 years). Data from selected studies were extracted, and meta-analysis was performed using random effects models to obtain the pooled prevalence with the 95% CI. The quality of the evidence was assessed with the Newcastle-Ottawa Scale.

Results: We identified 11 studies that met our inclusion criteria, involving 222,523 adults and 3 types of cancer: hepatocellular carcinoma (HCC), breast cancer, and other types of extrahepatic cancer. The overall pooled prevalence of NAFLD and cancer was 26% (95% CI 16%-35%), while 25% of people had NAFLD and HCC (95% CI 7%-42%). NAFLD and breast cancer had the highest prevalence out of the 3 forms of cancer at 30% (95% CI 14%-45%), while the pooled prevalence for NAFLD and other cancers was 21% (95% CI 12%-31%).

Conclusions: The review suggests that people with NAFLD may be at an increased risk of cancer that might not affect not only the liver but also other organs, such as the breast and bile duct. The findings serve as important evidence for policymakers to evaluate and recommend measures to reduce the prevalence of NAFLD through lifestyle and environmental preventive approaches.

Trial Registration: PROSPERO CRD42022321946; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=321946

(Interact J Med Res 2023;12:e40653) doi:10.2196/40653

KEYWORDS

fatty liver; nonalcoholic fatty liver; NAFL; nonalcoholic fatty liver disease; NAFLD; prevalence; cancers; cancer; extrahepatic; liver; carcinoma

Introduction

Nonalcoholic fatty liver disease (NAFLD) has a global prevalence of 25% and is a leading cause of cirrhosis and hepatocellular carcinoma [1]. NAFLD encompasses a disease continuum from steatosis with or without mild inflammation (nonalcoholic fatty liver; NAFL) to nonalcoholic steatohepatitis (NASH), which is characterized by necroinflammation and faster fibrosis progression than nonalcoholic fatty liver. NAFLD has a bidirectional association with components of the metabolic syndrome, and type 2 diabetes increases the risk of cirrhosis and related complications. It has become one of the most frequent chronic liver diseases. NAFLD is determined by the presence of ≥5% in the liver or more of hepatic steatosis [2] in liver magnetic resonance imaging proton density fat fraction findings or biopsies in the absence of secondary causes of hepatic fat accumulation, such as hepatitis C infection and glycogen storage disease [3]. At the same time, other approaches to diagnose NAFLD, such as clinical and laboratory scores, have been developed, even though the accuracy cannot be determined. Several biomarkers demonstrate better performance in the diagnosis of NAFLD. Therefore, many approaches are being combined with artificial intelligence to increase diagnostic performance [4]. Electronic health education, which includes the use of mobile communication devices (eg, smartphones and tablet computers), creates awareness and health alerts regarding the disease [5]. This technique is useful for educational purposes, such as promoting healthy behavior for community prevention and early screening.

NAFLD encompasses a spectrum of diseases including NAFL, which has a more benign course, and NASH, which can progress to cirrhosis and hepatocellular carcinoma (HCC) [6]. Due to the heterogeneous nature of the disease, the undetermined symptoms, and the high disease burden, there is increasing appreciation that NAFLD may also be becoming an important cause of HCC [7]. The reported global prevalence of NAFLD and HCC varies between 2% and 58.5% [8].

NAFLD is a complex, multifactorial disease caused by a sedentary lifestyle, obesity, poor dietary habits, intestinal flora, genetics, and other factors [9-11]. NAFLD typically occurs in patients with metabolic syndrome, ranging from simple hepatic steatosis to nonalcoholic steatohepatitis [12]. Visceral adiposity and insulin resistance are among important conditions associated with NAFLD that have been extensively studied [13-17]. Although NAFLD normally has a good prognosis, it can progress to nonalcoholic steatohepatitis, liver fibrosis, cirrhosis, HCC, and even breast cancer [18-20], with male patients having a high risk of such complications [21]. The presence of progressive liver disease is frequently detected only after it has advanced to a late stage. Patients with advanced liver disease usually do not respond well to intervention and have high risks of mortality and morbidity. The prevalence of NAFLD and its association with hepatic and extrahepatic complications have been reported in individual studies, many of which have been published recently. An up-to-date review of the current evidence to determine the global magnitude of the problem is warranted.

In this review, we systematically synthesized published online evidence on the global prevalence of NAFLD and quantified the magnitude of the association between NAFLD and the risk of extrahepatic cancers.

Methods

We conducted a systematic review of observational studies and report our findings in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Multimedia Appendix 1) and the GATHER (Guidelines for Accurate and Transparent Health Estimates Reporting) statement [22] (Multimedia Appendix 2).

Search Strategy

We searched various electronic databases, including PubMed, ProQuest, Scopus, and Web of Science, from inception to March 31, 2022, to identify relevant studies reporting the prevalence of NAFLD and cancers incident among adults (aged ≥18 years). A search strategy was generated systematically for PubMed and adapted for use in the other electronic databases. Keywords and comparable medical subject heading (MESH) terms were used in combination where appropriate, without restriction on language or publication year. Specific search terms were as follows: ("Fatty liver" OR "steatosis") AND ("cancer" OR "malignancy" OR "tumour" OR "carcinoma") AND "prevalence" for the PubMed search. Searches were restricted to human studies. Studies in languages other than English were excluded. Additionally, we screened cross-references from relevant original papers and review articles to identify primary studies not covered by the original database searches. We used a reference manager (EndNote; Clarivate Plc) to store all the studies and citations and developed a database for the included and excluded studies. We integrated our reporting with the MOOSE (Meta-Analysis of Observational Studies in Epidemiology) guidelines [23].

Criteria for Considering Studies for This Review

We included all observational studies that evaluated the prevalence of NAFLD and examined the association between NAFLD and the risk of developing cancers (Textbox 1). We included studies with an adult population aged 18 years and older of either sex without any restriction in terms of race or ethnicity in the meta-analysis. We excluded conference abstracts, case reports, reviews, commentaries, editorials, practice guidelines, and cross-sectional studies. In the case of multiple studies using the same cohort, the study with the most detailed information on the participants or the largest number of participants was selected. We included studies according to the population, intervention, comparator, outcome, study (PICOS) approach (Table S1 in Multimedia Appendix 3).

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Observational study
- Participants were adults aged 18 years and older

Exclusion criteria

- Conference abstracts, case reports, reviews, commentaries, editorials, practice guidelines, and cross-sectional studies
- Participants were younger than 18 years

Study Selection

Two authors (NHM and FNL) independently screened all the titles and abstracts to search for potential studies identified as a result of the search and coded them as “retrieve” (eligible, potentially eligible, or unclear) or “do not retrieve.” Another 2 authors (ZAM and SKB) independently retrieved the full-text study reports and publications to identify studies for inclusion, as well as identify and record reasons for the studies’ exclusion. In case of discrepancies, an agreement was reached by consensus and discussion with a third reviewer (NML).

Data Extraction Process

Two authors (IAR and MHAM) independently evaluated the methodological quality of each included study and extracted data using an electronic form in the reference manager, which was adapted from the Cochrane Handbook for Systematic Review of Interventions [24]; discrepancies were resolved by discussion with a third author (NAM). A standardized data collection form was used to extract the data. The extraction form contained information about the author, year of publication, setting (country of origin), population, study design, sample size, study period, presence of fatty liver, method of detection of fatty liver, type of cancer, and prevalence of NAFLD.

Risk of Bias in Included Studies

Methodological quality was determined using a domain-based tool adapted from the Newcastle-Ottawa Scale (NOS) to assess the risk of bias of each study (Multimedia Appendix 4). We classified the risk of bias as either low, moderate, high, or unclear across the following domains: selection of participants (selection bias), sample size justification (selection bias), outcome measurement (detection bias), and confounding adjustment. We assigned a score of 7 and above as good quality, and below 6 as having concerns related to determining the overall quality [25].

Data Synthesis and Analysis

We used Stata (version 16; StataCorp) for all statistical analysis. The pooled prevalence rates, as well as their 95% CIs, were calculated using a random effects model [26]. The I^2 statistic and Cochran Q test were used to assess heterogeneity among the studies [27]. The I^2 describes the percentage of the variability in effect estimates that is attributable to heterogeneity rather than sampling error. A value greater than 50% may be considered to indicate substantial heterogeneity, whereas a score of more than 75% indicates high heterogeneity [26,28].

Results

Identification of Studies

From the preliminary search, we identified 4687 studies across all databases, including PubMed (n=3190), ProQuest (n=89), Scopus (n=1290), and Web of Science (n=118). After removing duplicates, we screened a total of 158 titles and abstracts, of which 143 were excluded. We retrieved a total of 15 full texts for inclusion. We excluded 4 studies from the full texts and included 11 studies that met our inclusion criteria for this review. The most common reasons for exclusion were (1) no data on population prevalence of NFALD and cancer was reported and (2) the cancer condition was not included in the study. After removing duplicates, titles, and abstracts, 4 studies were excluded. We included 11 studies, as shown in Table 1 and Figure 1. Figure 2 shows the selection process to complete the PRISMA flow diagram.

Table 1. Characteristics of included studies.

Author/year	Country	Study design	Sample size, n	Setting	NAFLD[a] diagnosis	Type of cancer	Length of follow up	Prevalence of NAFLD (%)	Hazard ratio (95% CI)
Başaranoğlu et al (2014) [29]	Turkey	Prospective	105	Hospital	NAFLD defined by observing increased echogenicity and liver-kidney contrast using ultrasonography	Cancer in first-degree relatives	Not stated	27	Not stated
Lee et al (2017) [30]	Korea	Retrospective	104	Hospital	NAFLD defined by observing hepatic fat accumulation using magnetic resonance imaging	Breast	Not stated	18.3	Not stated
Chan et al (2017) [31]	China	Retrospective	270	Hospital	NAFLD defined as presence of steatosis ≥5%; steatohepatitis was defined as the presence of steatosis ≥5%, ballooning degeneration, and lobular inflammation in a liver biopsy	HCC[b]	3-12 months	39.6	6.84 (1.48-31.66)
Tian et al (2021) [32]	China	Cohort	263	Hospital	NAFLD investigated by ultrasound. Ten successful reads were required and the median was recorded. The ratio of the IQR divided by median (IQR/median) of all measurements less than 30% with a success rate (successful tests/total tests) ≥60% was regarded as a valid measurement and controlled attenuation parameter ≥240 was defined as hepatic steatosis	Breast	Not stated	41.8	Not stated
Tokushige et al (2011) [33]	Japan	Cohort	292	Hospital	Diagnosis of NAFLD was based on the following criteria: (1) detection of hepatic steatosis (or steatohepatitis) by liver biopsy or imaging; (2) intake of less than 20-30 grams of ethanol per day (as confirmed by the attending physician and family members who were in close contact with the patient); and (3) appropriate exclusion of other liver diseases (such as alcoholic liver disease, viral hepatitis, autoimmune hepatitis, drug-induced liver disease, primary biliary cirrhosis, primary sclerosing cholangitis, biliary obstruction, and metabolic liver diseases such as Wilson disease and hemochromatosis).	HCC	Not stated	2.0	Not stated
Lee et al (2019) [34]	Korea	Cohort	321	Hospital	NAFLD defined as the presence of ≥5% hepatic steatosis using ultrasound	HCC	3-6 months	8.2	3.005 (1.122-8.051)
Zarrinpar et al (2019) [35]	US	Retrospective	317	Hospital	NAFLD defined by having steatohepatitis based on the hospital file record	HCC	Not stated	24.0	Not stated

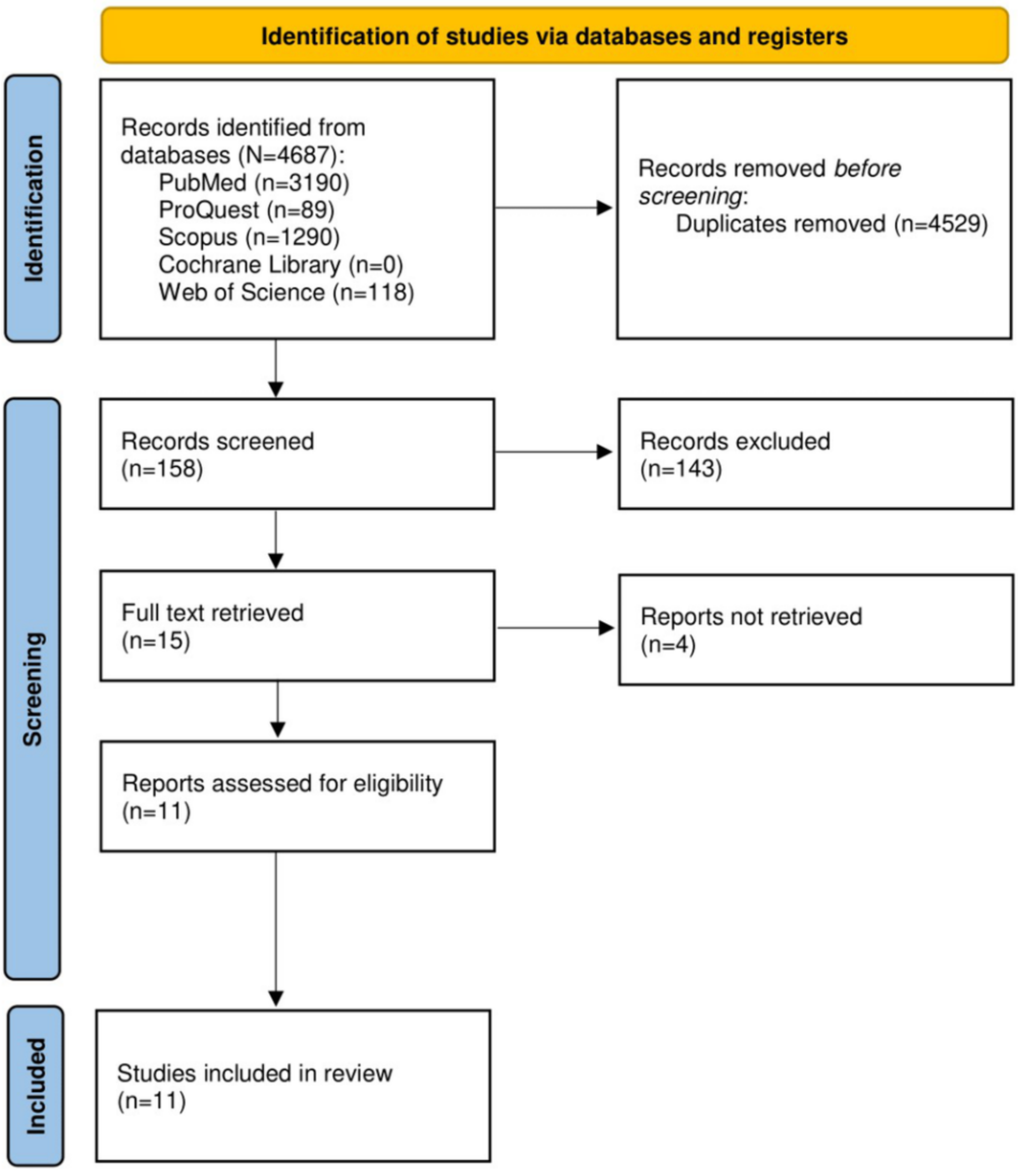
Author/year	Country	Study design	Sample size, n	Setting	NAFLD[a] diagnosis	Type of cancer	Length of follow up	Prevalence of NAFLD (%)	Hazard ratio (95% CI)
Reddy et al (2013) [36]	US	Cohort	181	Hospital	The underlying liver pathology reported in this study was based on a dedicated rereview of the pathological slides of each resection specimen by an experienced hepatobiliary pathologist at each respective center. Steatosis grade, lobular inflammation, hepatocyte ballooning, extent of fibrosis, and portal inflammation were described according to previous articles. Instead of the precise number of foci per high power field, lobular inflammation was reported as “none,” “rare/spotty,” “mild,” or “moderate/heavy.” Each of these terms was then coded in increasing severity from 0 to 3 to calculate the NAFLD activity score.	Intrahepatic; cholangiocarcinoma	Not stated	17.1	Not stated
Nseir et al (2017) [37]	Israel	Retrospective	133	Hospital	Presence of hepatic steatosis on abdominal computerized tomography examination and an attenuation of –5 to 10 Hounsfield units (calculated as liver attenuation minus spleen attenuation), no alcohol consumption (<20 g/day), negative serology for hepatitis B or C virus, negative to antibodies for autoimmune hepatitis, and no history of other known liver disease	Breast	Not stated	45.2	Not stated
Asfari et al (2020) [38]	US	Cross-sectional	218,950	Hospital	The study group was identified using the International Classification of Diseases 9th version code for nonalcoholic steatohepatitis	HCC	Not stated	50.0	Not stated
Lee et al (2019) [39]	Korea	Retrospective	1587	Hospital	NAFLD was diagnosed when the mean attenuation of the liver was lower than 40 Hounsfield units or 10 units lower than that of the spleen by using CT scanning, magnetic resonance imaging, or biopsy	Breast	Not stated	15.8	1.581 (1.038-2.410)

[a]NAFLD: nonalcoholic fatty liver disease.

[b]HCC: hepatocellular carcinoma.

Figure 1. Map of study sites in the included articles.

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of the selection process.



Characteristics of Included Studies

A total of 11 studies with 222,523 adult participants (aged ≥18 years) across 6 countries were included. A total of 3 studies were conducted in the United States and 8 studies were from 5 countries in Asia: South Korea (n=3), China (n=2), Turkey (n=1), Japan (n=1), and Israel (n=1). Cancer was classified into 3 types: HCC (5 studies), breast cancer (4 studies), and others (2 studies). The characteristics of the included studies are depicted in Table 1.

Risk of Bias of Included Studies

We assessed the quality of the included studies [29-39] as NOS quality, with modifications as shown in Table S2 in Multimedia Appendix 3. All 11 studies were observed to have good quality (NOS score 7 and above).

The Estimates of Pooled Prevalence of NAFLD and Cancer

The overall pooled prevalence of NAFLD among the adult population was 26% (95% CI 16%-35%), as shown in Figure 3, with a high level of heterogeneity between studies (I^2=99.2%; *P*=.001). The overall pooled prevalence of HCC was calculated based on these findings, and it was observed that the overall pooled prevalence of HCC was 25% (95% CI 7%-42%) with a high level of heterogeneity between studies (I^2=99.6%; *P*=.001) reported in Asia (4 publications) and the United States (1 publication; Figure 4). The overall pooled prevalence of breast cancer was 30% (95% CI 14%-45%) with a high level of heterogeneity between studies (I^2=96.9%; *P*=.001) reported in Asia (4 publications), as shown in Figure 5. In this review, cancer in first degree relatives and intrahepatic cholangiocarcinoma showed an overall pooled prevalence of 21% (95% CI 12%-31%) with a moderate level of heterogeneity between studies (I^2=71%; *P*=.001) reported in Asia and the United States, each with 1 publication (Figure 6).

Figure 3. Forest plot of overall prevalence of fatty liver and cancer [29-39].

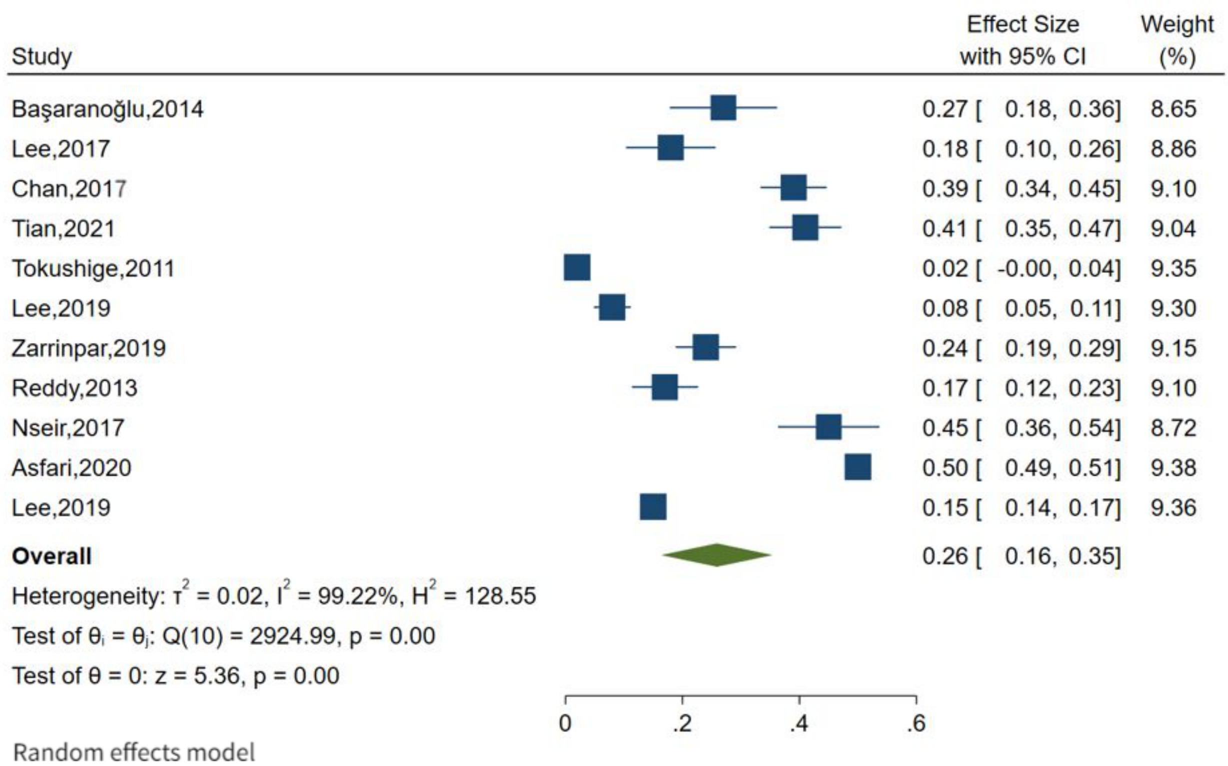


Figure 4. Forest plot of overall prevalence of fatty liver and hepatocellular carcinoma [31,33-35,38].

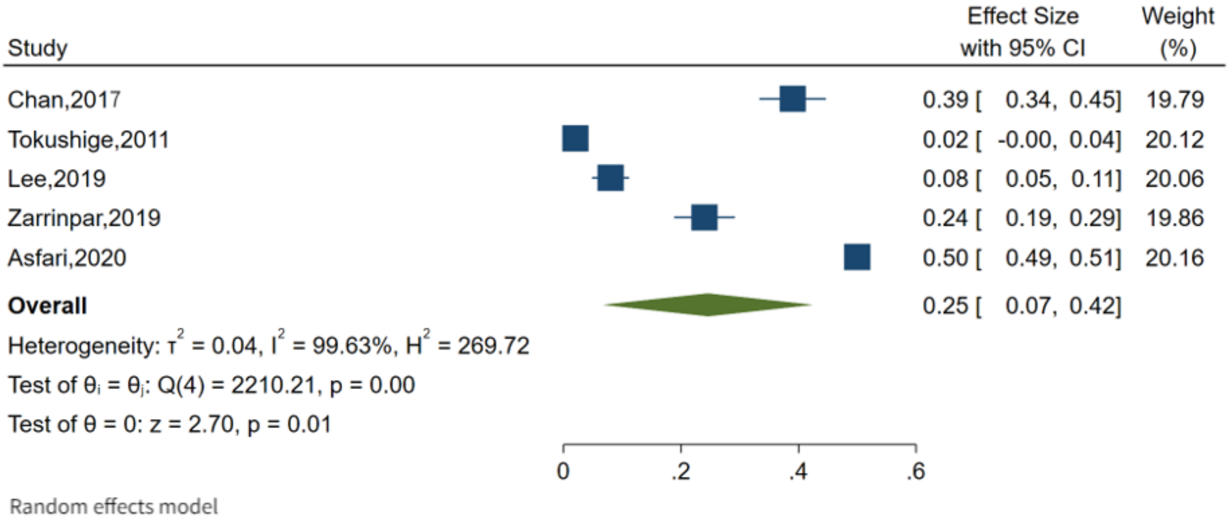


Figure 5. Forest plot of overall prevalence of fatty liver and breast cancer [30,32,37,39].

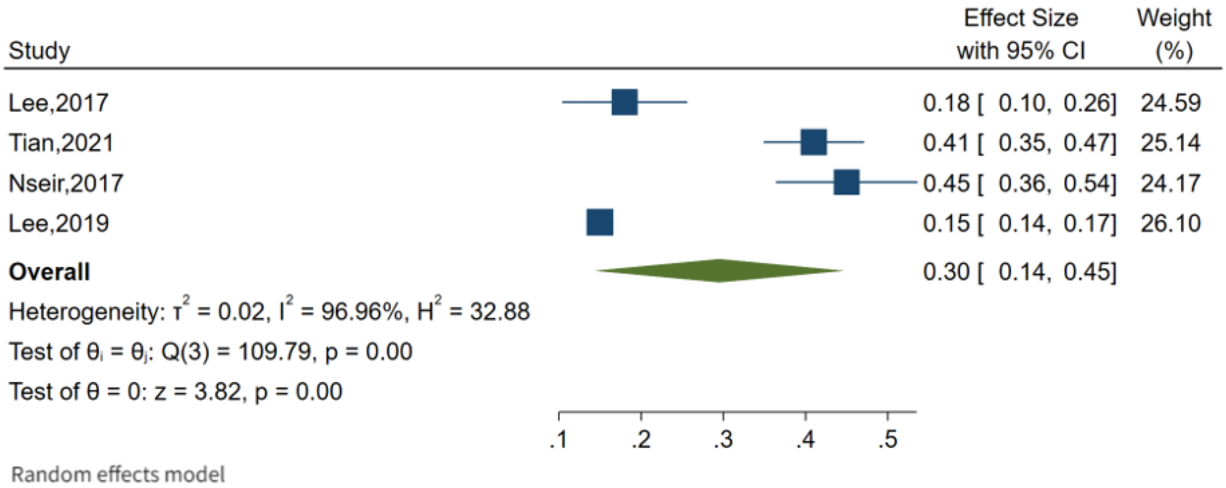
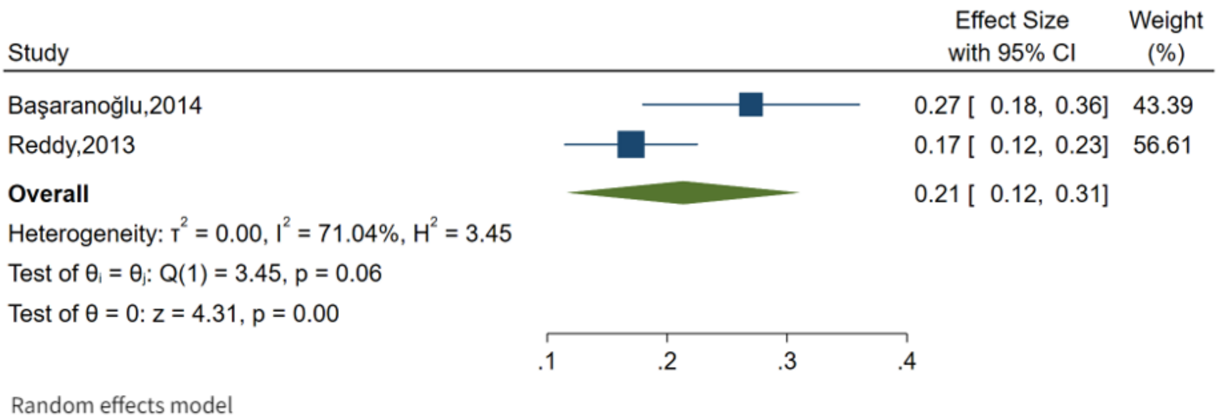


Figure 6. Forest plot of overall prevalence of fatty liver and other cancers [29,36].



Discussion

Principal Findings

This review elucidates an assessment of NAFLD as a worldwide epidemic that may contribute to chronic liver disease. In this systematic review and meta-analysis, we included 11 studies to comprehensively estimate the global prevalence and the association with cancers among the adult population with NAFLD. There is strong evidence that suggests the association of NAFLD with various diseases, such as cardiovascular disease, diabetes, and cancers such as bile duct, breast, and liver cancer [11,40]. In recent years, more research and discussion has been focused on the possible association between NAFLD and the risk of cancers. Therefore, this review will be useful in expounding the latest research progress on this issue [41].

In this review, the overall pooled prevalence of NAFLD and cancers was 26%. The pooled prevalence of NAFLD and breast cancer was higher than that of HCC and other cancers (30% vs 25% and 21%, respectively). This review also discovered that the highest prevalence of NAFLD is reported in breast cancer, which is consistent with the findings of studies by Kim et al [42] (2017) and Mantovani et al [43] (2022). According to the former report, NAFLD showed a strong association with the development of HCC and colorectal cancer in men and breast cancer in women [42]. The latter study, a meta-analysis, examined the risk of several prespecified cancers and found an increased risk of developing breast cancer [43].

A review by Thomas et al [44] (2022) examined NAFLD and the incidence of hepatic and extrahepatic cancers. The review found that the pooled incidence rate of HCC was 1.25 per 1000 person-years. However, our review examined the prevalence of HCC among the adult population with NFLD. Another review [45] found that the presence of NAFLD was independently associated with an 88% increased risk of HCC, as compared to the absence of NAFLD. This review is different from ours, which examined the prevalence of HCC with NAFLD.

Although the exact or detailed mechanism of this interaction remains unclear, cancer development in patients with NAFLD may be associated with a bidirectional interaction between NAFLD and metabolic syndrome [39]. According to Tiniakos et al [46] (2018), the increased susceptibility of the liver with steatosis to carcinogenic insults could be linked to metabolic derangements, such as metabolic syndrome, hyperinsulinemia, and the presence of insulin-like growth factor receptors in HCC, as well as systemic effects of deranged cytokines and adipokines, immune dysregulation, and changes in gut microbiota. Besides that, the genetic component has been named as another factor that contributes to an increased risk of HCC in individuals with NAFLD [46].

A few mechanisms for extrahepatic carcinogenesis of the fatty liver, such as breast cancer, have been proposed. First, high levels of inflammatory cytokines are closely associated with NAFLD, as they promote insulin resistance and elevated circulating triglycerides, influence growth, and increase apoptosis and tumor cell proliferation in many cancers [47]. Second, hyperinsulinemia and high levels of leptin have carcinogenic effects [48]. By binding to the circulating sex hormone–binding globulin, increasing insulin levels cause the elevated secretion of estrogen, and downstream signaling favors breast carcinogenesis [49]. Insulin may cross-bind to insulin-like growth factor-I (IGF-1) receptors on breast cells, and downstream signaling pathways provide proliferation stimuli to breast cancer cells [19]. Third, decreased levels of adiponectin lead to marked insulin resistance and subsequent increased levels of IGF-1. Insulin binds to IGF-1 receptors and plays an important role in cell proliferation, apoptosis, and increased production of vascular endothelial growth factors [37].

Strengths and Study Limitations

This study has several strengths. To our knowledge, this is a comprehensive and up-to-date systematic review and meta-analysis that reports the pooled prevalence of NAFLD and cancers. We examined a number of studies using stringent inclusion criteria. Following the broad search strategy, we were able to stratify hepatic and extrahepatic cancers. There are potential limitations in our review. The included studies largely had an observational design, and there are risks for a number of biases in term of design, selection of respondents, and sample size. The included studies were located only in Asia (n=8) or the United States (n=3), limiting the accuracy of the estimations. There is a paucity of data from other countries, and most studies are conducted among the urban population. Therefore, our findings might be biased and have limited generalizability beyond the countries included in this review. In short, there is a need to investigate NAFLD and cancer across different regions in the future.

Conclusion

This systematic review and meta-analysis shows that the pooled prevalence of NAFLD is closely associated with hepatic and extrahepatic cancers. This review provides evidence that a substantially high proportion of patients with NAFLD are associated with extrahepatic and hepatocellular carcinoma. However, the evidence remains inconclusive, and further studies are needed to confirm the association between NAFLD and cancers, as well as to improve surveillance strategies for patients with NAFLD who are at high risk of cancers. Since the global prevalence of NAFLD is increasing, policymakers must work toward reversing the current trends by increasing the awareness of NAFLD and promoting healthy lifestyles and environments.

Acknowledgments

We would like to thank the director-general of health of the Ministry of Health, Malaysia, for permission to publish this article.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' Contributions

NHM and NAM carried out study selection, data extraction, and statistical analysis and drafted the manuscript. ZAM and SKB drafted the manuscript, managed references, and completed the PROSPERO application. FNL and IAR participated in the study selection and data extraction and drafted the manuscript. TA evaluated the quality of the included studies. MRAH and NML participated in the discussion of any discrepancies and supervised the study. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 31 KB - ijmr_v12i1e40653_app1.docx]

Multimedia Appendix 2
GATHER (Guidelines for Accurate and Transparent Health Estimates Reporting) checklist.
[DOCX File , 31 KB - ijmr_v12i1e40653_app2.docx]

Multimedia Appendix 3
Table S1: The population, intervention, comparator, outcome, study (PICOS) approach. Table S2: Newcastle-Ottawa Quality Assessment Form for Cohort Studies.
[DOCX File , 16 KB - ijmr_v12i1e40653_app3.docx]

Multimedia Appendix 4
Newcastle-Ottawa Quality Assessment Form for Cohort Studies.
[DOCX File , 17 KB - ijmr_v12i1e40653_app4.docx]

References

1. Powell EE, Wong VW, Rinella M. Non-alcoholic fatty liver disease. Lancet 2021 Jun 05;397(10290):2212-2224. [doi: 10.1016/S0140-6736(20)32511-3] [Medline: 33894145]
2. Carr RM, Oranu A, Khungar V. Nonalcoholic fatty liver disease: Pathophysiology and management. Gastroenterol Clin North Am 2016 Dec;45(4):639-652 [FREE Full text] [doi: 10.1016/j.gtc.2016.07.003] [Medline: 27837778]
3. Kneeman JM, Misdraji J, Corey KE. Secondary causes of nonalcoholic fatty liver disease. Therap Adv Gastroenterol 2012 May;5(3):199-207 [FREE Full text] [doi: 10.1177/1756283X11430859] [Medline: 22570680]
4. Zhou L, Wang J, Yu S, Wu G, Wei Q, Deng Y, et al. Artificial intelligence in medical imaging of the liver. World J Gastroenterol 2019 Feb 14;25(6):672-682 [FREE Full text] [doi: 10.3748/wjg.v25.i6.672] [Medline: 30783371]
5. Majumder S, Deen MJ. Smartphone sensors for health monitoring and diagnosis. Sensors (Basel) 2019 May 09;19(9):2164 [FREE Full text] [doi: 10.3390/s19092164] [Medline: 31075985]
6. Pouwels S, Sakran N, Graham Y, Leal A, Pintar T, Yang W, et al. Non-alcoholic fatty liver disease (NAFLD): a review of pathophysiology, clinical management and effects of weight loss. BMC Endocr Disord 2022 Mar 14;22(1):63 [FREE Full text] [doi: 10.1186/s12902-022-00980-1] [Medline: 35287643]
7. Younossi ZM, Otgonsuren M, Henry L, Venkatesan C, Mishra A, Erario M, et al. Association of nonalcoholic fatty liver disease (NAFLD) with hepatocellular carcinoma (HCC) in the United States from 2004 to 2009. Hepatology 2015 Dec;62(6):1723-1730. [doi: 10.1002/hep.28123] [Medline: 26274335]
8. Huang DQ, El-Serag HB, Loomba R. Global epidemiology of NAFLD-related HCC: trends, predictions, risk factors and prevention. Nat Rev Gastroenterol Hepatol 2021 Apr;18(4):223-238 [FREE Full text] [doi: 10.1038/s41575-020-00381-6] [Medline: 33349658]
9. Chang Y, Jung H, Cho J, Zhang Y, Yun K, Lazo M, et al. Metabolically healthy obesity and the development of nonalcoholic fatty liver disease. Am J Gastroenterol 2016 Aug;111(8):1133-1140. [doi: 10.1038/ajg.2016.178] [Medline: 27185080]
10. Kwak M, Kim D, Chung GE, Kim W, Kim YJ, Yoon J. Role of physical activity in nonalcoholic fatty liver disease in terms of visceral obesity and insulin resistance. Liver Int 2015 Mar;35(3):944-952. [doi: 10.1111/liv.12552] [Medline: 24684289]
11. Ahmed IA, Mikail MA, Mustafa MR, Ibrahim M, Othman R. Lifestyle interventions for non-alcoholic fatty liver disease. Saudi J Biol Sci 2019 Nov;26(7):1519-1524 [FREE Full text] [doi: 10.1016/j.sjbs.2018.12.016] [Medline: 31762620]
12. Liu Y, Zhong GC, Tan HY, Hao FB, Hu JJ. Nonalcoholic fatty liver disease and mortality from all causes, cardiovascular disease, and cancer: a meta-analysis. Sci Rep 2019 Jul 31;9(1):11124 [FREE Full text] [doi: 10.1038/s41598-019-47687-3] [Medline: 31366982]
13. Cree-Green M, Ravi S, Carreau A, Sewell R, Baumgartner A, Coe G, et al. Nonalcoholic fatty liver disease in obese adolescent females is associated with multi-tissue insulin resistance and visceral adiposity markers. Metabol Open 2019 Jun;2:100011 [FREE Full text] [doi: 10.1016/j.metop.2019.100011] [Medline: 32812939]
14. Cariou B. The metabolic triad of non-alcoholic fatty liver disease, visceral adiposity and type 2 diabetes: Implications for treatment. Diabetes Obes Metab 2022 Feb;24 Suppl 2:15-27. [doi: 10.1111/dom.14651] [Medline: 35014161]
15. Eguchi Y, Eguchi T, Mizuta T, Ide Y, Yasutake T, Iwakiri R, et al. Visceral fat accumulation and insulin resistance are important factors in nonalcoholic fatty liver disease. J Gastroenterol 2006 May;41(5):462-469. [doi: 10.1007/s00535-006-1790-5] [Medline: 16799888]

16. Ercin CN, Dogru T, Genc H, Celebi G, Aslan F, Gurel H, et al. Insulin resistance but not visceral adiposity index is associated with liver fibrosis in nondiabetic subjects with nonalcoholic fatty liver disease. Metab Syndr Relat Disord 2015 Sep;13(7):319-325. [doi: 10.1089/met.2015.0018] [Medline: 26011302]
17. Kadowaki S, Tamura Y, Someya Y, Takeno K, Kaga H, Sugimoto D, et al. Fatty liver has stronger association with insulin resistance than visceral fat accumulation in nonobese Japanese men. J Endocr Soc 2019 Jul 01;3(7):1409-1416 [FREE Full text] [doi: 10.1210/js.2019-00052] [Medline: 31286107]
18. Ramesh S, Sanyal AJ. Evaluation and management of non-alcoholic steatohepatitis. J Hepatol 2005;42 Suppl(1):S2-12. [doi: 10.1016/j.jhep.2004.11.022] [Medline: 15777569]
19. Kwak M, Yim JY, Yi A, Chung G, Yang JI, Kim D, et al. Nonalcoholic fatty liver disease is associated with breast cancer in nonobese women. Dig Liver Dis 2019 Jul;51(7):1030-1035. [doi: 10.1016/j.dld.2018.12.024] [Medline: 30686716]
20. Su Q, Kumar V, Sud N, Mahato RI. MicroRNAs in the pathogenesis and treatment of progressive liver injury in NAFLD and liver fibrosis. Adv Drug Deliv Rev 2018 Apr;129:54-63. [doi: 10.1016/j.addr.2018.01.009] [Medline: 29391222]
21. Kim D, Kim W, Joo SK, Kim JH, Harrison SA, Younossi ZM, et al. Predictors of nonalcoholic steatohepatitis and significant fibrosis in non-obese nonalcoholic fatty liver disease. Liver Int 2019 Feb;39(2):332-341. [doi: 10.1111/liv.13983] [Medline: 30298568]
22. Stevens G, Alkema L, Black R, Boerma J, Collins G, Ezzati M, GATHER Working Group. Guidelines for Accurate and Transparent Health Estimates Reporting: the GATHER statement. PLoS Med 2016 Jun;13(6):e1002056 [FREE Full text] [doi: 10.1371/journal.pmed.1002056] [Medline: 27351744]
23. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000 Apr 19;283(15):2008-2012. [doi: 10.1001/jama.283.15.2008] [Medline: 10789670]
24. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M. Cochrane Handbook for Systematic Reviews of Interventions, version 6.3. URL: https://training.cochrane.org/handbook [accessed 2023-06-14]
25. The Newcastle-Ottawa Scale (NOS) for assessing the quality of non-randomized studies in meta-analysis. Ottawa Hospital Research Institute. URL: https://www.ohri.ca/programs/clinical_epidemiology/oxford.asp [accessed 2023-06-14]
26. Riley RD, Higgins JPT, Deeks JJ. Interpretation of random effects meta-analyses. BMJ 2011 Feb 10;342:d549 [FREE Full text] [doi: 10.1136/bmj.d549] [Medline: 21310794]
27. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. Res Synth Methods 2010 Apr;1(2):97-111. [doi: 10.1002/jrsm.12] [Medline: 26061376]
28. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003 Sep 06;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
29. Başaranoğlu M, Canbakan B, Yıldız K, Ceylan B, Baysal B, Uysal O, et al. Fatty liver disease might increase the risk of abdominal operation in patients with fatty liver and the prevalence of cancer in first-degree relatives. Turk J Gastroenterol 2014 Dec;25 Suppl 1:138-141 [FREE Full text] [doi: 10.5152/tjg.2014.7674] [Medline: 25910292]
30. Lee S, Jung Y, Bae Y, Yun SP, Kim S, Jo H, et al. Prevalence and risk factors of nonalcoholic fatty liver disease in breast cancer patients. Tumori 2017 Mar 24;103(2):187-192. [doi: 10.5301/tj.5000536] [Medline: 27647227]
31. Chan AWH, Wong GLH, Chan H, Tong JHM, Yu Y, Choi PCL, et al. Concurrent fatty liver increases risk of hepatocellular carcinoma among patients with chronic hepatitis B. J Gastroenterol Hepatol 2017 Mar;32(3):667-676. [doi: 10.1111/jgh.13536] [Medline: 27547913]
32. Tian S, Li H, Li R, Ran L, Li S, Wu J, et al. Prevalence of hepatic steatosis and metabolic associated fatty liver disease among female breast cancer survivors. Chin Med J (Engl) 2022 Oct 05;135(19):2372-2374 [FREE Full text] [doi: 10.1097/CM9.0000000000002121] [Medline: 36535013]
33. Tokushige K, Hashimoto E, Horie Y, Taniai M, Higuchi S. Hepatocellular carcinoma in Japanese patients with nonalcoholic fatty liver disease, alcoholic liver disease, and chronic liver disease of unknown etiology: report of the nationwide survey. J Gastroenterol 2011 Oct;46(10):1230-1237. [doi: 10.1007/s00535-011-0431-9] [Medline: 21748549]
34. Lee YB, Ha Y, Chon YE, Kim MN, Lee JH, Park H, et al. Association between hepatic steatosis and the development of hepatocellular carcinoma in patients with chronic hepatitis B. Clin Mol Hepatol 2019 Mar;25(1):52-64 [FREE Full text] [doi: 10.3350/cmh.2018.0040] [Medline: 30360031]
35. Zarrinpar A, Faltermeier CM, Agopian VG, Naini BV, Harlander-Locke MP, Kaldas FM, et al. Metabolic factors affecting hepatocellular carcinoma in steatohepatitis. Liver Int 2019 Mar;39(3):531-539 [FREE Full text] [doi: 10.1111/liv.14002] [Medline: 30427105]
36. Reddy SK, Hyder O, Marsh JW, Sotiropoulos GC, Paul A, Alexandrescu S, et al. Prevalence of nonalcoholic steatohepatitis among patients with resectable intrahepatic cholangiocarcinoma. J Gastrointest Surg 2013 Apr;17(4):748-755 [FREE Full text] [doi: 10.1007/s11605-013-2149-x] [Medline: 23355033]
37. Nseir W, Abu-Rahmeh Z, Tsipis A, Mograbi J, Mahamid M. Relationship between non-alcoholic fatty liver disease and breast cancer. Isr Med Assoc J 2017 Apr;19(4):242-245 [FREE Full text] [Medline: 28480679]
38. Asfari M, Talal Sarmini M, Alomari M, Lopez R, Dasarathy S, McCullough A. The association of nonalcoholic steatohepatitis and hepatocellular carcinoma. Eur J Gastroenterol Hepatol 2020 Dec;32(12):1566-1570 [FREE Full text] [doi: 10.1097/MEG.0000000000001681] [Medline: 32073443]

39. Lee Y, Lee H, Chang S, Lee C, Kim J, Jung Y, et al. Underlying nonalcoholic fatty liver disease is a significant factor for breast cancer recurrence after curative surgery. Medicine (Baltimore) 2019 Sep;98(39):e17277. [doi: 10.1097/MD.0000000000017277] [Medline: 31574842]
40. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. Hepatology 2016 Jul;64(1):73-84. [doi: 10.1002/hep.28431] [Medline: 26707365]
41. Liu S, Ma X, Zhao J, Du S, Zhang J, Dong M, et al. Association between nonalcoholic fatty liver disease and extrahepatic cancers: a systematic review and meta-analysis. Lipids Health Dis 2020 May 31;19(1):118 [FREE Full text] [doi: 10.1186/s12944-020-01288-6] [Medline: 32475354]
42. Kim G, Lee HC, Choe J, Kim M, Lee MJ, Chang H, et al. Association between non-alcoholic fatty liver disease and cancer incidence rate. J Hepatol 2017 Nov 02;68(1):140. [doi: 10.1016/j.jhep.2017.09.012] [Medline: 29150142]
43. Mantovani A, Petracca G, Beatrice G, Csermely A, Tilg H, Byrne CD, et al. Non-alcoholic fatty liver disease and increased risk of incident extrahepatic cancers: a meta-analysis of observational cohort studies. Gut 2022 Apr;71(4):778-788. [doi: 10.1136/gutjnl-2021-324191] [Medline: 33685968]
44. Thomas JA, Kendall BJ, Dalais C, Macdonald GA, Thrift AP. Hepatocellular and extrahepatic cancers in non-alcoholic fatty liver disease: A systematic review and meta-analysis. Eur J Cancer 2022 Sep;173:250-262. [doi: 10.1016/j.ejca.2022.06.051] [Medline: 35944373]
45. Petrelli F, Manara M, Colombo S, De Santi G, Ghidini M, Mariani M, et al. Hepatocellular carcinoma in patients with nonalcoholic fatty liver disease: A systematic review and meta-analysis: HCC and steatosis or steatohepatitis. Neoplasia 2022 Aug;30:100809 [FREE Full text] [doi: 10.1016/j.neo.2022.100809] [Medline: 35636146]
46. Tiniakos D, Maurício J, Reeves H. Fatty liver disease and hepatocellular carcinoma: The pathologist's view. Adv Exp Med Biol 2018;1032:55-69. [doi: 10.1007/978-3-319-98788-0_4] [Medline: 30362090]
47. Shoelson SE, Lee J, Goldfine AB. Inflammation and insulin resistance. J Clin Invest 2006 Jul;116(7):1793-1801 [FREE Full text] [doi: 10.1172/JCI29069] [Medline: 16823477]
48. Tilg H, Diehl AM. NAFLD and extrahepatic cancers: have a look at the colon. Gut 2011 Jun;60(6):745-746 [FREE Full text] [doi: 10.1136/gut.2011.239392] [Medline: 21454382]
49. Khan S, Shukla S, Sinha S, Meeran SM. Role of adipokines and cytokines in obesity-associated breast cancer: therapeutic targets. Cytokine Growth Factor Rev 2013 Dec;24(6):503-513. [doi: 10.1016/j.cytogfr.2013.10.001] [Medline: 24210902]

Abbreviations

GATHER: Guidelines for Accurate and Transparent Health Estimates
HCC: hepatocellular carcinoma
IGF-1: insulin-like growth factor-I
MESH: medical subject heading
MRI-PDFF: magnetic resonance imaging proton density fat fraction
NAFL: nonalcoholic fatty liver
NAFLD: nonalcoholic fatty liver disease
NASH: nonalcoholic steatohepatitis
NOS: Newcastle Ottawa Scale
PICOS: population, intervention, comparator, outcome, study
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by A Mavragani; submitted 30.06.22; peer-reviewed by MZ Johari, L Tetlow, C Sun; comments to author 06.02.23; revised version received 22.02.23; accepted 19.04.23; published 19.07.23.

Please cite as:
Muhamad NA, Maamor NH, Leman FN, Mohamad ZA, Bakon SK, Abdul Mutalip MH, Rosli IA, Aris T, Lai NM, Abu Hassan MR
The Global Prevalence of Nonalcoholic Fatty Liver Disease and its Association With Cancers: Systematic Review and Meta-Analysis
Interact J Med Res 2023;12:e40653
URL: https://www.i-jmr.org/2023/1/e40653
doi:10.2196/40653
PMID:37467012

Review

Improving Predictability and Effectiveness in Preventive Digital Health Interventions: Scoping Review

Keld Pedersen[1], PhD; Bjarne Rerup Schlichter[1], PhD

Information Systems, Department of Management, Aarhus University, Aarhus C, Denmark

Corresponding Author:
Keld Pedersen, PhD
Information Systems
Department of Management
Aarhus University
Bartholins Allé 14
Aarhus C, 8000
Denmark
Phone: 45 51157818
Email: keldp@mgmt.au.dk

Abstract

Background: Lifestyle-related diseases caused by inadequate diet and physical activity cause premature death, loss of healthy life years, and increased health care costs. Randomized controlled trial (RCT) studies indicate that preventive digital health interventions (P-DHIs) can be effective in preventing these health problems, but the results of these studies are mixed. Adoption studies have identified multiple factors related to individuals and the context in which they live that complicate the transfer of positive results from RCT studies to practical use. Implementation studies have revealed barriers to the large-scale implementation of mobile health (mHealth) solutions in general. Consequently, there is no clear path to delivering predictable outcomes from P-DHIs and achieving effectiveness when scaling up interventions to reduce health problems in society.

Objective: This research aimed to expand our understanding of how to increase the outcome predictability of P-DHIs by focusing on physical activity and diet behaviors and amplify our understanding of how to improve effectiveness in large-scale implementations.

Methods: The research objective was pursued through a multidisciplinary scoping review. This scoping review used the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) as a guide. A comprehensive search of Web of Science and PubMed limited to English-language journal articles published before January 2022 was conducted. Google Scholar was used for hand searches. Information systems theory was used to identify key constructs influencing outcomes of IT in general. Public health and mHealth literature were used to identify factors influencing the adoption of, outcomes from, and implementation of P-DHIs. Finally, the P-DHI investment model was developed based on information systems constructs and factors from the public health and mHealth literature.

Results: In total, 203 articles met the eligibility criteria. The included studies used a variety of methodologies, including literature reviews, interviews, surveys, and RCT studies. The P-DHI investment model suggests which constructs and related factors should be emphasized to increase the predictability of P-DHI outcomes and improve the effectiveness of large-scale implementations.

Conclusions: The research suggests that outcome predictability could be improved by including descriptions of the constructs and factors in the P-DHI investment model when reporting from empirical studies. Doing so would increase our understanding of when and why P-DHIs succeed or fail. The effectiveness of large-scale implementations may be improved by using the P-DHI investment model to evaluate potential difficulties and possibilities in implementing P-DHIs to create better environments for their use before investing in them and when designing and implementing them. The cost-effectiveness of large-scale implementations is unknown; implementations are far more complicated than just downloading and using apps, and there is uncertainty accompanying implementations given the lack of coordinated control over the constructs and factors that influence the outcome.

(Interact J Med Res 2023;12:e40205) doi:10.2196/40205

KEYWORDS

mobile health; mHealth; digital interventions; adoption; implementation; prevention; physical activity; diet; mobile phone; scoping review; review

Introduction

Lifestyle-related diseases caused by inadequate diet and physical activity (PA) are a major problem in many societies, resulting in premature death, loss of healthy life years, and increased health care costs [1]. Facilitated by the widespread adoption of smartphones and wearables, preventive digital health interventions (P-DHIs) can present a more cost-effective approach to reach larger populations than traditional approaches [2,3]. However, randomized controlled trials (RCTs), adoption research, and implementation research indicate that predictable, positive outcomes from P-DHIs and the large-scale implementation of these solutions are difficult to achieve. Reviews of RCTs on P-DHIs that focus on PA and diet have revealed mixed results. Studies have reported no outcomes [4,5], small outcomes [6,7], outcomes that diminish over time [8], limited evidence for positive outcomes [9-11], positive outcomes for some individuals in some settings [12], mixed outcomes [13-15], promising results [16,17], and effectiveness of P-DHIs [18-22]. Beyond RCT studies, adoption research [23-26] indicates a range of factors related to the technology, individuals, and the context in which they live that complicates the transfer of results achieved in RCT studies to other persons in general. Furthermore, implementation research [27,28] identifies implementation barriers in health care organizations and society that hinder large-scale implementation of mobile health (mHealth) solutions, including the P-DHIs studied in this review.

Although low methodological quality may account for some of the uncertainty regarding the outcomes reported by RCT studies [14], the mixed results and adoption and implementation difficulties are unsurprising from an information systems perspective. We know that organizations experience quite different outcomes when investing in similar IT as the outcomes depend on many factors other than the IT [29]. Richardson and Zmud [30] emphasize that “The salient question, then, is not ‘Does IT pay off?’ but rather ‘Under what conditions does IT pay off?’” This is the core question behind this research as well. Specifically, this study investigated how to increase the predictability of outcomes of P-DHIs focusing on PA and diet behaviors and how to improve the effectiveness of large-scale implementations of P-DHIs.

This research was conducted as a scoping review as evidence regarding how to improve predictability and effectiveness from large-scale implementations of P-DHIs is unclear and a broad multidisciplinary understanding of this issue is needed. Investigating these issues is crucial for successfully exploiting P-DHIs as part of large-scale public health initiatives that demand both predictability and effectiveness. The development of the P-DHI investment model addresses these issues. The model is based on general constructs from information systems theory known to strongly influence outcomes from IT investments and factors closely linked to the specific use of IT in P-DHIs. The factors studied in this review relate to the adoption of P-DHIs, behavior change supported by P-DHIs, and the implementation of P-DHIs in society, thereby influencing P-DHI outcomes.

The concept of *health care organization* in this study references a wide variety of organizations, including public and private sector organizations at national, regional, and community levels that are engaged in public health and provide preventive initiatives.

The concept of *community* refers to both the physical community in which a person lives and their social community, which may include web-based social networks established through social media applications.

Tailoring refers to the process of adapting a P-DHI to the specific context (including specific persons) for which its use is intended. Similar research also uses *customization*, *individualization*, and *personalization* to name this process. *Tailoring* is used in this paper as it is a broader concept.

The study of P-DHIs in this review includes the use of smartphone apps and wearables as a key component, as well as additional resources accessed through the apps (eg, web-based social networks with other persons facing similar health-related concerns or knowledge provided by health care experts). A P-DHI is perceived as consisting of both IT (the IT investment) and additional investments (the non-IT investments) made to implement and benefit from the IT.

Methods

Overview

This research is based on a systematic multidisciplinary scoping literature review [31,32] using the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) as a guide (Multimedia Appendix 1). The method used in this review included the steps outlined in Textbox 1.

Textbox 1. Method used in this review.

Identifying relevant literature streams

- Given the multidisciplinary nature of the factors influencing the outcomes of preventive digital health interventions (P-DHIs; eg, self-efficacy [33], software quality [34], and factors in the context [35]), it was appropriate to explore a broad range of literature streams.

Identifying articles within the literature streams

- This was done primarily by identifying high-impact theoretical models and literature reviews and secondarily by identifying individual empirical studies.
- Analyzing articles, coding, and categorizing the constructs and factors from the articles and developing the P-DHI investment model

Identifying Relevant Literature Streams

The choice of literature streams and articles was guided by the concerns outlined in Textbox 2.

Textbox 2. Concerns that guided the choice of literature streams and articles.

Which overall constructs influence the outcomes of using IT in general?

- The information systems literature focusing on value creation from IT was reviewed to identify these constructs.

Which factors influence the outcomes of prevention initiatives in general?

- To address this question, articles describing the most frequently used theoretical models in the public health intervention literature were reviewed.

Which specific factors influence preventive digital health intervention (P-DHI) outcomes?

- The articles included to address this question related to (1) adoption (factors influencing the degree to which persons adopt P-DHIs), (2) health outcome (factors related to P-DHI effectiveness in terms of influencing behavior and improving health), and (3) implementation (factors influencing the degree to which it is possible to implement P-DHIs in health care organizations and society).

Identifying Articles: Eligibility Criteria

Textbox 3 presents the eligibility criteria for articles in the literature streams.

Textbox 3. Eligibility criteria.

General eligibility criteria for all literature streams

- Peer-reviewed journal articles
- Articles written in English
- Articles published before January 2022

Eligibility criteria for information systems articles

- Review articles (including models based on reviews) focusing on IT business value creation
- Articles that identify constructs that influence outcomes of IT in general
- Articles focusing specifically on technologies or specific industries not related to the research question were excluded.

Eligibility criteria for theoretical models used in public health

- Articles describing the most frequently used theories to research health-related behaviors within public health
- Articles that identify factors that influence health-related behaviors and behavior change
- In 2015, the following were identified as the most frequently used theories [36]: the transtheoretical model of change (used in 91/276, 33% of the identified articles), the theory of planned behavior (36/276, 13%), social cognitive theory (29/276, 10.5%), the Information–Motivation–Behavioral Skills Model (18/276, 6.5%), the Health Belief Model (9/276, 3.3%), self-determination theory (9/276, 3.3%), the Health Action Process Approach (8/276, 2.9%), and social learning theory (6/276, 2.2%). Even though the socioecological model is not among the most frequently used models, it was included as it offers insight into the relationship between the context, behaviors, and health not provided by the other theories.

Eligibility criteria for mobile health (mHealth) articles

- Articles identifying factors influencing the adoption of preventive digital health interventions (P-DHIs)
  - Reviews and empirical articles focusing on the adoption and use of P-DHIs, including interventions focusing on physical activity (PA) and diet
- Articles identifying factors that influence P-DHI health outcomes in terms of influencing behavior and health
  - Reviews of mHealth articles focusing on outcomes reported in PA or diet randomized controlled trial studies
  - Articles that identified factors influencing the outcome in terms of behavior change and health (such as the inclusion of behavior change techniques in the design)
  - Articles suggesting standards and taxonomies describing factors that influence the outcome in terms of behavior change and health
- Articles identifying factors influencing the large-scale implementation of P-DHIs
  - Only review articles focusing on the implementation of mHealth were included. Given that many of the barriers are assumed to be independent of the specific purpose of the mHealth solution, this literature review included articles that addressed mHealth implementation in general as well as articles specifically focusing on prevention related to diet and PA.
- Articles focusing on very specific issues, such as specific diseases (eg, sexual health), a very specific geographical area, or a very specific technology (eg, blockchain), were excluded.

Identifying Articles: Search Strategy

Search strategies for each of the literature streams were developed using search strings based on keywords, as described in Textbox 4.

Web of Science was used because of the multidisciplinary nature of the literature review. PubMed was used to specifically search for mHealth articles, thereby removing the risk of relevant mHealth articles not being found through Web of Science. Google Scholar was used to identify highly cited theoretical models used in public health (Textbox 4) based on the study by Davis et al [36]. The search strings used in Web of Science and PubMed were tested in pilot searches. Keywords were added and removed to determine whether a broader search would include relevant articles that were not identified using narrower search strings. For example, “fitness app” was added to include a higher number of relevant articles. In the mHealth adoption search string, the term “review” was removed as including this keyword excluded significant insights in this literature stream.

Textbox 4. Search strings.

Information systems articles

- Search string: TS=((“it” OR “information technology” OR “is” OR “information system”) AND “business value” AND review)
- Both “IT” and “IS” (information system) were included as especially older articles use the concept of “information system,” not “information technology.”
- Database: Web of Science

Theoretical models used in public health

- Keywords: “Transtheoretical Model of Change,” “Theory of Planned Behavior,” “Social Cognitive Theory,” “Information-Motivation-Behavioral-Skills Model,” “Health Believe Model,” “Self-determination Theory,” “Health Action Process Approach,” “Social Learning Theory,” “Socio-Ecological Model”
- Database: the keywords were used in individual searches in Google Scholar to identify highly cited articles describing the models.

Mobile health articles

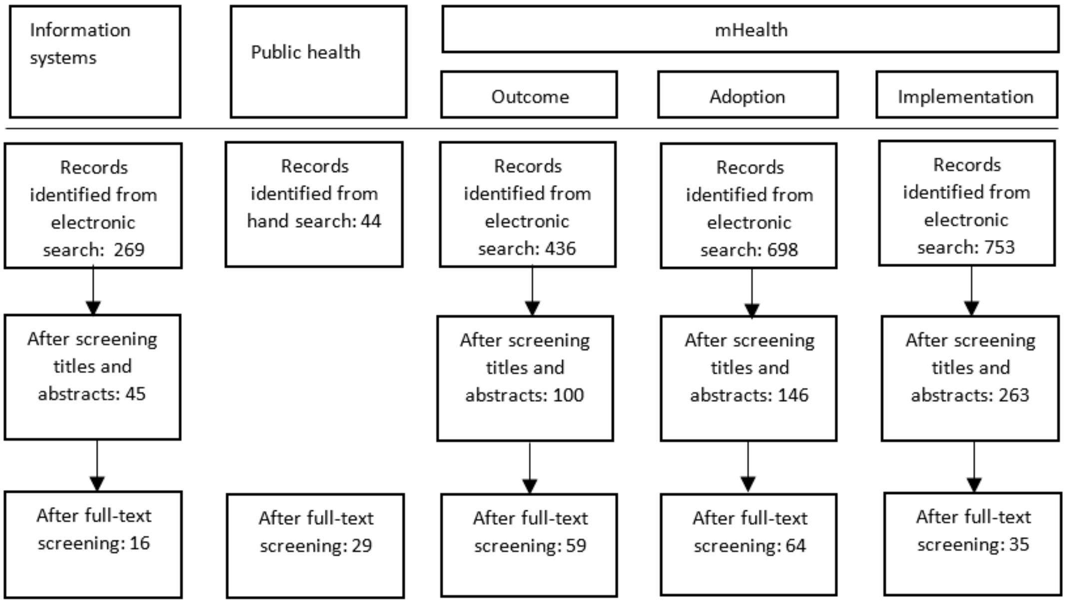
- Articles identifying factors that influence adoption and use of preventive digital health interventions (P-DHIs)
  - Search string: TS=((mHealth OR m-health OR “mobile health” OR smartphone OR “mobile app*” OR “mobile application*” OR “fitness app” or “diet app”) AND (“physical inactivity” OR overweight OR obesity OR nutrition OR diet OR “physical activity” OR fitness OR prevent* OR “chronic disease”) AND (adoption OR “technology acceptance model” OR TAM* OR “unified theory of acceptance” OR UTAUT* OR “use of technology” OR “IS success model”))
  - This search string did not include the concept “review” as initial searches using this concept returned too few articles and left out significant contributions.
- Articles identifying factors that influence P-DHI outcomes in terms of influencing behavior and health
  - Search string: TS=((mHealth OR m-health OR “mobile health” OR smartphone OR “mobile apps” OR “mobile applications” OR “fitness app” or “diet app”) AND (prevent* OR “behavioral change” OR “behavior change”) AND (“physical inactivity” OR overweight OR obesity OR nutrition OR diet OR “physical activity” OR fitness) AND (review))
- Articles identifying factors influencing the large-scale implementation of P-DHIs
  - Search string: TS=((mHealth OR m-health OR “mobile health” OR “fitness app” or “diet app”) AND implement* AND review)
  - The search criteria were broader than in previous searches as it was assumed that many implementation issues are general and not specific to apps focusing on diet and physical activity.
  - Database: Web of Science and PubMed. Searches in Web of Science used TS (topic). Searches in PubMed used “Title/Abstract”. The same keywords were used in both databases.

Identifying Articles: Screening and Eligibility

The search results from Web of Science and PubMed were downloaded to Microsoft Excel (Microsoft Corp). There was a substantial overlap between the mHealth search results from Web of Science and PubMed. The number of records identified from the electronic search reported in Figure 1 was after the removal of duplicates. Both authors screened the articles independently of each other, and subsequently, the results were compared and discussed.

Titles and abstracts were screened against the eligibility criteria (Textbox 3). The excluded articles were labeled with reasons for exclusion. Only articles clearly outside the scope of interest were excluded in this step. The inclusion and exclusion criteria were updated during the screening process. A total of 554 articles were extracted for full-text screening. The same inclusion and exclusion criteria were used for full-text screening. During full-text screening, additional papers were included based on a backward search using Google Scholar. A total of 203 articles met the inclusion criteria. Several articles identified and described the same factors. In particular, adoption research reported many identical factors as the studies were based on the same information system adoption models. For example, social influence (subjective norms) was emphasized in many adoption research papers. To reduce the number of references, not all articles that emphasized, for example, social influence were included in the references. The same inclusion and exclusion criteria were used for title and abstract and full-text screening. Owing to the nature of the review, bias concerns were not used to exclude articles. Figure 1 illustrates the screening process.

Figure 1. The screening process. The mobile health (mHealth) articles identified from the electronic search contained some duplicates across the 3 searches. The numbers in the after screening titles and abstracts section in the diagram are the numbers of unique articles.



Data Charting Process: Analyzing Articles, Extracting, and Structuring Constructs and Factors

The identified articles were imported into NVivo (QSR International) and analyzed, and the constructs and factors identified in the articles were coded in an iterative process.

The first step was to use information systems theory to identify constructs that influence outcomes from investments in IT in general. For example, "context" was identified as a key construct.

The second step was to identify and categorize factors from the public health and mHealth literature influencing (1) adoption, (2) outcomes (eg, behavior change and health) specifically from P-DHIs, and (3) the possibilities for large-scale implementation. For example, multiple sources within the public health and mHealth literature emphasized the importance of "social influence" for both adoption and behavior change.

In the third step, the factors from the public health and mHealth literature were categorized using the key construct identified in information systems theory—IT and non-IT investments establishing P-DHIs, the context in which P-DHIs are implemented and used, and the lag effects that influence when outcomes from P-DHIs are realized. For example, "social influence" was categorized as a part of the "context." Additional lower-level constructs were included as well (eg, specific parts of the context)—information systems theory is concerned with organizational processes, not processes in a person's life, and the context for health-related behavior change includes a person's changing behavior and the community in which they live.

Finally, the P-DHI investment model was developed based on the general information systems constructs and related factors from the public health and mHealth literature.

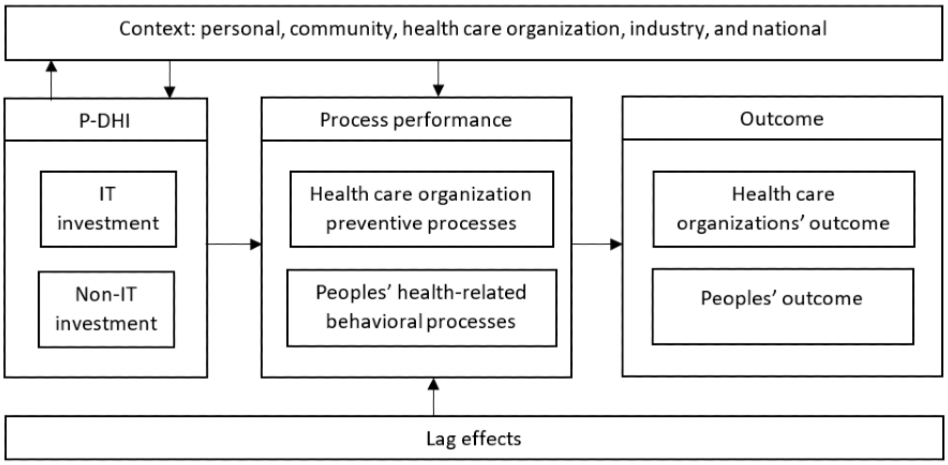
One coder (the first author) coded all the articles, making it easier to ensure consistency but introducing validity and reliability concerns. Another coder (the second author) independently coded approximately 10% of the articles (22/203, 10.8% of the articles), and the coding was subsequently compared to reduce validity and reliability concerns. Only minor discrepancies in coding were identified, discussed, and resolved. The level of uncertainty and subjective interpretation when coding the text from these articles was low. Regarding the information systems articles, constructs influencing the outcome were coded using the concepts (eg, lag effects) in the articles. Regarding the public health and mHealth articles, the factors in these models (eg, perceived self-efficacy) influencing behavior change and health were coded using the concepts from the articles. During the coding process, fewer and fewer new concepts were added because of the conceptual consensus across the articles. Subsequently, additional categories were included to group the factors and establish a more general understanding (eg, some factors related to the capabilities of individuals).

Results

Overview

The results in terms of the P-DHI investment model are illustrated in Figure 2. The model illustrates how outcomes from using P-DHIs are created and, consequently, how predictability and large-scale effectiveness might be improved. First, the P-DHI investment model is explained. Second, the constructs and factors in the model are presented.

Figure 2. The preventive digital health intervention (P-DHI) investment model.



The Model

The P-DHI investment model is based on the constructs used in information systems theory to explain outcomes from the use of IT [29,37,38], specifically based on the model by Schryen [39]. The basic logic is that an outcome is created, in this case, for example, weight loss, by changing the performance of the processes involved. In the case of weight loss, it is processes in health care organizations delivering preventive health care services and processes in terms of individual health-related behaviors associated with weight loss. Changes are achieved using a combination of IT and non-IT investments. IT investments are investments in apps and wearables integrated with health care systems such as a database collecting data from individual P-DHI users. Non-IT investments are additional investments in changes in health care organizations, the services they provide, and society in general necessary for delivering prevention using the P-DHIs, as well as additional investments made by P-DHI users to change behavior, such as investments in fitness equipment, time, and energy needed for behavior change.

The P-DHI should be tailored to match the characteristics of the context, and the context may support or complicate process changes within health care organizations and in P-DHI users' lives. For example, health care professionals might resist using this kind of technology, and P-DHI users might experience a lack of social support or find it difficult to change their behaviors for other reasons. Consequently, the P-DHI might provide support for overcoming barriers and exploiting resources in the context. Lag effects, for example, learning how to use P-DHIs, can delay how long it takes to create positive outcomes. The following sections explain the factors in various parts of the model that influence the outcome.

P-DHIs: IT Investment

The 5 factors in Textbox 5 are important as they are closely related to the improvement of people's health-related behaviors and, thereby, the outcome.

Textbox 5. IT investment factors that influence the outcome of preventive digital health interventions (P-DHIs).

1. Complying with software quality requirements: the degree to which the P-DHI app complies with basic software quality requirements and is integrated with other relevant IT systems
2. Tailored to the context: the degree to which the P-DHI app is tailored to an individual's context and personal needs
3. Developing personal capabilities: the degree to which the P-DHI app supports the development of the personal capabilities needed for behavior change using a P-DHI app
4. Behavior change support: the degree to which the P-DHI app provides theory-based behavior change support
5. Provides additional personal help: the degree to which the P-DHI app provides additional help through access to web-based social networks and health care professionals during the behavior change process

The logic behind the first 5 factors is that, if a P-DHI app lacks basic software quality characteristics, such as being easy to use, it will reduce adoption and use, and to provide relevant support for the individual users, it needs to be tailored to the context in which they live and their individual needs. Using P-DHIs to change health-related behaviors requires behavior change capabilities in general as well as capabilities related to specific behaviors (eg, exercise-related capabilities) and capabilities in the use of this kind of technology. If a person lacks these capabilities, they need support from the P-DHI to develop them. The actual behavior change process is best supported using behavior change theory, such as the transtheoretical model of health behavior that supports the entire behavior change process and behavior change techniques (BCTs) that support individual steps in this process. During the behavior change process, there might be a need for web-based personal help from health care professionals and peers taking part in similar change processes. The remainder of this section describes the theoretical foundation behind these 5 factors.

The software quality requirements can be summarized as being easy to use [21,26,34,40-46], engaging and interesting

[34,47-49], trustworthy (eg, in terms of being credible [34,50] and secure [26,27,46,51]), technically robust [26,52], and affordable [27,47,53-57]. The strong focus on usability is not surprising given that perceived ease of use [58] is influential on a person's intention to adopt P-DHI apps [59]. However, the review by Milne-Ives et al [11] found no relationship between usability and app effectiveness, that is, although usability ratings were high in the reviewed RCT studies, app effectiveness remained low. Being easy to use is important for adoption and use but not enough to create positive outcomes.

P-DHI apps should be integrated with dependable and locally adapted food databases [56] and other relevant apps [26]. In particular, mHealth implementation research emphasizes the need for integrating apps with health care information systems used by health care professionals [28,35,60-66].

Similar to any other software, P-DHI apps should be user-centered and tailored to the context [22,26,40,42,46,67-77]. For example, we know that P-DHI app use of BCTs should be tailored as their efficacy depends on users' sociodemographic [10,78] and psychosocial factors [10]. However, tailoring includes adapting the P-DHI app to the entire context. Overcoming barriers [79-84] and exploiting specific resources in the context [80,85-88] are vital parts of behavior change. The following section about the context describes contextual factors that should be considered for tailoring. Being user-focused and developing tailored P-DHI apps seems quite complicated compared with the development of IT systems to be used within organizations because of the number of users, variations across users, and variations in users' contexts.

Behavior change requires the capabilities to change and perform new behaviors and belief in these capabilities (perceived self-efficacy) [33,64,80,89,90]. This review identified 3 different types of capabilities: capabilities related to behavior change in general, such as capabilities to self-regulate [91]; capabilities specifically related to the execution of new behaviors (eg, to exercise or prepare healthy food) [90]; and capabilities to use technology as a part of health-related behavior change, for example, in terms of eHealth literacy [4]. Consequently, P-DHI apps should support the improvement of people's capabilities for health-related behavior change as well as their level of perceived self-efficacy if needed. This can be accomplished in many different ways, such as providing guidance [44,83,84], opportunities for practice [83,84], experience-based learning [33,75,81,92], social models [88,91,92], and mastery experiences (graded tasks) [83,84,91]. Wester et al [4] suggested identifying and improving low eHealth literacy before the actual behavior change intervention to make interventions more successful for people with low socioeconomic status.

The core of a P-DHI app is, of course, the features that support the behavior change process. Research indicates that theory-based P-DHIs are more effective [26,42,46,81]. This review identified 4 types of behavioral theories that seem especially relevant. The first type of theory is the models (transtheoretical model of change [89] and Health Action Process Approach [79]) that cover the entire behavior change process and divide the process into stages; the second type is the taxonomies that describe the specific BCTs (eg, goal setting) to be used during the stages [84]; and the third type of theory focuses on specific factors, such as perceived self-efficacy, that have a large impact on the possibilities for successful behavior change [33]. The last kind of theory, the socioecological models, focuses on how the context influences health-related behaviors, behavior change, and health. For example, how the physical environment influences PA behaviors [85,93-95] and how the food environment influences diet behaviors [96,97]. The implication of being "theory-based" is that a P-DHI app should provide different kinds of support depending on the person's current stage in the behavior change process and that the functionality should support the use of BCTs, attempt to influence the key factors such as perceived self-efficacy positively, and be context aware and help users exploit resources and overcome barriers in the context. There is some uncertainty about which BCTs to use [8,10,20], how best to combine them [8,10], and how best to implement them within apps [11,98]. Some RCT reviews [11,21] reported no association between the number of BCTs applied in apps and app effectiveness, whereas another review of P-DHIs focusing on PA reported that initiatives failed for people with low socioeconomic status irrespective of the BCTs used [4]. The inclusion of BCTs in apps is important but not sufficient to achieve effectiveness [11].

During behavior change, 2 different sources might provide additional support that complements the support provided by the P-DHI app functionality—support from social networks and health care professionals. Providing access to social networks through P-DHIs is generally perceived as beneficial [15,99]. Social networks potentially provide access to many resources that we know influence the adoption of P-DHIs, continued use of P-DHIs, and behavior change—social models [84,88,92], social influence [26,53,59,84,100-102], social support [15,80,85,88,89,103], social comparisons [12,83,84], and social connectedness and a feeling of relatedness to a community [45,104,105]. Although social networks can provide access to many resources and reportedly increase engagement with PA interventions [106-108], research findings [16,26,44,109-111] regarding the benefit of social networks and features are mixed. Some discrepancies may be attributed to differences in social comparison orientation [109]. Koönig et al [26] reported that social features serve to motivate some, whereas participation in competitions or observing others' success is demotivating to others. Excessive competition in social networks [15] and lack of alignment of behavioral goals between the person attempting to change their behavior and other participants in the network [16] can negate the positive effect of social networks.

Combining P-DHI app use with support from health care professionals seems to offer greater effectiveness than stand-alone interventions [10,51,64,112], and therefore, it may be beneficial to provide some degree of access to health care professionals through the P-DHI. Some uncertainty remains regarding the relative contribution of coaching delivered by health care professionals and the app itself [15,18]. Using apps as stand-alone solutions increased PA in some studies but failed to do so in others [15].

P-DHIs: Non-IT Investment

Generally, non-IT investments are investments in organizational changes that are required to benefit from IT investments [39,113]. This study used a broader understanding of non-IT investments as the purpose of P-DHIs is to change individuals' behaviors. The first 3 factors in Textbox 6 influence the performance of health care organizations' preventive processes, whereas factor 4 influences the improvement of people's health-related behaviors.

Textbox 6. Non-IT investment factors that influence the outcome of preventive digital health interventions (P-DHIs).

1. Integrating the use of P-DHIs into health care organizations: the degree to which the use of P-DHIs is properly integrated into health care organizations' preventive processes
2. Recruiting and engaging P-DHI users: the degree to which health care organizations succeed in recruiting and engaging P-DHI users
3. Providing additional services for P-DHI users: the degree to which health care organizations provide additional services, such as workshops that support behavior change
4. P-DHI users' investments: the degree to which P-DHI users invest the required resources in terms of engagement and motivation, time, and money, among others, for changing their behavior

The logic behind the factors in Textbox 6 is that, even if a highly sophisticated P-DHI app is bought or developed, it does not produce positive outcomes for society before it is properly integrated into health care organizations' processes focusing on preventing chronic diseases in the population. Furthermore, it requires investment in targeted campaigns that recruit and engage people in society who are at risk of developing chronic diseases caused by lifestyle-related problems. Finally, positive outcomes require more from P-DHI users than downloading an app; they require considerable engagement, time, and money from the user. The remainder of this section describes the theoretical foundation behind the aforementioned 4 factors.

Similar to any other technology, the integration of P-DHIs in health care organizations requires organizational changes to succeed. The literature specifically on the implementation of P-DHIs is very sparse, but the literature on mHealth implementation in general emphasizes the same kind of changes as the implementation of IT in general (eg, new incentives [43,62], policies [43,67], ways of working [28,60-63,114] and collaborating [40,54], and new capabilities [27,63]). Integration also includes establishing the facilities needed for providing training and support for using P-DHI apps to change behavior [49,66,102,115,116]. The suggested ways to implement these organizational changes are also similar to what we know from the implementation of other types of IT systems—formulating implementation strategies [62,67], managing organizational resistance [67], and training internal users [40,61,62,67,117].

In addition to these internal changes, implementation involves choosing and using engagement and recruitment strategies, such as promotion and marketing campaigns and clinical endorsement [41]. Both the transtheoretical model of change [89] and the Health Action Process Approach [79] distinguish between initial stages, in which potential users of a P-DHI do not even acknowledge that they need to change their behavior or are uncertain about engaging in behavior change even though they acknowledge the need, and later stages, in which actual behavior change actions are executed and potential users attempt to maintain behaviors. Therefore, different campaigns are needed for potential users depending on their stage in the process.

Research indicates that multicomponent interventions involving additional services are generally more effective than stand-alone app interventions [10,18]. The RCT reviews studied in this paper described additional services in the included RCT studies, such as workshops and group sessions [5,8,18], individual education and coaching sessions [4,8,15,18,19], mindfulness sessions [8], personalized feedback from health care professionals [19,21], and motivational interviewing [19]. Kozik et al [118] suggested that training of individual persons should be tailored to the context in terms of addressing inequity issues through the provision of special onboarding sessions for advanced-age and low-education populations. This kind of personal support from health care professionals creates scalability problems and may make P-DHIs less attractive from an economic perspective.

We know little about the non-IT investments for P-DHI users and how they affect the outcome. Collecting data on this type of investment is not addressed in the P-DHI research reviewed in this study. From the behavior change theory and models [79,89] and the BCT taxonomy [84] that describes the behavior change process and specific activities (eg, self-monitoring and regulation), we can deduce that P-DHI users must invest considerable time, energy, mental and physical resources, and money to motivate themselves, plan and execute behavior change activities, bounce back from setbacks, and rearrange their life and context so that they support the new behaviors.

The Context

Generally, the same kind of IT system can lead to different results in different contexts [29,37-39], and this also applies to P-DHIs. The socioecological theory emphasizes how the context influences behaviors and behavior change [88,119], the behavior change wheel [120] emphasizes the importance of persons' opportunities for successful behavior change, and Han and Lee [22] report that the use of P-DHIs in different situations for different persons may lead to varied outcomes.

A wide range of contextual factors influences how P-DHIs should be designed and tailored, and the factors in Textbox 7 influence the outcome of P-DHIs.

Textbox 7. Contextual factors that influence the outcome of preventive digital health interventions (P-DHIs).

1. Individual users' characteristics: the degree to which the P-DHI matches individuals' current stage in the behavior change process and their characteristics and influences these characteristics positively
2. Community-level characteristics: the degree to which the P-DHI supports behavior change in the P-DHI user's specific community
3. Health care organization readiness: the degree to which the necessary resources are in place to support the implementation of P-DHIs in health care organizations
4. Health care sector requirements: the degree to which P-DHIs comply with core health care sector requirements
5. National-level support for the use of P-DHIs: the degree to which national-level funding, policies and regulations, and technological infrastructures support the use of P-DHIs

The logic behind these 5 factors is that P-DHI users who are different and live in different communities that provide different barriers and possibilities for behavior change need diverse kinds of support. Even if the P-DHI matches these characteristics and provides the right support, success still depends on health care organizations being ready to implement P-DHIs. Furthermore, to become a part of health care services, P-DHIs need to comply with the formal requirements that we expect from health care services, such as being evidence-based, and large-scale implementations require national-level support and sufficient technological infrastructures in society. The remainder of this section describes the theoretical foundation behind these 5 factors.

P-DHI users at dissimilar stages in the behavior change process need diverse kinds of support. The main difference is between the initial stages, in which potential users are developing intentions to change, and later stages, in which they attempt to change or maintain the behavior [79,89]. Some of the personal characteristics that the P-DHI needs to be designed for and tailored toward are unmodifiable in the sense that they cannot be changed as a part of the behavior change process. Demographics and socioeconomic status influence individuals' acceptance of and use of P-DHIs [17,26,42,55,57,118,121-124] and the outcome of interventions [4,12,125,126]. The degree to which a P-DHI app is consistent with personal values [116] and culture [42,121,127] also influences individuals' acceptance.

Other characteristics are modifiable, and the P-DHI should attempt to influence them to improve the possibilities of successful behavior change. These characteristics are related to the users' intentions for behavior change, their capabilities for behavior change, and their situation in life. Potential P-DHI users' level of health consciousness [128-130], perception of their own health and health risks [80,82,124], expectations regarding the outcome of changing behaviors [80], attitudes toward new behaviors [131], and self-efficacy beliefs have a strong influence on intentions for behavior change. The required capabilities were described in the previous section. Current life situations include the degree to which P-DHI users face issues such as stress [91], feeling tired [91], being depressed [91], temptations to deviate from new behaviors [89], lack of time [26,44,85], or competing priorities that make behavior change difficult [41].

The resources available in a specific context have a large impact on health-related behavior, behavior change, and health [80,85-88]. For example, we know that access to community-level health care resources is important [86].

The social context in terms of social influence is important for the general acceptance of P-DHI apps [23], the intention to adopt these apps [26,132,133], the intention to use these apps [25,49,53,100,134], the actual use of these apps [103], and the continued use of these apps [24,44,50,111,135]. The physical context influences both food intake [96,97] and PA [85,93-95,136]. As previously described, P-DHIs should help P-DHI users overcome barriers and exploit community-level resources. Tonkin et al [76] suggested that P-DHI apps could provide information about local stores offering healthy food options and assist in creating a healthier food environment, might help find appropriate fitness partners [77], and generally help rearrange the context to support new healthy behaviors [81].

As previously described, under non-IT investments, the widespread use of P-DHIs requires changes to preventive processes within health care organizations. We know little specifically about the implementation of P-DHIs, but we do know that the successful implementation of mHealth in general requires adequate management resources [62], financial resources [60-63,67,117], and IT resources [28,35,43,62,137]. Furthermore, different types of organizational capabilities (eg, project management capabilities) are needed when implementing these solutions [27,28,35,51,60,63,67,138]. One of the most important issues is health care professionals' attitudes toward these solutions [14,28,63], their outcome expectancies [14,40,43,62,66,139], their resistance to change [28,63,67], and their perception of the organization's readiness to use mHealth [62,138]. Some sources mention that difficulties in implementation can be attributed to health care organizations' relatively slow adoption of new technologies such as mHealth [67,118], their organizational culture [35,67] and norms [140], and frequent budget deficits [62].

P-DHIs should comply with the requirements posed by the health care sector in general. However, there is generally a lack of regulation (eg, Food and Drug Administration approval or other kinds of certification) of mHealth apps [63,65,118,141], there are differences in medical and clinical practices across state or country lines that can complicate the use of mHealth apps [60], and there is a lack of evidence regarding the effectiveness of mHealth in practical use that one would normally expect from elements in health care services [63].

Government support for the use of mHealth [35,40]; government involvement [27,57]; funding [27,40,57]; and mHealth policies, strategies, and guidance [27,40,57,61,63] are important for P-DHIs to be used as a central part of national public health initiatives. Furthermore, P-DHIs and other mHealth solutions require widespread access to mobile technology [27,35,40,41,43,74,118,122,138,142] and reliable technological infrastructures [14,27,43,54,63,137,138,142,143].

Outcomes, Process Changes, and Lag Effects

The outcomes for health care organizations and individuals are divided into 2 categories: tangible and intangible. The outcomes listed in this section are the possible outcomes mentioned in the reviewed literature. We know that P-DHIs in some cases change PA and diet behaviors [7,8,10,13,16-18,21,22], and we know that the behavior changes from using P-DHIs positively influence health [7,12,13,18,19]. On the basis of this review, little is known about the impact on intangible outcomes such as P-DHI users' capabilities for long-term health management. Similarly, we know little about the impact on health care organization outcomes (eg, in terms of reduced costs).

Tangible outcomes include improved health-related behaviors and health, other impacts on individuals (eg, improved convenience), changes to health care professionals' work (eg, workload), and impact on health care organizations (eg, improved cost-effectiveness). Improved health-related behaviors and health are, quite naturally, emphasized in the reviewed literature [6,11,14,21,22,117]. Health-related outcomes can also involve improved appearance, regaining past fitness, or complying with job requirements [111]. There are also more practical outcomes for individuals, such as easier access to health care [14,16,27,62,117], improved convenience [14,51] and communication [14,27,117], and lower costs for individuals using mHealth [27,117]. Although these outcomes are positive, P-DHI apps pose a risk of discriminating against people with low socioeconomic status [4,125,126] on the wrong side of the "digital divide" [42,126,144,145]. There are also concerns about P-DHI apps leading to unhealthy behaviors such as food choices based on ease of registration within the app, extreme calorie restriction, and eating disorders [26].

Health care organizations might experience improved cost-effectiveness [14,16] by reaching more persons at a lower cost [59], using more scalable health care services [16], and improving patient care [62]. P-DHIs may influence several aspects of the health care professional experience both positively and negatively. The reviewed literature mentioned aspects such as workload [7,14,43,62,66,139], record maintenance [14], job security [40,62], efficiency, job autonomy, and effectiveness [62].

Intangible outcomes include increased awareness about health, increased motivation for changing health-related behaviors, external acknowledgment, psychological development and well-being in general, and improved health and behavior change capabilities. There are several positive intangible outcomes that may increase individuals' possibilities for long-term health outcomes: increased awareness of health-related issues [26,74,82,89,91,139], increased motivation for changing health-related behaviors [12,26], acknowledgment from social networks [91], psychological development and well-being [12], increased capabilities (eg, more knowledge about health [26,80,146] and behavior change–related skills [26]), and improved self-efficacy [7,12]. Furthermore, users may experience greater empowerment and improved daily life autonomy [7]. On the negative side, attempts to change behavior can also lead to negative feelings such as guilt, disappointment, anxiety when failing, or feeling neurotic about one's own body image [26]. mHealth might increase social isolation among older adults, who might prefer direct in-person contact with health care professionals [17].

Owing to lag effects, outcomes from IT investments are generally not realized immediately [39]. IT investments can even lead to worse performance in the interim because of learning-by-doing effects [147]. Stephenson et al [8] found a decrease in behavior changes from P-DHIs, specifically reduced sedentary behavior (in RCT studies), over time, whereas Emberson et al [12] reported that, with regard to PA (in RCT studies), interventions of longer durations led to better results than those of shorter durations. The meta-analysis by Moönninghoff et al [148] found that, although the effects of PA interventions were maintained in the long term, the size of the effect diminished over time. Generally, we lack knowledge about the long-term effectiveness [6,13,148] and cost-effectiveness of P-DHIs that promote PA and sedentary behavior changes [6]. There is no research among the studies in this review specifically exploring lag effects in the context of time elapsed between the implementation of a large-scale P-DHI and changes to process performance in health care organization prevention processes and people's health-related behaviors being realized. In addition, there is no research explicitly addressing the factors that contribute to lag effects.

Using the Model in the Prevention of Lifestyle-Related Health Problems

The P-DHI investment model and its 14 factors can be used by health care organizations when considering, designing, and implementing P-DHIs.

When considering using a P-DHI, the P-DHI investment model can be used to perform an initial feasibility study to assess the likelihood that a P-DHI will succeed for a specific target group. Using the constructs and factors, it is possible to identify situations in which a P-DHI would most likely succeed or fail and what it would require from the P-DHI in terms of IT and non-IT investments to succeed. For example, success would be difficult in a situation in which the target group has a low socioeconomic status, has a personal life situation characterized by high levels of stress, lacks readiness for health-related behavior change, and inhabits communities that provide little support for healthy behaviors and behavior change, and in which health care organizations lack capabilities in providing services based on the use of P-DHIs and the technological infrastructures in society are unreliable. Textbox 8 describes key questions related to the constructs and factors in the P-DHI investment model that health care organizations should address when considering, designing, and implementing P-DHIs.

Textbox 8. Using the model in practice.

Outcome

- What kind of outcome do we want to achieve for the targeted persons (eg, reducing the risk of cardiovascular disease for a specific target group characterized by a high risk of developing cardiovascular disease)?
- What kind of outcome do we want to achieve for the involved health care organizations (eg, lowering costs and making prevention initiatives easier to access and more attractive for the target group)?

Context

- What are the major characteristics of the context?
- In what ways does the context support or hinder the target group's behavior change and process changes in health care organizations?
- Individual users' characteristics: the degree to which the preventive digital health intervention (P-DHI) matches individuals' current stage in the behavior change process and their characteristics and influences these characteristics positively
    - Do we attempt to support potential users who have little or no intention of changing their behavior?
    - Do we attempt to support potential users who have the intention but need support to successfully change their behavior?
    - Do we attempt to support potential users who have succeeded in changing their behavior but need support to maintain the new behaviors?
    - What characterizes the potential users, and how should we design the P-DHI to increase the likelihood of adoption and use? How might we support the development of personal characteristics (eg, their awareness about health) in ways that increase the possibilities for success?
- Community-level characteristics: the degree to which the P-DHI supports behavior change in the user's specific community
    - What characterizes the potential users' communities in terms of typical barriers and resources, and how might we support the users in overcoming barriers and exploiting resources?
    - How difficult is it going to be to achieve the behavior change–related outcome for the individual P-DHI users in this community?
- Health care organization readiness: the degree to which the necessary resources are in place to support the implementation of P-DHIs in health care organizations
    - Do we have the needed resources for implementing a P-DHI, or do we need to prepare and invest in specific resources before we invest in a specific P-DHI?
    - How difficult is it going to be to achieve the outcome for the health care organization?
- Health care sector requirements: the degree to which P-DHIs comply with core health care sector requirements
    - Which health care sector requirements do we need to comply with regarding regulation, medical practice, and being evidence-based?
    - How are we going to achieve compliance?
- National-level support for the use of P-DHIs: the degree to which national-level funding, policies and regulations, and technological infrastructures support the use of P-DHIs
    - What are the possibilities for funding?
    - Which policies and regulations (eg, the General Data Protection Regulation) do we need to comply with?
    - What characterizes the technological infrastructures that we rely on for the P-DHI, and which limitations and possibilities do they represent?

Changes

- What specific changes to the target group's behaviors represent the easiest way to successfully achieve individual P-DHI users' outcomes given the context characteristics?
- What specific changes to health care organization processes represent the easiest way to successfully achieve health care organizations' outcomes given the context characteristics?

Lag effects

- When can we realistically expect to experience outcomes from individual-level behavior changes, and which factors drive the lag effects?
- When can we realistically expect to experience outcomes from changes to health care organizations, and which factors drive the lag effects?

P-DHI IT investment

- Given the outcomes and changes that we attempt to achieve in these specific contexts, what are the key requirements for the P-DHI apps and how might we address these requirements?

- Complying with software quality requirements: the degree to which the P-DHI app complies with basic software quality requirements and is integrated with other relevant IT systems
  - Given the expected outcomes, the behavior changes that we are aiming for, and the requirements we can derive from the context characteristics, how should we define and fulfill the software quality requirements for this specific P-DHI app? For example, what does user friendly mean for this specific app when it is used by these specific users in this specific context?
- Tailored to the context: the degree to which the P-DHI app is tailored to individuals' context and personal needs
  - Given the expected outcomes, the behavior changes that we are aiming for, and the requirements we can derive from the context characteristics, how should we tailor this specific app to make it fit the individual users and their context? What can be achieved through static and dynamic tailoring?
- Developing personal capabilities: the degree to which the P-DHI app supports the development of the personal capabilities needed for behavior change using a P-DHI app
  - Given the expected outcomes, the behavior changes that we are aiming for, and the requirements we can derive from the context characteristics, how should we support the P-DHI users in improving relevant capabilities?
- Behavior change support: the degree to which the P-DHI app provides theory-based behavior change support
  - Given the expected outcomes, the behavior changes that we are aiming for, and the requirements we can derive from the context characteristics, which model (eg, the transtheoretical model), behavior change techniques, and other theories should we use as the foundation for the design of the P-DHI app? How could we use the theory in the best way?
- Provides additional personal help: the degree to which the P-DHI app provides additional help through access to web-based social networks and health care professionals during the behavior change process
  - Given the expected outcomes, the behavior changes that we are aiming for, and the requirements we can derive from the context characteristics, to what extent is personal help from health care professionals needed? How might we use web-based social networks to support the behavior change process? How might we minimize the costs?

P-DHI non-IT investment

- Given the outcomes and changes that we attempt to achieve in these specific contexts, what are the key requirements for the P-DHI non-IT investments and how might we address these requirements?
- Integrating the use of P-DHIs into health care organizations: the degree to which the use of P-DHIs is properly integrated into health care organizations' preventive processes
  - Given the outcomes and changes that we attempt to achieve in these specific contexts, what kind of non-IT investments do we need to integrate the P-DHI into the health care organizations' processes?
- Recruiting and engaging P-DHI users: the degree to which health care organizations succeed in recruiting and engaging users
  - Given the outcomes and changes that we attempt to achieve in these specific contexts, how might we recruit and engage potential P-DHI users?
- Providing additional services for P-DHI users: the degree to which health care organizations provide additional services, such as workshops that support behavior change
  - Given the outcomes and changes that we attempt to achieve in these specific contexts, what kind of additional services do we need to realize the outcomes?
- P-DHI users' investments: the degree to which users invest the needed resources in terms of engagement and motivation, time, and money, among others, for changing behaviors
  - Given the outcomes and changes that we attempt to achieve in these specific contexts, what and how much do we expect that the P-DHI users need to invest in terms of money, time, engagement, equipment, and other resources to succeed? Can we reduce these investments to make it easier for the P-DHI users?

Using the model, the actual design process starts with deciding which types of tangible and intangible outcomes for the target group and the health care organizations should be offered by this P-DHI (Textbox 8). For example, types of outcomes might be the improvement of individuals' long-term capabilities for managing their own health, helping individuals achieve short-term outcomes (eg, in terms of reduced body weight within a few weeks), or improving the cost-effectiveness of health care organizations. Deciding on the types of outcomes offered by the P-DHI and understanding how difficult they are to achieve requires insight into the context. Textbox 8 describes key questions that health care organizations need to consider regarding the different parts of the context, for example, the kind of barriers that P-DHI users might experience.

The next step is to identify the easiest means of achieving these outcomes by selecting which individual behaviors and internal processes in health care organizations to change and in which way. Some behaviors may be easier to change than others, and the same applies to organizational processes in health care organizations. The goal is to identify the path of least resistance, that is, identify the set of changes that might achieve these types of outcomes in the easiest way given the insights into individuals in the target group and the health care organizations. The more these behaviors and preventive processes vary across individuals and health care organizations, the more they will require in terms of tailoring possibilities.

The last step is designing the simplest P-DHI—consisting of both IT and non-IT investments—that might achieve these changes. When doing so, it is important to strike the most efficient and effective balance between IT and non-IT investments and consider the lag effects to develop a realistic expectation of when these changes might be accomplished.

The design of a P-DHI using the P-DHI investment model is more comprehensive than simply designing an app as the P-DHI contains both IT and non-IT investments. During the design process, the 5 factors related to the P-DHI app and the 4 factors related to the non-IT investments should be considered, and the design should comply with the requirements that can be deduced based on insights into the context in which it is to be used (Textbox 8).

The design of the non-IT investment includes designing the organizational changes in health care organizations to offer P-DHIs (eg, new policies and ways of working), the implementation process (eg, how to facilitate user acceptance within the organization), the recruitment strategies, and the design of additional services (eg, how to provide training and support). Furthermore, design also includes considerations regarding the personal non-IT investments needed from users for implementing, using, and benefitting from the P-DHI to change their behaviors, for example, when and how they will use the app (eg, how much and how they will use the resources provided by the app), how they will allocate the necessary time and resources and rearrange the immediate context to better support healthy behaviors, and how these investments might be reduced to make it easier to change behaviors.

The model can also be used to evaluate the difficulties and possibilities of implementing P-DHIs in relation to the various aspects of the context and create environments conducive to the use of P-DHIs before investing in a P-DHI as part of a public health initiative. This can be accomplished by reducing barriers and improving supportive resources in the context before the intervention, if possible, or by tailoring the P-DHI to help individuals overcome barriers and exploit resources throughout the intervention.

The model illustrates the complexity and uncertainty related to the use of P-DHIs as a major part of public health initiatives. Developing apps and making them accessible on mobile devices sounds easy; however, developing apps in compliance with the requirements stated by the P-DHI model is quite complicated because of the variety of personal, technical, organizational, and social requirements. Furthermore, implementation is difficult, non-IT investments are considerable, and positive outcomes found in RCT research cannot be assumed to easily translate to large-scale implementations as there is little coordinated control over the factors influencing the outcome. Although some factors may be controlled to an extent by individuals, other factors are under the control of local communities and social networks, health care organizations, government agencies, and private sector companies.

Discussion

Principal Findings

The research in this paper set out to amplify our understanding of how to increase the predictability of outcomes from P-DHIs focusing on PA and diet behaviors as well as expand our understanding of how to improve the effectiveness of large-scale implementations. The P-DHI investment model presented in Figure 2 addresses both concerns and helps us understand "under what conditions P-DHIs pay off."

Predictability

The P-DHI investment model can be used to increase predictability in P-DHI research and practice as it includes the major constructs and factors that influence outcomes. The RCT reviews examined in this paper did not include descriptions of the many factors influencing outcomes, which is likely because these descriptions are missing from the individual studies. The reviews typically provide information about the use of BCTs but do not disclose how well the apps support individuals' capability development or the use of social networks, tailoring, software quality, or the use of general mHealth app quality rating scales [34] and app quality rating scales specifically for health behavior change [81,149]. They do not include information about non-IT investments made by individuals to change behaviors. The RCT research reviewed in this paper provides information about changes in people's health-related process performance (eg, increased PA) and tangible outcomes for people, such as weight loss. Lag effects were not reported, and the same applies to most factors related to the personal context, with the exception of demographic factors such as age, sex, profession, and health. Community-level contextual factors that influence behavior change (such as social support) were not reported. In addition, factors related to other parts of the context were not reported, but these factors likely do not influence the outcomes of RCT studies. Furthermore, it was not reported how well the constructs were aligned (eg, how well P-DHI apps match individuals and the context in which they live). Similar concerns were raised by RCT reviews emphasizing a need for improved intervention reporting in RCT studies [6,11,14,16,20,150] and for more studies that advance our knowledge on the contribution of the different parts of P-DHIs (eg, BCTs and personal contact with health care professionals) [8,10,11,15,18].

Reporting information about the constructs, the relationships between them, and the categories of factors in the P-DHI investment model when publishing empirical studies would help explain why some P-DHIs fail or succeed for some persons and, thereby, increase outcome predictability and create opportunities for improvement.

Large-Scale Effectiveness

The cost-effectiveness of large-scale implementations is reportedly unknown [6], and the literature review conducted in this paper found no information regarding the cost-effectiveness of large-scale implementations. However, the previous section described how the P-DHI investment model can be used during design and implementation to increase large-scale effectiveness.

Future Research

This research also points toward areas that need further study. There is a need for more empirical research on the contribution of the different parts of P-DHI apps; individuals' non-IT investments; lag effects; and the many different types of potential outcomes of P-DHI use that extend past the tangible health outcomes, for example, how P-DHIs can be used to increase individual capabilities necessary to experience long-term health benefits. Furthermore, this review identified a need for research that can clarify some of the uncertainties regarding how to best use BCTs and web-based social networks in P-DHIs. Future research could also benefit from including theories from the socioecological tradition to investigate how P-DHIs can not only support individual behavior change but also improve the context in which the behavior takes place. Regarding future literature reviews, the literature review presented in this paper could inspire other researchers to conduct multidisciplinary reviews combining knowledge from different fields. The use of P-DHIs is a multidisciplinary approach, but this does not seem to be reflected in the current research on P-DHIs. The P-DHI model may inspire researchers to address some of the uncertainties raised in this study by exploiting other streams of literature.

Limitations

This research has limitations related to the way in which the literature review was conducted. The scope of the mHealth implementation literature is broader than that of PA and diet P-DHIs, which introduces the risk that some of the identified factors are less relevant for the PA and diet P-DHIs studied in this review. The argument for using this broader scope is that the major difficulties in implementing these solutions (eg, the existence of supportive policies and infrastructures) are likely independent of the specific types of apps. Another limitation is the breadth of the literature review, which does not cover all the factors in detail. However, the goal was to establish a broad understanding of the constructs and factors influencing outcomes rather than exploring the individual factors in detail. The restricted use of the public health socioecological perspective in the model is another limitation. The reviewed mHealth literature almost exclusively addressed how to support individuals in changing health-related behaviors, but other kinds of mHealth apps with greater focus on changing the context to support healthy behaviors would also add value. Furthermore, there are other literature streams that would be valuable to study to address the research objective, for example, literature on nudging. Finally, the reviewed mHealth RCT research was predominantly based on empirical studies from high-income countries, whereas the mHealth implementation research reviewed was predominantly based on empirical studies from lower-income countries.

Conclusions

This research suggests that outcome predictability could be improved by including descriptions of the constructs and factors in the P-DHI investment model when reporting empirical studies. Doing so would increase our understanding of when and why P-DHIs succeed or fail. The effectiveness of large-scale implementations may be improved by using the P-DHI investment model to evaluate potential difficulties and possibilities in implementing P-DHIs to create better environments for the use of P-DHIs before investing in them and when designing and implementing them. The cost-effectiveness of large-scale implementations is unknown; implementations are far more complicated than just downloading and using apps, and there is uncertainty accompanying implementations given the lack of coordinated control over the constructs and factors that influence the outcome.

Acknowledgments

This research was funded internally by the university.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.
[DOCX File , 85 KB - ijmr_v12i1e40205_app1.docx]

References

1. Bolnick HJ, Bui AL, Bulchis A, Chen C, Chapin A, Lomsadze L, et al. Health-care spending attributable to modifiable risk factors in the USA: an economic attribution analysis. Lancet Public Health 2020 Oct;5(10):e525-e535 [FREE Full text] [doi: 10.1016/s2468-2667(20)30203-6]
2. Düking P, Tafler M, Wallmann-Sperlich B, Sperlich B, Kleih S. Behavior change techniques in wrist-worn wearables to promote physical activity: content analysis. JMIR Mhealth Uhealth 2020 Nov 19;8(11):e20820 [FREE Full text] [doi: 10.2196/20820] [Medline: 33211023]

3. McKay FH, Wright A, Shill J, Stephens H, Uccellini M. Using health and well-being apps for behavior change: a systematic search and rating of apps. JMIR Mhealth Uhealth 2019 Jul 04;7(7):e11926 [FREE Full text] [doi: 10.2196/11926] [Medline: 31274112]
4. Western MJ, Armstrong ME, Islam I, Morgan K, Jones UF, Kelson MJ. The effectiveness of digital interventions for increasing physical activity in individuals of low socioeconomic status: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2021 Nov 09;18(1):148 [FREE Full text] [doi: 10.1186/s12966-021-01218-4] [Medline: 34753490]
5. Böhm B, Karwiese SD, Böhm H, Oberhoffer R. Effects of mobile health including wearable activity trackers to increase physical activity outcomes among healthy children and adolescents: systematic review. JMIR Mhealth Uhealth 2019 Apr 30;7(4):e8298. [doi: 10.2196/mhealth.8298] [Medline: 31038460]
6. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017 Apr 18;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
7. Cavero-Redondo I, Martinez-Vizcaino V, Fernandez-Rodriguez R, Saz-Lara A, Pascual-Morena C, Álvarez-Bueno C. Effect of behavioral weight management interventions using lifestyle mHealth self-monitoring on weight loss: a systematic review and meta-analysis. Nutrients 2020 Jul 03;12(7):1977 [FREE Full text] [doi: 10.3390/nu12071977] [Medline: 32635174]
8. Stephenson A, McDonough SM, Murphy MH, Nugent CD, Mair JL. Using computer, mobile and wearable technology enhanced interventions to reduce sedentary behaviour: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2017 Aug 11;14(1):105 [FREE Full text] [doi: 10.1186/s12966-017-0561-4] [Medline: 28800736]
9. McCarroll R, Eyles H, Ni Mhurchu C. Effectiveness of mobile health (mHealth) interventions for promoting healthy eating in adults: a systematic review. Prev Med 2017 Dec;105:156-168. [doi: 10.1016/j.ypmed.2017.08.022] [Medline: 28882743]
10. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec 07;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
11. Milne-Ives M, Lam C, De Cock C, Van Velthoven MH, Meinert E. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review. JMIR Mhealth Uhealth 2020 Mar 18;8(3):e17046 [FREE Full text] [doi: 10.2196/17046] [Medline: 32186518]
12. Emberson MA, Lalande A, Wang D, McDonough DJ, Liu W, Gao Z. Effectiveness of smartphone-based physical activity interventions on individuals' health outcomes: a systematic review. Biomed Res Int 2021 Aug 6;2021:6296896 [FREE Full text] [doi: 10.1155/2021/6296896] [Medline: 34409104]
13. Dobbie LJ, Tahrani A, Alam U, James J, Wilding J, Cuthbertson DJ. Exercise in obesity-the role of technology in health services: can this approach work? Curr Obes Rep 2022 Sep 17;11(3):93-106 [FREE Full text] [doi: 10.1007/s13679-021-00461-x] [Medline: 34791611]
14. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
15. Langarizadeh M, Sadeghi M, As'habi A, Rahmati P, Sheikhtaheri A. Mobile apps for weight management in children and adolescents; an updated systematic review. Patient Educ Couns 2021 Sep;104(9):2181-2188. [doi: 10.1016/j.pec.2021.01.035] [Medline: 33573915]
16. Domin A, Spruijt-Metz D, Theisen D, Ouzzahra Y, Vögele C. Smartphone-based interventions for physical activity promotion: scoping review of the evidence over the last 10 years. JMIR Mhealth Uhealth 2021 Jul 21;9(7):e24308 [FREE Full text] [doi: 10.2196/24308] [Medline: 34287209]
17. Schorr EN, Gepner AD, Dolansky MA, Forman DE, Park LG, Petersen KS, American Heart Association Cardiovascular Disease in Older Populations Committee of the Council on Clinical CardiologyCouncil on CardiovascularStroke Nursing; Council on Arteriosclerosis, ThrombosisVascular Biology;Council on LifestyleCardiometabolic Health. Harnessing mobile health technology for secondary cardiovascular disease prevention in older adults: a scientific statement from the american heart association. Circ Cardiovasc Qual Outcomes 2021 May;14(5):e000103 [FREE Full text] [doi: 10.1161/HCQ.0000000000000103] [Medline: 33793309]
18. Wang E, Abrahamson K, Liu PJ, Ahmed A. Can mobile technology improve weight loss in overweight adults? A systematic review. West J Nurs Res 2020 Sep 25;42(9):747-759. [doi: 10.1177/0193945919888224] [Medline: 31762402]
19. Kim HN, Seo K. Smartphone-based health program for improving physical activity and tackling obesity for young adults: a systematic review and meta-analysis. Int J Environ Res Public Health 2019 Dec 18;17(1):15 [FREE Full text] [doi: 10.3390/ijerph17010015] [Medline: 31861359]
20. Fiedler J, Eckert T, Wunsch K, Woll A. Key facets to build up eHealth and mHealth interventions to enhance physical activity, sedentary behavior and nutrition in healthy subjects - an umbrella review. BMC Public Health 2020 Oct 23;20(1):1605 [FREE Full text] [doi: 10.1186/s12889-020-09700-7] [Medline: 33097013]
21. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]

22. Han M, Lee E. Effectiveness of mobile health application use to improve health behavior changes: a systematic review of randomized controlled trials. Healthc Inform Res 2018 Jul;24(3):207-226 [FREE Full text] [doi: 10.4258/hir.2018.24.3.207] [Medline: 30109154]
23. Kim S, Choudhury A. Comparison of older and younger adults' attitudes toward the adoption and use of activity trackers. JMIR Mhealth Uhealth 2020 Oct 22;8(10):e18312 [FREE Full text] [doi: 10.2196/18312] [Medline: 33090116]
24. Puri A, Kim B, Nguyen O, Stolee P, Tung J, Lee J. User acceptance of wrist-worn activity trackers among community-dwelling older adults: mixed method study. JMIR Mhealth Uhealth 2017 Nov 15;5(11):e173 [FREE Full text] [doi: 10.2196/mhealth.8211] [Medline: 29141837]
25. Okumus B, Ali F, Bilgihan A, Ozturk AB. Psychological factors influencing customers’ acceptance of smartphone diet apps when ordering food at restaurants. Int J Hosp Manag 2018 Jun;72:67-77. [doi: 10.1016/j.ijhm.2018.01.001]
26. König LM, Attig C, Franke T, Renner B. Barriers to and facilitators for using nutrition apps: systematic review and conceptual framework. JMIR Mhealth Uhealth 2021 Apr 01;9(6):e20037 [FREE Full text] [doi: 10.2196/20037] [Medline: 34254938]
27. Aranda-Jan CB, Mohutsiwa-Dibe N, Loukanova S. Systematic review on what works, what does not work and why of implementation of mobile health (mHealth) projects in Africa. BMC Public Health 2014 Feb 21;14(1):188 [FREE Full text] [doi: 10.1186/1471-2458-14-188] [Medline: 24555733]
28. Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: a systematic review. J Telemed Telecare 2016 Oct 16;24(1):4-12. [doi: 10.1177/1357633x16674087]
29. Dedrick J, Gurbaxani V, Kraemer KL. Information technology and economic performance: a critical review of the empirical evidence. ACM Comput Surv 2003 Mar;35(1):1-28. [doi: 10.1145/641865.641866]
30. Richardson VJ, Zmud RW. The value relevance of information technology investment announcements: incorporating industry strategic IT role. In: Proceedings of the 35th Annual Hawaii International Conference on System Sciences. 2002 Presented at: 35th Annual Hawaii International Conference on System Sciences; Jan 7-10, 2002; Big Island, HI. [doi: 10.1109/hicss.2002.994251]
31. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Sep;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]
32. Munn Z, Peters MD, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Nov 19;18(1):143 [FREE Full text] [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
33. Bandura A. Self-efficacy. In: Ramachaudran VS, editor. Encyclopedia of Human Behavior. New York, NY: Academic Press; 1994.
34. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
35. Leon N, Schneider H, Daviaud E. Applying a framework for assessing the health system challenges to scaling up mHealth in South Africa. BMC Med Inform Decis Mak 2012 Nov 05;12(1):123 [FREE Full text] [doi: 10.1186/1472-6947-12-123] [Medline: 23126370]
36. Davis R, Campbell R, Hildon Z, Hobbs L, Michie S. Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. Health Psychol Rev 2015 Aug 08;9(3):323-344 [FREE Full text] [doi: 10.1080/17437199.2014.941722] [Medline: 25104107]
37. Dehning B, Richardson VJ. Returns on investments in information technology: a research synthesis. J Inf Syst 2002;16(1). [doi: 10.2139/ssrn.302139]
38. Melville N, Kraemer K, Gurbaxani V. Review: information technology and organizational performance: an integrative model of IT business value. MIS Q 2004;28(2):283-322. [doi: 10.2307/25148636]
39. Schryen G. Revisiting IS business value research: what we already know, what we still need to know, and how we can get there. Eur J Inf Syst 2017 Dec 19;22(2):139-169. [doi: 10.1057/ejis.2012.45]
40. Ginige JA, Maeder AJ, Long V. Evaluating success of mobile health projects in the developing world. Stud Health Technol Inform 2014;206:7-19. [Medline: 25365667]
41. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
42. McGarrigle L, Todd C. Promotion of physical activity in older people using mHealth and ehealth technologies: rapid review of reviews. J Med Internet Res 2020 Dec 29;22(12):e22201 [FREE Full text] [doi: 10.2196/22201] [Medline: 33372894]
43. Opoku D, Stephani V, Quentin W. A realist review of mobile phone-based health interventions for non-communicable disease management in sub-Saharan Africa. BMC Med 2017 Feb 06;15(1):24 [FREE Full text] [doi: 10.1186/s12916-017-0782-z] [Medline: 28162090]
44. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158 [FREE Full text] [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]

45. Baretta D, Perski O, Steca P. Exploring users' experiences of the uptake and adoption of physical activity apps: longitudinal qualitative study. JMIR Mhealth Uhealth 2019 Feb 08;7(2):e11636 [FREE Full text] [doi: 10.2196/11636] [Medline: 30735143]
46. Lyzwinski LN, Caffery LJ, Bambling M, Edirippulige S. Consumer perspectives on mHealth for weight loss: a review of qualitative studies. J Telemed Telecare 2017 Feb 09;24(4):290-302. [doi: 10.1177/1357633x17692722]
47. Yuan S, Ma W, Kanthawala S, Peng W. Keep using my health apps: discover users' perception of health and fitness apps with the UTAUT2 model. Telemed J E Health 2015 Sep;21(9):735-741. [doi: 10.1089/tmj.2014.0148] [Medline: 25919238]
48. Lee C, Lee K, Lee D. Mobile healthcare applications and gamification for sustained health maintenance. Sustainability 2017 May 08;9(5):772. [doi: 10.3390/su9050772]
49. Ferreira Barbosa H, García-Fernández J, Pedragosa V, Cepeda-Carrion G. The use of fitness centre apps and its relation to customer satisfaction: a UTAUT2 perspective. Int J Sports Mark Spons 2021 May 20;23(5):966-985. [doi: 10.1108/ijsms-01-2021-0010]
50. Lee HE, Cho J. What motivates users to continue using diet and fitness apps? Application of the uses and gratifications approach. Health Commun 2017 Dec 29;32(12):1445-1453. [doi: 10.1080/10410236.2016.1167998] [Medline: 27356103]
51. Okorodudu DE, Bosworth HB, Corsino L. Innovative interventions to promote behavioral change in overweight or obese individuals: a review of the literature. Ann Med 2015 May 10;47(3):179-185 [FREE Full text] [doi: 10.3109/07853890.2014.931102] [Medline: 25011006]
52. Turner T, Spruijt-Metz D, Wen CK, Hingle MD. Prevention and treatment of pediatric obesity using mobile and wireless technologies: a systematic review. Pediatr Obes 2015 Dec 12;10(6):403-409 [FREE Full text] [doi: 10.1111/ijpo.12002] [Medline: 25641770]
53. Cajita MI, Hodgson NA, Budhathoki C, Han HR. Intention to use mHealth in older adults with heart failure. J Cardiovasc Nurs 2017;32(6):E1-E7 [FREE Full text] [doi: 10.1097/JCN.0000000000000401] [Medline: 28248747]
54. Odendaal WA, Anstey Watkins J, Leon N, Goudge J, Griffiths F, Tomlinson M, et al. Health workers' perceptions and experiences of using mHealth technologies to deliver primary healthcare services: a qualitative evidence synthesis. Cochrane Database Syst Rev 2020 Mar 26;3(3):CD011942 [FREE Full text] [doi: 10.1002/14651858.CD011942.pub2] [Medline: 32216074]
55. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
56. Vasiloglou MF, Christodoulidis S, Reber E, Stathopoulou T, Lu Y, Stanga Z, et al. What healthcare professionals think of "Nutrition and diet" apps: an international survey. Nutrients 2020 Jul 24;12(8):2214 [FREE Full text] [doi: 10.3390/nu12082214] [Medline: 32722339]
57. Kaboré SS, Ngangue P, Soubeiga D, Barro A, Pilabré AH, Bationo N, et al. Barriers and facilitators for the sustainability of digital health interventions in low and middle-income countries: a systematic review. Front Digit Health 2022 Nov 28;4:1014375 [FREE Full text] [doi: 10.3389/fdgth.2022.1014375] [Medline: 36518563]
58. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
59. Bettiga D, Lamberti L, Lettieri E. Individuals' adoption of smart technologies for preventive health care: a structural equation modeling approach. Health Care Manag Sci 2020 Jun 26;23(2):203-214. [doi: 10.1007/s10729-019-09468-2] [Medline: 30684067]
60. Whittaker R. Issues in mHealth: findings from key informant interviews. J Med Internet Res 2012 Oct 02;14(5):e129 [FREE Full text] [doi: 10.2196/jmir.1989] [Medline: 23032424]
61. Labrique AB, Wadhwani C, Williams KA, Lamptey P, Hesp C, Luk R, et al. Best practices in scaling digital health in low and middle income countries. Global Health 2018 Nov 03;14(1):103 [FREE Full text] [doi: 10.1186/s12992-018-0424-z] [Medline: 30390686]
62. Jacob C, Sanchez-Vazquez A, Ivory C. Social, organizational, and technological factors impacting clinicians' adoption of mobile health tools: systematic literature review. JMIR Mhealth Uhealth 2020 Feb 20;8(2):e15935 [FREE Full text] [doi: 10.2196/15935] [Medline: 32130167]
63. Zakerabasali S, Ayyoubzadeh SM, Baniasadi T, Yazdani A, Abhari S. Mobile health technology and healthcare providers: systemic barriers to adoption. Healthc Inform Res 2021 Oct;27(4):267-278 [FREE Full text] [doi: 10.4258/hir.2021.27.4.267] [Medline: 34788907]
64. Feldman DI, Theodore Robison W, Pacor JM, Caddell LC, Feldman EB, Deitz RL, et al. Harnessing mHealth technologies to increase physical activity and prevent cardiovascular disease. Clin Cardiol 2018 Jul 20;41(7):985-991 [FREE Full text] [doi: 10.1002/clc.22968] [Medline: 29671879]
65. Kong T, Scott MM, Li Y, Wichelman C. Physician attitudes towards-and adoption of-mobile health. Digit Health 2020 Feb 20;6:2055207620907187 [FREE Full text] [doi: 10.1177/2055207620907187] [Medline: 32128235]
66. Wattanapisit A, Amaek W, Wattanapisit S, Tuangratananon T, Wongsiri S, Pengkaew P. Challenges of implementing an mHealth application for personalized physical activity counselling in primary health care: a qualitative study. Int J Gen Med 2021 Jul;Volume 14:3821-3831. [doi: 10.2147/ijgm.s317241]

67. Matthew-Maich N, Harris L, Ploeg J, Markle-Reid M, Valaitis R, Ibrahim S, et al. Designing, implementing, and evaluating mobile health technologies for managing chronic conditions in older adults: a scoping review. JMIR Mhealth Uhealth 2016 Jun 09;4(2):e29 [FREE Full text] [doi: 10.2196/mhealth.5127] [Medline: 27282195]
68. Tong HL, Quiroz JC, Kocaballi AB, Fat SC, Dao KP, Gehringer H, et al. Personalized mobile technologies for lifestyle behavior change: a systematic review, meta-analysis, and meta-regression. Prev Med 2021 Jul;148:106532. [doi: 10.1016/j.ypmed.2021.106532] [Medline: 33774008]
69. Dao KP, De Cocker K, Tong HL, Kocaballi AB, Chow C, Laranjo L. Smartphone-delivered ecological momentary interventions based on ecological momentary assessments to promote health behaviors: systematic review and adapted checklist for reporting ecological momentary assessment and intervention studies. JMIR Mhealth Uhealth 2021 Nov 19;9(11):e22890 [FREE Full text] [doi: 10.2196/22890] [Medline: 34806995]
70. Mattila E, Korhonen I, Salminen JH, Ahtinen A, Koskinen E, Sarela A, et al. Empowering citizens for well-being and chronic disease management with wellness diary. IEEE Trans Inform Technol Biomed 2010 Mar;14(2):456-463. [doi: 10.1109/titb.2009.2037751]
71. Stehr P, Karnowski V, Rossmann C. The multi-faceted usage patterns of nutrition apps: a survey on the appropriation of nutrition apps among German-speaking users of MyFitnessPal. BMC Med Inform Decis Mak 2020 Oct 28;20(1):279 [FREE Full text] [doi: 10.1186/s12911-020-01294-9] [Medline: 33115444]
72. Wang Y, Collins WB. Systematic evaluation of mobile fitness apps: apps as the tutor, recorder, game companion, and cheerleader. Telemat Inform 2021 Jun;59:101552. [doi: 10.1016/j.tele.2020.101552]
73. Claudel SE, Ceasar JN, Andrews MR, El-Toukhy S, Farmer N, Middleton KR, et al. Time to listen: a mixed-method study examining community-based views of mobile technology for interventions to promote physical activity. BMJ Health Care Inform 2020 Aug 23;27(3):e100140 [FREE Full text] [doi: 10.1136/bmjhci-2020-100140] [Medline: 32830106]
74. de Korte EM, Wiezer N, Janssen JH, Vink P, Kraaij W. Evaluating an mHealth app for health and well-being at work: mixed-method qualitative study. JMIR Mhealth Uhealth 2018 Mar 28;6(3):e72 [FREE Full text] [doi: 10.2196/mhealth.6335] [Medline: 29592846]
75. Georgsson M, Staggers N. Patients' perceptions and experiences of a mHealth diabetes self-management system. Comput Inform Nurs 2017 Mar;35(3):122-130. [doi: 10.1097/CIN.0000000000000296] [Medline: 27748662]
76. Tonkin E, Jeffs L, Wycherley TP, Maher C, Smith R, Hart J, et al. A smartphone app to reduce sugar-sweetened beverage consumption among young adults in Australian remote indigenous communities: design, formative evaluation and user-testing. JMIR Mhealth Uhealth 2017 Dec 12;5(12):e192 [FREE Full text] [doi: 10.2196/mhealth.8651] [Medline: 29233803]
77. Ghanvatkar S, Kankanhalli A, Rajan V. User models for personalized physical activity interventions: scoping review. JMIR Mhealth Uhealth 2019 Jan 16;7(1):e11098 [FREE Full text] [doi: 10.2196/11098] [Medline: 30664474]
78. Brannon EE, Cushing CC. A systematic review: is there an app for that? Translational science of pediatric behavior change for physical activity and dietary interventions. J Pediatr Psychol 2015 May;40(4):373-384. [doi: 10.1093/jpepsy/jsu108] [Medline: 25502745]
79. Schwarzer R, Luszczynska A. How to overcome health-compromising behaviors. Eur Psychol 2008 Jan;13(2):141-151. [doi: 10.1027/1016-9040.13.2.141]
80. Schwarzer R, Lippke S, Luszczynska A. Mechanisms of health behavior change in persons with chronic illness or disability: the Health Action Process Approach (HAPA). Rehabil Psychol 2011 Aug;56(3):161-170. [doi: 10.1037/a0024509] [Medline: 21767036]
81. McKay FH, Slykerman S, Dunn M. The app behavior change scale: creation of a scale to assess the potential of apps to promote behavior change. JMIR Mhealth Uhealth 2019 Jan 25;7(1):e11130 [FREE Full text] [doi: 10.2196/11130] [Medline: 30681967]
82. Janz NK, Becker MH. The Health Belief Model: a decade later. Health Educ Q 1984 Sep 04;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]
83. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
84. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
85. Sallis JF, Hovell MF, Hofstetter CR, Barrington E. Explanation of vigorous physical activity during two years using social learning variables. Soc Sci Med 1992 Jan;34(1):25-32. [doi: 10.1016/0277-9536(92)90063-v] [Medline: 1738853]
86. McLeroy KR, Bibeau D, Steckler A, Glanz K. An ecological perspective on health promotion programs. Health Educ Q 1988 Sep 04;15(4):351-377. [doi: 10.1177/109019818801500401] [Medline: 3068205]
87. Green LW, Richard L, Potvin L. Ecological foundations of health promotion. Am J Health Promot 1996 Aug 26;10(4):270-281. [doi: 10.4278/0890-1171-10.4.270] [Medline: 10159708]
88. Bauman AE, Reis RS, Sallis JF, Wells JC, Loos RJ, Martin BW. Correlates of physical activity: why are some people physically active and others not? Lancet 2012 Jul;380(9838):258-271. [doi: 10.1016/s0140-6736(12)60735-1]
89. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997 Aug 26;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]

90. Fisher WA, Fisher JD, Harman J. The information-motivation-behavioral skills model: a general social psychological approach to understanding and promoting health behavior. In: Suls J, Wallston KA, editors. Social Psychological Foundations of Health and Illness. Hoboken, NJ: Wiley; 2003.
91. Bandura A. Health promotion from the perspective of social cognitive theory. Psychol Health 1998;13(4):623-649. [doi: 10.1080/08870449808407422]
92. Bandura A. Social Learning Theory. Hoboken, NJ: Prentice-Hall; 1977.
93. Duncan M, Mummery K. Psychosocial and environmental factors associated with physical activity among city dwellers in regional Queensland. Prev Med 2005 Apr;40(4):363-372. [doi: 10.1016/j.ypmed.2004.06.017] [Medline: 15530589]
94. McNeill LH, Kreuter MW, Subramanian SV. Social environment and physical activity: a review of concepts and evidence. Soc Sci Med 2006 Aug;63(4):1011-1022. [doi: 10.1016/j.socscimed.2006.03.012] [Medline: 16650513]
95. Thornton CM, Kerr J, Conway TL, Saelens BE, Sallis JF, Ahn DK, et al. Physical activity in older adults: an ecological approach. Ann Behav Med 2017 Apr 28;51(2):159-169 [FREE Full text] [doi: 10.1007/s12160-016-9837-1] [Medline: 27680568]
96. Robinson T. Applying the socio-ecological model to improving fruit and vegetable intake among low-income African Americans. J Community Health 2008 Dec 2;33(6):395-406. [doi: 10.1007/s10900-008-9109-5] [Medline: 18594953]
97. Di Cesare M, Sorić M, Bovet P, Miranda JJ, Bhutta Z, Stevens GA, et al. The epidemiological burden of obesity in childhood: a worldwide epidemic requiring urgent action. BMC Med 2019 Nov 25;17(1):212 [FREE Full text] [doi: 10.1186/s12916-019-1449-8] [Medline: 31760948]
98. Dugas M, Gao GG, Agarwal R. Unpacking mHealth interventions: a systematic review of behavior change techniques used in randomized controlled trials assessing mHealth effectiveness. Digit Health 2020 Feb 20;6:2055207620905411 [FREE Full text] [doi: 10.1177/2055207620905411] [Medline: 32128233]
99. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012 May 08;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]
100. Huang CY, Yang MC. Empirical investigation of factors influencing consumer intention to use an artificial intelligence-powered mobile application for weight loss and health management. Telemed J E Health 2020 Oct 01;26(10):1240-1251. [doi: 10.1089/tmj.2019.0182] [Medline: 31971883]
101. Liu D, Maimaitijiang R, Gu J, Zhong S, Zhou M, Wu Z, et al. Using the unified theory of acceptance and use of technology (UTAUT) to investigate the intention to use physical activity apps: cross-sectional survey. JMIR Mhealth Uhealth 2019 Aug 22;7(9):e13127 [FREE Full text] [doi: 10.2196/13127] [Medline: 31507269]
102. Ndayizigamiye P, Kante M, Shingwenyana S. An adoption model of mHealth applications that promote physical activity. Cogent Psychol 2020 May 12;7(1). [doi: 10.1080/23311908.2020.1764703]
103. Cho J, Kim S, Jeong G, Kim C, Seo JK. Investigation of influential factors of predicting individuals' use and non-use of fitness and diet apps on smartphones: application of the machine learning algorithm (XGBoost). Am J Health Behav 2021 Jan 01;45(1):111-124. [doi: 10.5993/ajhb.45.1.9]
104. Tsai TH, Chang YS, Chang HT, Lin YW. Running on a social exercise platform: applying self-determination theory to increase motivation to participate in a sporting event. Comput Hum Behav 2021 Jan;114:106523. [doi: 10.1016/j.chb.2020.106523]
105. Deci EL, Ryan RM. The "what" and "why" of goal pursuits: human needs and the self-determination of behavior. Psychol Inq 2000 Oct;11(4):227-268. [doi: 10.1207/s15327965pli1104_01]
106. Petersen JM, Prichard I, Kemps E. A comparison of physical activity mobile apps with and without existing web-based social networking platforms: systematic review. J Med Internet Res 2019 Aug 16;21(8):e12687 [FREE Full text] [doi: 10.2196/12687] [Medline: 31420956]
107. Ehlers DK, Huberty JL. Middle-aged women's preferred theory-based features in mobile physical activity applications. J Phys Act Health 2014 Sep;11(7):1379-1385. [doi: 10.1123/jpah.2012-0435] [Medline: 24368818]
108. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS One 2016 May 23;11(5):e0156164 [FREE Full text] [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]
109. Arigo D, Brown MM, Pasko K, Suls J. Social comparison features in physical activity promotion apps: scoping meta-review. J Med Internet Res 2020 Mar 27;22(3):e15642 [FREE Full text] [doi: 10.2196/15642] [Medline: 32217499]
110. Li L, Peng W, Kononova A, Bowen M, Cotten SR. Factors associated with older adults' long-term use of wearable activity trackers. Telemed J E Health 2020 Jun 01;26(6):769-775. [doi: 10.1089/tmj.2019.0052] [Medline: 31553281]
111. Molina MD, Myrick JG. The ‘how’ and ‘why’ of fitness app use: investigating user motivations to gain insights into the nexus of technology and fitness. Sport Soc 2020 Apr 07;24(7):1233-1248. [doi: 10.1080/17430437.2020.1744570]
112. Windasari NA, Lin FR, Kato-Lin YC. Continued use of wearable fitness technology: a value co-creation perspective. Int J Inf Manag 2021 Apr;57:102292. [doi: 10.1016/j.ijinfomgt.2020.102292]
113. Brynjolfsson E, Hitt LM. Computing productivity: firm-level evidence. Rev Econ Stat 2003 Nov;85(4):793-808. [doi: 10.1162/003465303772815736]
114. Lobelo F, Kelli HM, Tejedor SC, Pratt M, McConnell MV, Martin SS, et al. The wild wild west: a framework to integrate mHealth software applications and wearables to support physical activity assessment, counseling and interventions for

cardiovascular disease risk reduction. Prog Cardiovasc Dis 2016 May;58(6):584-594 [FREE Full text] [doi: 10.1016/j.pcad.2016.02.007] [Medline: 26923067]

115. Yang Y, Koenigstorfer J. Determinants of fitness app usage and moderating impacts of education-, motivation-, and gamification-related app features on physical activity intentions: cross-sectional survey study. J Med Internet Res 2021 Jul 13;23(7):e26063 [FREE Full text] [doi: 10.2196/26063] [Medline: 34255656]
116. Jeon E, Park HA. Factors affecting acceptance of smartphone application for management of obesity. Healthc Inform Res 2015 Apr;21(2):74-82 [FREE Full text] [doi: 10.4258/hir.2015.21.2.74] [Medline: 25995959]
117. Early J, Gonzalez C, Gordon-Dseagu V, Robles-Calderon L. Use of mobile health (mHealth) technologies and interventions among community health workers globally: a scoping review. Health Promot Pract 2019 Nov 10;20(6):805-817. [doi: 10.1177/1524839919855391] [Medline: 31179777]
118. Kozik M, Isakadze N, Martin SS. Mobile health in preventive cardiology: current status and future perspective. Curr Opin Cardiol 2021 Sep 01;36(5):580-588 [FREE Full text] [doi: 10.1097/HCO.0000000000000891] [Medline: 34224437]
119. Richard L, Gauvin L, Raine K. Ecological models revisited: their uses and evolution in health promotion over two decades. Annu Rev Public Health 2011 Apr 21;32(1):307-326. [doi: 10.1146/annurev-publhealth-031210-101141] [Medline: 21219155]
120. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6(1):42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
121. Alsswey A, Al-Samarraie H. Elderly users' acceptance of mHealth user interface (UI) design-based culture: the moderator role of age. J Multimodal User Interfaces 2019 Jul 20;14(1):49-59. [doi: 10.1007/s12193-019-00307-w]
122. Chae J. A comprehensive profile of those who have health-related apps. Health Educ Behav 2018 Aug 20;45(4):591-598. [doi: 10.1177/1090198117752784] [Medline: 29353544]
123. Shaw RJ, Yang Q, Barnes A, Hatch D, Crowley MJ, Vorderstrasse A, et al. Self-monitoring diabetes with multiple mobile health devices. J Am Med Inform Assoc 2020 May 01;27(5):667-676 [FREE Full text] [doi: 10.1093/jamia/ocaa007] [Medline: 32134447]
124. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017 Apr 19;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]
125. Chandrasekaran R, Katthula V, Moustakas E. Patterns of use and key predictors for the use of wearable health care devices by US adults: insights from a national survey. J Med Internet Res 2020 Oct 16;22(10):e22443 [FREE Full text] [doi: 10.2196/22443] [Medline: 33064083]
126. Zhang Y, Lin Z, Li X, Xiaoming T. Factors affecting ICT use in health communication among the older population in Jiangsu, China. Libri 2019;69(1):41-53. [doi: 10.1515/libri-2018-0103]
127. Oyibo K, Olagunju AH, Olabenjo B, Adaji I, Deters R, Vassileva J. BEN'FIT: design, implementation and evaluation of a culture-tailored fitness app. In: Proceedings of the Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization. 2019 Presented at: UMAP '19: 27th Conference on User Modeling, Adaptation and Personalization; Jun 9-12, 2019; Larnaca, Cyprus.
128. Dutta-Bergman MJ. Primary sources of health information: comparisons in the domain of health attitudes, health cognitions, and health behaviors. Health Commun 2004;16(3):273-288. [doi: 10.1207/s15327027hc1603_1]
129. Chen MF, Lin NP. Incorporation of health consciousness into the technology readiness and acceptance model to predict app download and usage intentions. Internet Res 2018 Apr 04;28(2):351-373. [doi: 10.1108/intr-03-2017-0099]
130. Damberg S. Predicting future use intention of fitness apps among fitness app users in the United Kingdom: the role of health consciousness. Int J Sports Mark Spons 2021 Aug 17;23(2):369-384. [doi: 10.1108/ijsms-01-2021-0013]
131. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
132. Zhang Y, Liu C, Luo S, Xie Y, Liu F, Li X, et al. Factors influencing patients' intentions to use diabetes management apps based on an extended unified theory of acceptance and use of technology model: web-based survey. J Med Internet Res 2019 Aug 13;21(8):e15023 [FREE Full text] [doi: 10.2196/15023] [Medline: 31411146]
133. Miao R, Wu Q, Wang Z, Zhang X, Song Y, Zhang H, et al. Factors that influence users' adoption intention of mobile health: a structural equation modeling approach. Int J Prod Res 2017 Jun 12;55(19):5801-5815. [doi: 10.1080/00207543.2017.1336681]
134. Van Rhoon L, McSharry J, Byrne M. Development and testing of a digital health acceptability model to explain the intention to use a digital diabetes prevention programme. Br J Health Psychol 2022 Sep 31;27(3):716-740. [doi: 10.1111/bjhp.12569] [Medline: 34719099]
135. Beldad AD, Hegner SM. Expanding the technology acceptance model with the inclusion of trust, social influence, and health valuation to determine the predictors of German users' willingness to continue using a fitness app: a structural equation modeling approach. Int J Hum Comput Interact 2017 Nov 30;34(9):882-893. [doi: 10.1080/10447318.2017.1403220]
136. Sallis JF, Bauman A, Pratt M. Environmental and policy interventions to promote physical activity. Am J Prev Med 1998 Nov;15(4):379-397. [doi: 10.1016/s0749-3797(98)00076-2] [Medline: 9838979]

137. Aamir J, Ali SM, Kamel Boulos MN, Anjum N, Ishaq M. Enablers and inhibitors: a review of the situation regarding mHealth adoption in low- and middle-income countries. Health Policy Technol 2018 Mar;7(1):88-97. [doi: 10.1016/j.hlpt.2017.11.005]
138. Krah EF, de Kruijf JG. Exploring the ambivalent evidence base of mobile health (mHealth): a systematic literature review on the use of mobile phones for the improvement of community health in Africa. Digit Health 2016 Nov 24;2:2055207616679264 [FREE Full text] [doi: 10.1177/2055207616679264] [Medline: 29942576]
139. Glynn LG, Glynn F, Casey M, Wilkinson LG, Hayes PS, Heaney D, et al. Implementation of the SMART MOVE intervention in primary care: a qualitative study using normalisation process theory. BMC Fam Pract 2018 May 02;19(1):48 [FREE Full text] [doi: 10.1186/s12875-018-0737-2] [Medline: 29720084]
140. Okazaki S, Castañeda JA, Sanz S, Henseler J. Factors affecting mobile diabetes monitoring adoption among physicians: questionnaire study and path model. J Med Internet Res 2012 Dec 21;14(6):e183 [FREE Full text] [doi: 10.2196/jmir.2159] [Medline: 23257115]
141. Byambasuren O, Beller E, Glasziou P. Current knowledge and adoption of mobile health apps among Australian general practitioners: survey study. JMIR Mhealth Uhealth 2019 Jun 03;7(6):e13199 [FREE Full text] [doi: 10.2196/13199] [Medline: 31199343]
142. Seyyedi N, Rahimi B, Farrokh Eslamlou HR, Timpka T, Lotfnezhad Afshar H. Mobile phone applications to overcome malnutrition among preschoolers: a systematic review. BMC Med Inform Decis Mak 2019 Apr 05;19(1):83 [FREE Full text] [doi: 10.1186/s12911-019-0803-2] [Medline: 30953497]
143. Kiberu VM, Mars M, Scott RE. Barriers and opportunities to implementation of sustainable e-Health programmes in Uganda: a literature review. Afr J Prim Health Care Fam Med 2017 May 29;9(1):e1-10 [FREE Full text] [doi: 10.4102/phcfm.v9i1.1277] [Medline: 28582996]
144. Cho J, Kim S. Personal and social predictors of use and non-use of fitness/diet app: application of random forest algorithm. Telemat Inform 2020 Dec;55:101301. [doi: 10.1016/j.tele.2019.101301]
145. Gordon NP, Hornbrook MC. Older adults' readiness to engage with eHealth patient education and self-care resources: a cross-sectional survey. BMC Health Serv Res 2018 Mar 27;18(1):220 [FREE Full text] [doi: 10.1186/s12913-018-2986-0] [Medline: 29587721]
146. Fisher JD, Fisher WA. Changing AIDS-risk behavior. Psychol Bull 1992 May;111(3):455-474. [doi: 10.1037/0033-2909.111.3.455] [Medline: 1594721]
147. Brynjolfsson E, Rock D, Syverson C. Artificial intelligence and the modern productivity paradox: a clash of expectations and statistics. In: Agrawal A, Gans J, Goldfarb A, editors. The Economics of Artificial Intelligence An Agenda. Chicago, IL: University of Chicago Press; 2019.
148. Mönninghoff A, Kramer JN, Hess AJ, Ismailova K, Teepe GW, Tudor Car L, et al. Long-term effectiveness of mHealth physical activity interventions: systematic review and meta-analysis of randomized controlled trials. J Med Internet Res 2021 Apr 30;23(4):e26699 [FREE Full text] [doi: 10.2196/26699] [Medline: 33811021]
149. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: a comprehensive quality and therapeutic potential evaluation tool for mobile and web-based eHealth interventions. J Med Internet Res 2017 Mar 21;19(3):e82 [FREE Full text] [doi: 10.2196/jmir.7270] [Medline: 28325712]
150. Zaslavsky O, Roopsawang I, Chen AT. Promoting behavioral change in mobile health interventions for older adults: a scoping review. Res Gerontol Nurs 2020 Mar 01;13(2):102-116 [FREE Full text] [doi: 10.3928/19404921-20191031-01] [Medline: 31697395]

Abbreviations

BCT: behavior change technique
mHealth: mobile health
PA: physical activity
P-DHI: preventive digital health intervention
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RCT: randomized controlled trial

Edited by T de Azevedo Cardoso; submitted 10.06.22; peer-reviewed by T Baranowski, MS Aslam, A Finny; comments to author 28.08.22; revised version received 01.11.22; accepted 09.06.23; published 20.07.23.

Please cite as:
Pedersen K, Schlichter BR
Improving Predictability and Effectiveness in Preventive Digital Health Interventions: Scoping Review
Interact J Med Res 2023;12:e40205
URL: https://www.i-jmr.org/2023/1/e40205
doi:10.2196/40205
PMID:37471129

Review

Studies on HIV/AIDS Among Students: Bibliometric Analysis

Na Wang[1,2*], MD; Runxi Zhang[2*], MBBS; Zeyan Ye[2], MPH; Guanghua Lan[1], MD; Qiuying Zhu[1], MS; Huanhuan Chen[1], MS; Xiangjun Zhang[3], MD; Shengkui Tan[2], MD; Yuhua Ruan[1], MD; Mei Lin[1], MD

[1]Guangxi Key Laboratory of Major Infectious Disease Prevention Control and Biosafety Emergency Response, Guangxi Center for Disease Control and Prevention, Nanning, China

[2]Guangxi Key Laboratory of Environmental Exposomics and Entire Lifecycle Health, School of Public Health, Guilin Medical University, Guilin, China

[3]Department of Clinical Pharmacy and Translational Science, College of Pharmacy, University of Tennessee Health Science Center, Memphis, TN, United States

[*]these authors contributed equally

Corresponding Author:
Mei Lin, MD
Guangxi Key Laboratory of Major Infectious Disease Prevention Control and Biosafety Emergency Response
Guangxi Center for Disease Control and Prevention
18 Jinzhou Road
Nanning, 530028
China
Phone: 86 07712518689
Email: gxlinmei@126.com

Abstract

Background: In recent years, HIV infection in students has been an ongoing concern worldwide. A large number of articles have been published; however, statistical analysis of the data presented in these publications is lacking.

Objective: This study aimed to detect and analyze emerging trends and collaborative networks in research on HIV/AIDS among students.

Methods: Research publications on HIV/AIDS among students from 1985 to 2022 were collected from the Web of Science Core Collection. A topic search was used for this study, and articles in English were included. CiteSpace was used to generate visual networks of countries/regions, institutions, references, and keywords. Citation analysis was used to discover milestones in the field and trace the roots of the knowledge base. Keyword analysis was used to detect research hotspots and predict future trends.

Results: A total of 2726 publications met the inclusion criteria. Over the past 38 years, the number of publications annually has been on the rise overall. The United States had the highest number of publications (n=1303) and the highest centrality (0.91). The University of California system was the core institution. The main target population of studies on HIV/AIDS among students were medical and university students. These studies focused on students' knowledge, attitudes, risk behaviors, and education about HIV/AIDS. The recent bursting keywords (gay, sexual health, adherence, barriers, mental health, HIV testing, stigma, and antiretroviral therapy) revealed research trends and public interest on this topic.

Conclusions: This study identified countries/regions and institutions contributing to the research area of HIV/AIDS among students and revealed research hotspots and emerging trends. The field of research on HIV/AIDS among students was growing rapidly. The United States was at the center, and the University of California system was the core institution. However, academic collaboration should be strengthened. Future research may focus on exploring gay students, sexual health, adherence, barriers, mental health, HIV testing, stigma, and antiretroviral therapy.

(Interact J Med Res 2023;12:e46042) doi:10.2196/46042

KEYWORDS

bibliometric analysis; HIV; acquired immunodeficiency syndrome; AIDS; student; university; college; postsecondary; bibliometric; communicable; sexually transmitted disease; STD; sexual transmission; sexually transmitted infection; STI

Introduction

HIV/AIDS is a chronic infection that affects not only physical health but also social relationships, mental health, quality of life, and economic aspects. Students are the hope of their families and the future of a nation. Approximately 4000 individuals aged 15 years and older become newly infected with HIV every day worldwide, with 27.5% of them aged 15-24 years [1]. In 2017, approximately 19% of individuals aged 15-24 years living with HIV/AIDS in China were students [2]. Students living with HIV/AIDS could be experiencing body image issues; negative feelings; poor self-esteem; and especially at the university level, poor thinking, learning, memory, and concentration [3]. Therefore, the prevention and control of HIV infection in students must receive close attention.

In recent years, HIV infection among students has been an ongoing concern worldwide, such as knowledge of HIV/AIDS, risk behaviors, and HIV prevention education [4]. However, there is no systematic study of global research trends and guidelines in this area. Bibliometric analysis is a branch of quantitative science that has been used as a powerful tool for understanding emerging trends and knowledge structures in research fields and fostering new research ideas [5].

CiteSpace is an essential bibliometric analysis tool that facilitates the detection of emerging trends and mutations in a field [6]. It has been applied to research in more than 60 different scientific fields [7]. It plays an important role in describing keyword co-occurrence and cocited reference networks. CiteSpace can not only predict emerging trends of spatial epidemiology in infectious diseases [8] but also analyze patterns of relationships between nanosciences, health, and biology [9].

There may be articles that use other bibliometric analysis software; however, they only focused on specific students, such as college students [10]. This bibliometric analysis clearly illustrated the milestones and hotspots of research on HIV/AIDS among students from 1985 to 2022. Articles on HIV/AIDS among students were searched using the Web of Science Core Collection (WoSCC). Afterward, CiteSpace was used to perform statistical calculations and generate visual networks to reveal hotspots and frontiers of research.

Methods

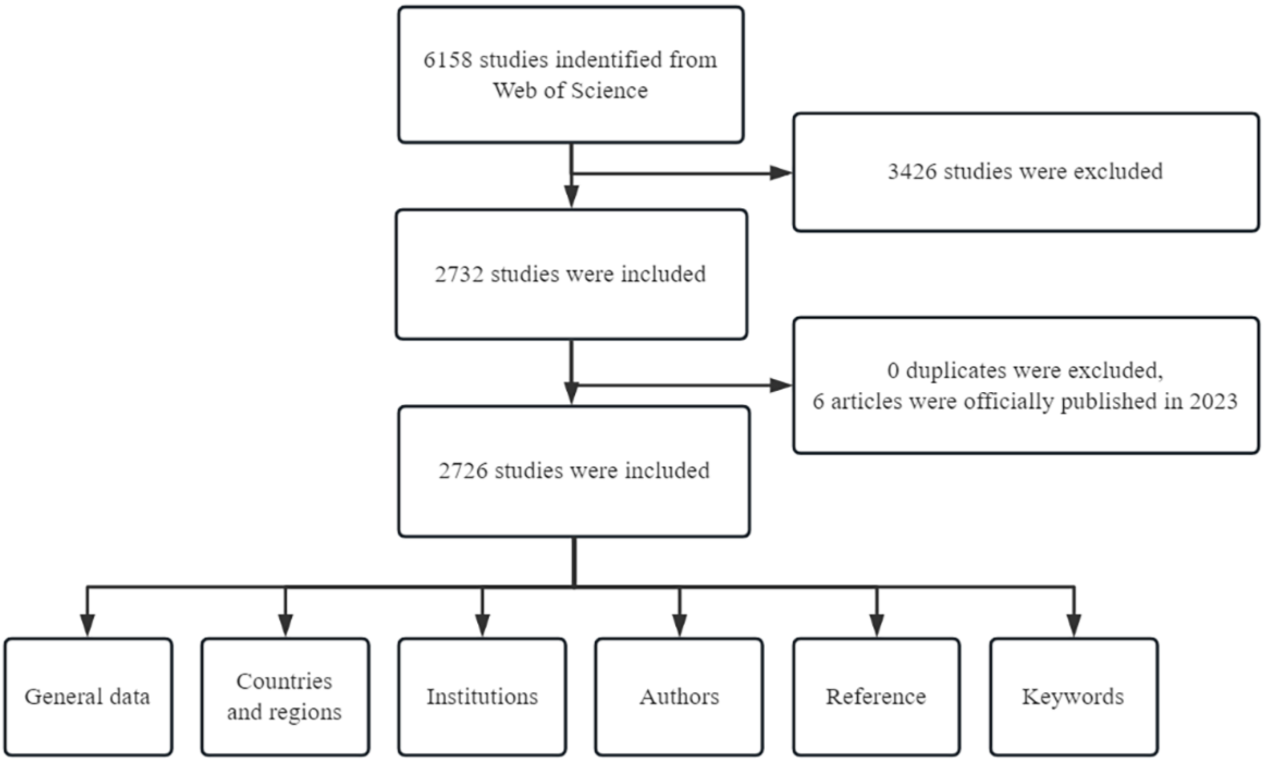
Data Sources and Search Strategies

The Web of Science database is an authoritative citation information source with the most selective journal coverage [11]. The data search was conducted using WoSCC on March 20, 2023. The research strategies were as follows: *TS=Topic, (TS="student$") AND ((TS="HIV") OR (TS="AIDS") OR (TS="Acquired Immune Deficiency Syndrome"))*, over the period from 1985 to 2022. A total of 6158 articles were obtained, but 3426 of them were manually excluded for not being relevant to the research content. Table 1 shows the inclusion and exclusion criteria, Figure 1 shows the study flowchart, and Multimedia Appendix 1 shows the complete research strategies and results. Study selection and data extraction were performed independently by 2 authors. Differences of opinion were settled by discussion or referral to a third author.

Table 1. Inclusion and exclusion criteria.

	Inclusion criteria	Exclusion criteria
Article type	Article	Reviews, book chapters, editorials, letters, commentaries, meeting abstracts, duplicate literature, etc
Language	English	Spanish, Portuguese, French, German, Russian, etc
Content	HIV/AIDS among students	Aid

Figure 1. Flowchart of the search strategy and bibliometric analysis process.



Data Preprocessing

We downloaded all records and references from WoSCC, including authors, titles, journals, years, keywords, abstracts, dates, countries, institutions, and references. Four folders were created, named "Input," "Output," "Data," and "Projects." All records and references were placed in the "Input" folder and then imported into CiteSpace to remove duplicates. After removing the duplicates, the articles appeared in the "Output" folder by year. Finally, the data in the "Output" folder were copied to the "Data" folder to be ready for analysis by CiteSpace.

Bibliometric Analysis

CiteSpace (version 6.2.R3) [12] was used to identify countries/regions and institutions contributing to research on HIV/AIDS among students and to reveal research hotspots and emerging trends. CiteSpace parameters were set as follows: (1) time slice from 1985 to 2022; (2) year per slice=1; and (3) pruning=pathfinder or pruning the merged network. Other parameters were set to default values.

Nodes indicated the object of analysis, including countries/regions, institutions, references, and keywords. The more frequently an object appears in the data set, the larger the node. A link between 2 nodes represents a copublishing partnership between 2 countries/regions or institutions [7]. In a network of keyword co-occurrence, a link represents the co-occurrence of 2 keywords in different articles [7]. It implied the association of 2 research contents. The thicker the line, the closer the relationship is between the 2 nodes.

The centrality of a node is a property that quantifies the importance of the node's position in a network [6]. Betweenness centrality is one of the most commonly used centrality metrics [13]. It measures the percentage of the shortest paths in the network to which a given node belongs [14]. A node with strong betweenness centrality can show a purple ring on the outside [6].

The analysis of keyword bursts can identify hotspots and frontiers that could have an impact on future research [15]. The analysis of citation bursts can reveal articles that had a significant impact in the field [16].

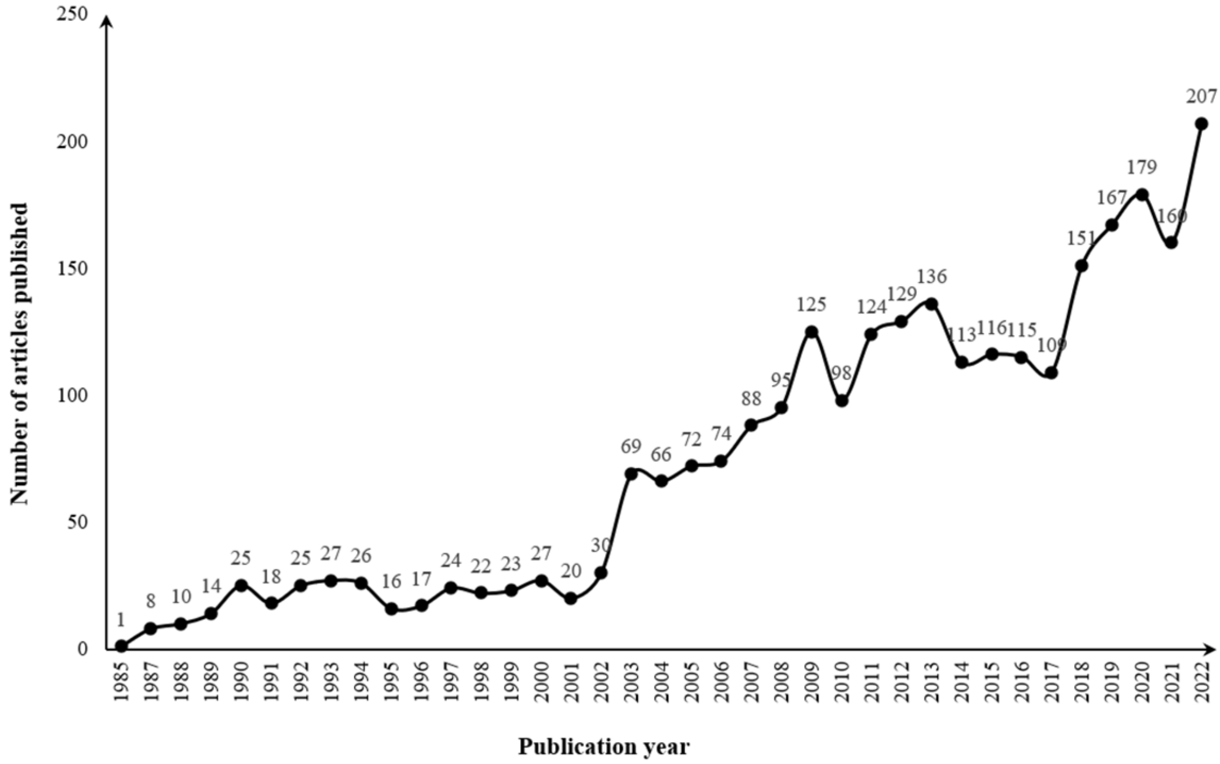
Results

General Data

A total of 2726 articles were included in this study. The trend in the number of articles reflects the popularity of research and the speed of knowledge growth [17]. The general trend in research on HIV/AIDS among students was on the rise, especially after 2002. The number of articles peaked in 2022, with 207 articles (Figure 2).

Research on HIV/AIDS among students can be divided into 3 stages. The preliminary stage was from 1985 to 2002, with 1 to 30 articles per year. It lasted 18 years but accounted for 12.22% (n=333) of the total number of articles. The research at this time laid the foundation and guided future research. In 1985, Price et al [18] published the first academic article on the assessment of high school students' perceptions and misperceptions of AIDS. It played an influential and leading role in research on HIV/AIDS among students.

Figure 2. Distribution of articles by publication year, from 1985 to 2022.



The developmental stage was from 2003 to 2017, with 66 to 136 articles per year. The total number of articles from this time period was 1529 (56.09%). A large number of articles had been accumulated during this stage of research. It can be regarded as a transition between low-level and high-level research. More and more professors and scholars had been attracted to this field, and enthusiasm for this research was high.

The superior stage was from 2018 to 2022, with 151 to 207 articles per year. The total number of articles from this time period was 864 (31.69%). Although there were small fluctuations in this period, studies on HIV/AIDS among students were generally on the rise.

Research Collaboration

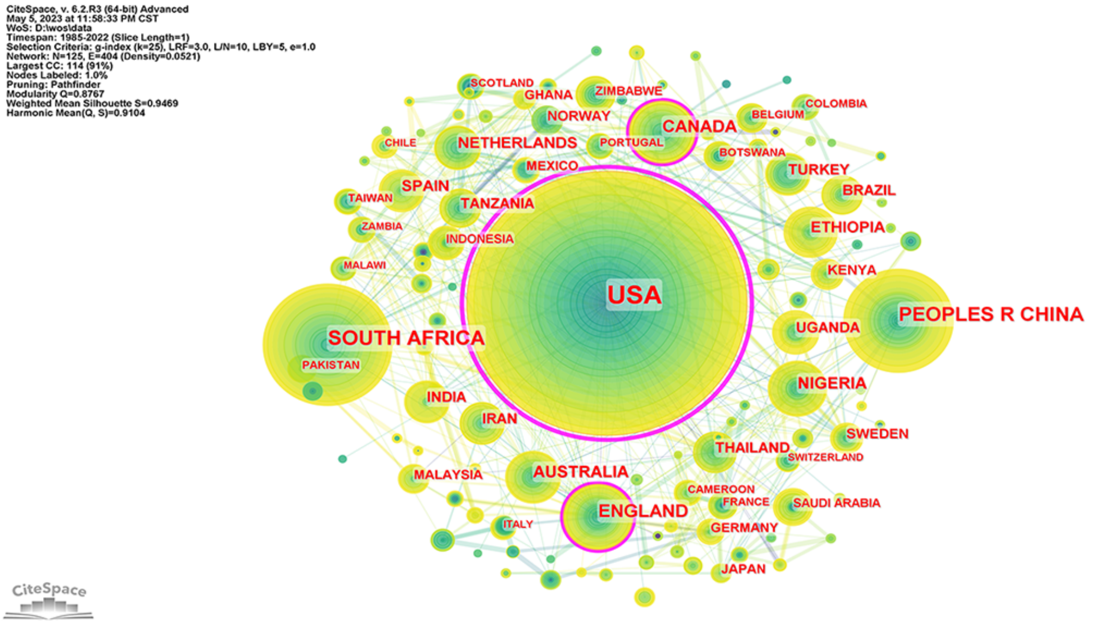
Countries/Regions

An analysis of the geographical distribution of published articles reflects the academic collaboration between countries/regions [19] (Figure 3). The size of the node indicates the number of articles published in different countries/regions [20]. The thicker the link, the closer the cooperation between the countries/regions. The United States contributed the most in terms of the number of articles (n=1303). South Africa (n=295) ranked second, and China (n=209) ranked third. These top 3 countries/regions accounted for 66.29% (1807/2726) of the total number of articles (Multimedia Appendix 2).

The United States was in the lead, with 47.8% (1303/2726) of the total articles and a betweenness centrality of 0.91. It had research collaborations with 69 countries/regions—much more than any other country. The world's first article on HIV/AIDS among students was published in 1985 by the University of Toledo in the United States [18]. It sparked the beginning of studies on HIV/AIDS among students. Research on HIV/AIDS among students in the United States not only began early but was of high quality.

South Africa had the second-highest number of articles, accounting for 10.82% (295/2726), but had a much lower betweenness centrality (0.09) than the United States. It had research collaborations with 24 countries/regions. The large number of people living with HIV/AIDS in South Africa has attracted the focused attention of experts. Experts have conducted extensive research on students and published numerous articles. In addition, countries/regions heavily affected by HIV/AIDS, such as Nigeria and Ethiopia, had published a large number of articles in this field.

Figure 3. Network of collaborative relationships among countries/regions. CC: co-citations; CST: Central Standard Time; LBY: look back year; L/N: maximum links per node; LRF: link retaining factor; WoS: Web of Science.



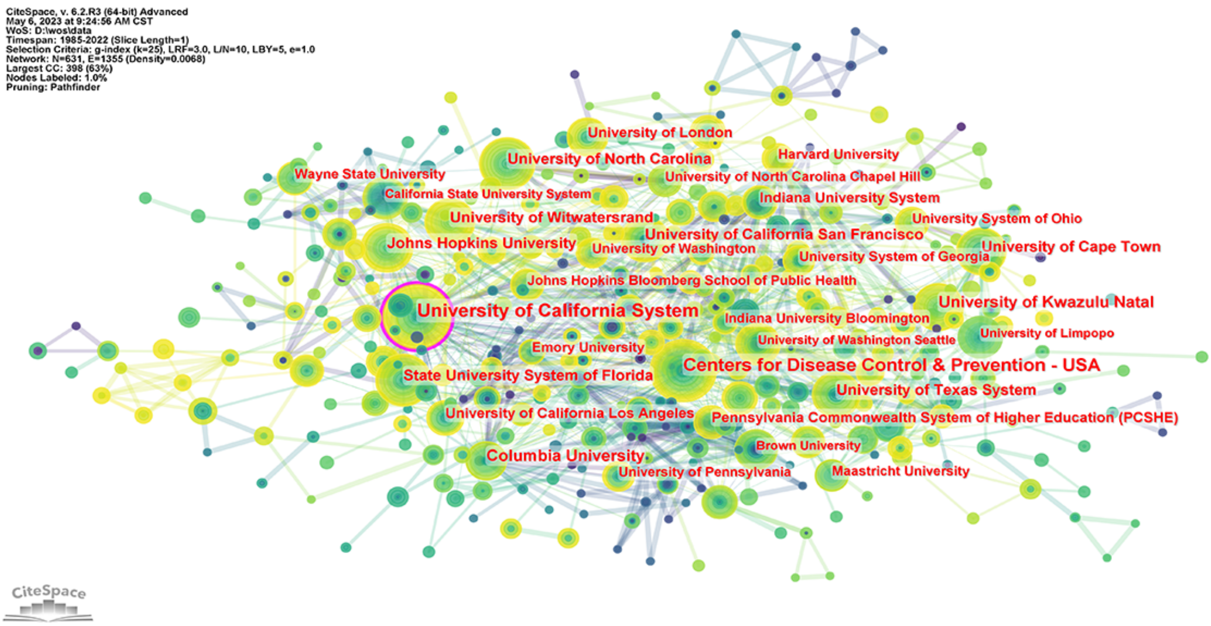
China ranked third with 7.67% (209/2726) of the total articles and a betweenness centrality of 0.07. It cooperated with 19 countries/regions. The first Chinese article on HIV/AIDS among students was published by the University of Hong Kong in 1999 [21]. The proportion of students with HIV in China increased year by year [2]. China's research in this field started late, but its quality has improved rapidly.

The nodes in the United States, England, and Canada had purple rings, which meant that they had high betweenness centrality. Experts from these countries had extensive international cooperation. In general, the number of articles was higher in countries/regions with high medical standards, such as the United States and England. Some countries/regions with high HIV prevalence also had a large number of articles, such as South Africa, Nigeria, and Ethiopia. However, the top 3 countries/regions for betweenness centrality were all medically advanced. The United States had the most cooperation with other countries/regions. A network of the US-centered academic collaborations has been formed, but collaboration among countries/regions needs to be strengthened.

Institutions

Figure 4 showed the major institutions in research on HIV/AIDS among students. The size of the node indicates the number of articles the institution had published [20]. The thicker the line, the closer the cooperation is between the 2 institutions [20]. Nodes with purple rings have high betweenness centrality.

Figure 4. Network of collaborative relationships among institutions. CC: co-citations; CST: Central Standard Time; LBY: look back year; L/N: maximum links per node; LRF: link retaining factor; WoS: Web of Science.



The University of California system not only had the most publications (n=125) but also had the highest betweenness centrality (0.12). Nine of the top 10 most productive institutions were universities, and the other was the Centers for Disease Control and Prevention. Of these, 8 institutions are from the United States and 2 are from South Africa. Institutions with a high number of publications are mainly concentrated in the United States (Multimedia Appendix 2).

In terms of research collaboration, primary cooperation networks have been formed. The University of California system was the only node with a purple ring, which had the highest betweenness centrality. This implies that its academic influence was so high that it was a central institution in this field. Recently, institutions have been working together more frequently and more closely than ever before.

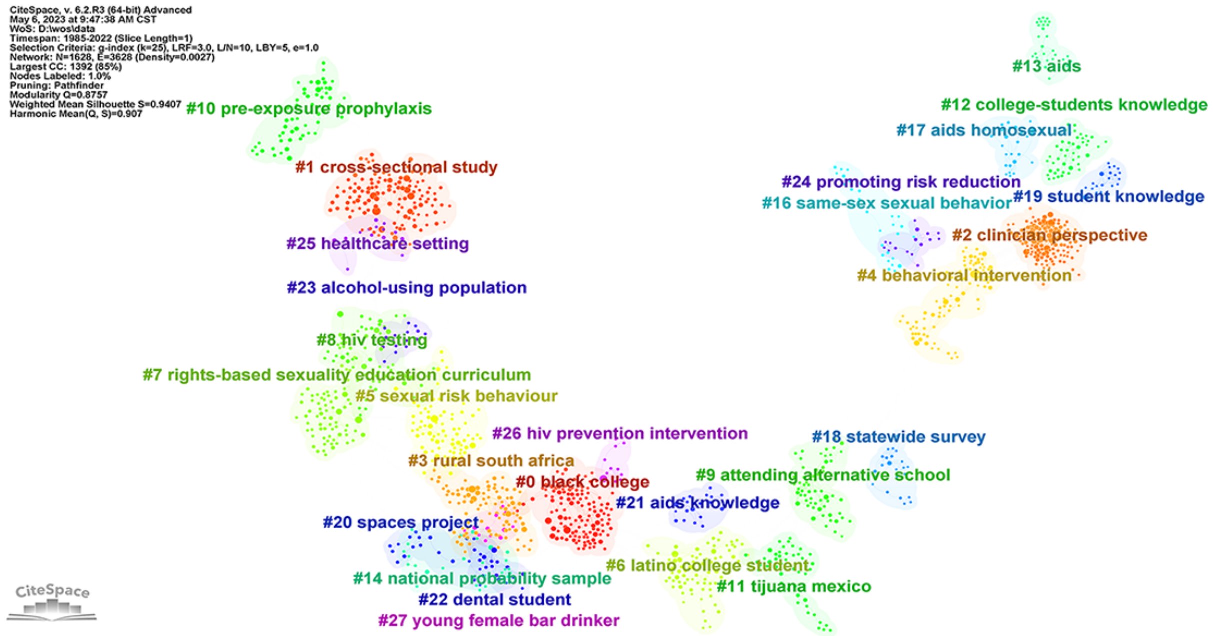
Knowledge Base Analysis

Documents cocitation analysis (DCA) refers to the frequency of 2 documents cited in the joint citation list [22]. The network formed by the cocited references can capture the research priorities of the basic science community [23]. Through DCA, we can discover milestones in the field and trace the roots of the knowledge base.

Reference Cocitation

The DCA network consisted of 1628 nodes and 3628 links, and a total of 27 major clusters were formed (Figure 5). Modularity Q can reflect the network structure and the clarity of clustering [24]. It ranges between 0 and 1. The closer the value is to 1, the better the modularity of the network. The silhouette is an indicator of the homogeneity of the members of the entire cluster [24]. It ranges between –1 and 1. The closer the value is to 1, the more homogeneous the cluster members are. In this network, modularity Q was 0.88 and silhouette was 0.94. This suggested that these clusters had analytical significance. CiteSpace provides 3 algorithms to calculate cluster labels: latent semantic indexing, log-likelihood ratio (LLR), and mutual information. Among them, LLR is the best choice to identify the most unique terms to the cluster [12]. Labels extracted by latent semantic indexing tend to capture implicit semantic relationships across data sets, whereas labels selected by LLR and mutual information tend to reflect a unique aspect of a cluster [25]. In the process of clustering, the results obtained by LLR were the most appropriate and most in line with the actual situation. Therefore, the algorithm used in this clustering was LLR.

Figure 5. Clusters of reference co-citation. CC: co-citations; CST: Central Standard Time; LBY: look back year; L/N: maximum links per node; LRF: link retaining factor; WoS: Web of Science.



Clusters #0, #6, #17, #22, #23, and #27 are about the research objects for HIV/AIDS among students. Clusters #3 and #11 are mainly about high-incidence areas of HIV/AIDS among students. Clusters #1, #2, #14, #18, and 20 are related to methods of research on HIV/AIDS among students. Clusters #4, #5, #7, #8, #9, #10, #12, #16, #19, #21, #24, #25, and #26 suggested that the content of the studies was focused on HIV/AIDS knowledge, risk behavior, education, and prevention.

Most-Cited Articles

The most-cited article in our data set is Li et al [26] with 24 citations, followed by Hingson et al [27] with 21 citations. Walter and Vaughan [28], Shisana et al [29], and Weinstock et al [30] are tied for third with 19 citations each.

These most-cited articles had much in common. They were almost always related to HIV/AIDS knowledge, attitudes, and behaviors. However, they were studied from different angles. Li et al [26] focused on making suggestions for HIV/AIDS prevention among students from a policy perspective. Walter and Vaughan [28] divided students into an intervention group and a comparison group to evaluate the effect of a HIV curriculum on reducing HIV risk among students. The other 3 articles [27,29,30] analyzed knowledge, attitudes, and behaviors about HIV/AIDS in the form of surveys.

Citation Burst

If an article is cited frequently over a short time period, it is considered a reference with strong citation burst [6]. The first article with strong citation burst is Li et al [26] with a burst strength of 12.76. The number of citations to this article increased substantially in 2020 and continues to the present.

Sigma

Sigma is a value used in CiteSpace to measure the novelty of a node, which combines the importance of the node in the network structure (betweenness centrality) and the importance of the node in time (bursts) [23]. Nodes with great betweenness centrality and bursts have higher sigma values. The pioneering article by Walter and Vaughan [28] has the highest sigma of 80.96, which means it has both strong betweenness centrality and citation burst.

Research Hotspots and Evolutionary Trends

Analysis of Research Hotspots

Keywords are a summary of the content of the article. The analysis of keywords can reveal the development process, explore research hotspots, and especially predict the future development trend of a field [31]. The basic principle of keyword co-occurrence analysis is to calculate the co-occurrence frequency of keywords in different articles and use the co-occurrence frequency to measure the relationship between keywords [8]. There were 752 nodes and 4484 links in this network. Nodes represent keywords. The thicker the line, the more times 2 keywords appear together in different articles (Figure 6).

The keyword with the largest betweenness centrality was “United States” (0.09), whereas the keyword with the greatest count was “HIV” (n=494). These words were mainly related to students’ knowledge, attitude, risk behavior, and prevention of HIV/AIDS. The results implied that the hotspots were closely related to students’ knowledge, attitudes, and risk behaviors toward HIV (Table 2).

Figure 6. Network of keyword co-occurrence. CC: co-citations; CST: Central Standard Time; LBY: look back year; L/N: maximum links per node; LRF: link retaining factor; WoS: Web of Science.

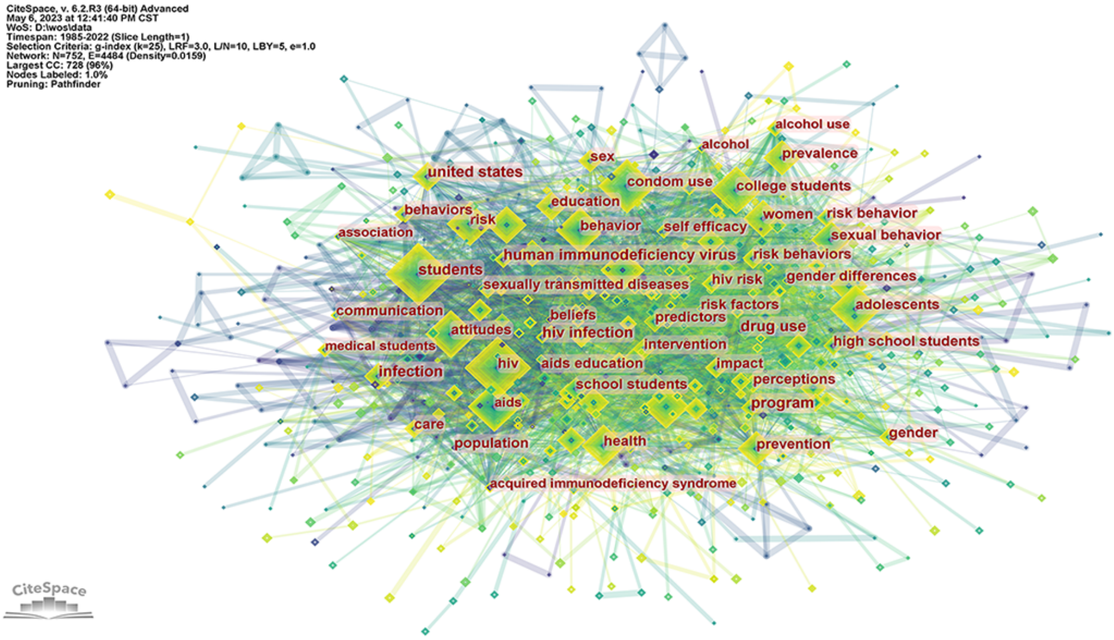


Table 2. Top 20 high-frequency keywords in studies on HIV/AIDS among students, from 1985 to 2022.

Keyword	Count, n	Centrality
HIV	494	0.04
Students	472	0.05
Attitudes	393	0.03
Condom use	387	0.04
AIDS	386	0.03
Adolescents	369	0.02
Knowledge	346	0.02
Risk	341	0.04
College students	326	0.05
Behavior	286	0.05
HIV/AIDS	276	0.02
Health	239	0.05
Sexual behavior	238	0.05
Prevention	233	0.05
Women	229	0.04
Prevalence	199	0.03
Education	198	0.06
HIV prevention	198	0.03
University students	174	0.02
Infection	172	0.08

Keyword Clustering Analysis

CiteSpace can be used for keyword clustering [15]. Similar keywords can be grouped into a cluster (Figure 7). Normally, modularity Q>0.3 and silhouette>0.7 indicate that map clustering is appropriate [15]. In this network, the modularity Q was 0.34 and the silhouette value was 0.70, implying that the clustering was analytically meaningful. The results obtained by LLR were the most appropriate and most in line with the actual situation. Therefore, this study used the LLR algorithm to extract the clustering labels from the keywords of articles. The results revealed that dental students and university students were the main targets of research on HIV/AIDS among students. The content was focused on knowledge, risk behavior, and prevention (Figure 7 and Table 3).

Figure 7. Clusters of keywords. CC: co-citations; CST: Central Standard Time; LBY: look back year; L/N: maximum links per node; LRF: link retaining factor; WoS: Web of Science.

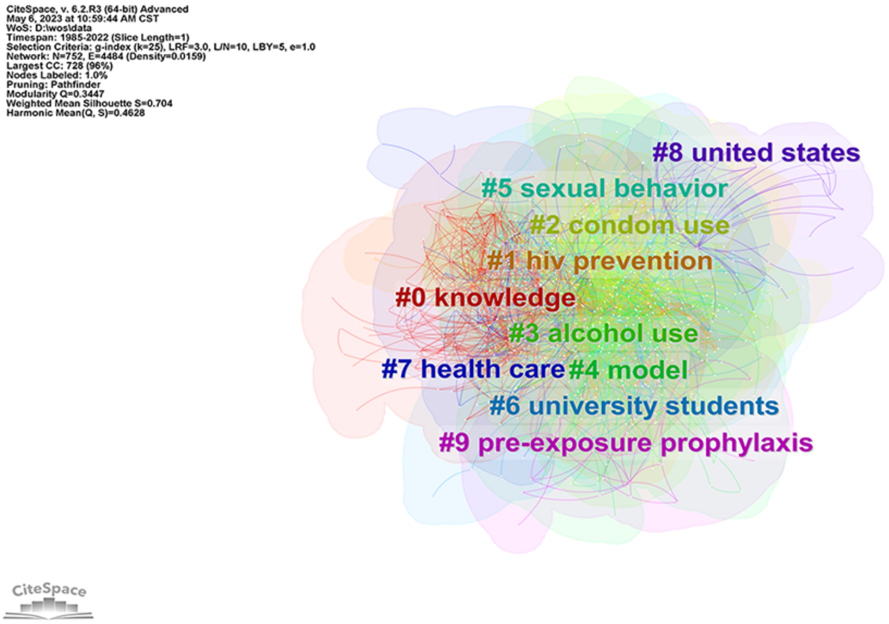


Table 3. Keyword clusters in studies on HIV/AIDS among students, from 1985 and 2022.

Cluster ID	Label (LLR[a])	Size, n	Silhouette	Major included keywords	Mean years
0	Knowledge	110	0.74	Dental students, attitudes, nursing students, and stigma	2003
1	HIV prevention	103	0.67	Peer education, program, sex education, and South Africa	2005
2	Condom use	96	0.71	Risk, behavior, college students, and HIV infection	1997
3	Alcohol use	90	0.72	Substance use, mental health, sensation seeking, and HIV risk	2005
4	Model	83	0.61	Determinants, health belief model, sex, and planned behavior	2006
5	Sexual behavior	70	0.67	Sexually transmitted infections, risk factors, high school students, and stigma	2007
6	University students	65	0.66	Sexual behavior, risk perception, HIV testing, and sexually transmitted infection	2006
7	Health care	54	0.74	Women, parent-child communication, serious games, and health communication	2012
8	United States	30	0.86	HIV/AIDS, children, sexual minority, and judgments	2002

[a]LLR: log-likelihood ratio.

Hotspot Research Objects

In studies of HIV/AIDS among students, the study population consisted mainly of medical students and university students. Medical students have more contact with patients living with HIV/AIDS. Their attitude toward patients living with HIV/AIDS affects the quality of care provided to these patients. They are also at high risk of occupational exposure. Additionally, common university students are at risk of HIV infection through unprotected sex [32].

Attitude, Knowledge, and Prevention

Attitudes can help medical students overcome fear and discrimination about HIV/AIDS [33]. For the general university student, HIV-related stigma was one of the strongest barriers to HIV testing and treatment [34]. Through HIV/AIDS education, students gained sufficient knowledge to increase awareness of HIV prevention, reduce the risk of infection, and reduce AIDS-related stigma [35,36].

"Risk Behavior" Including Clusters #2, #3, and #5

A cross-sectional study on the risk of HIV transmission among medical students found that 29.13% reported occupational injuries due to needle exposure [37]. Occupational exposure increased the risk of HIV/AIDS infection among medical students. For the common university student, risk behaviors for HIV infection are mainly unprotected sex and substance use [38].

Keyword Burst Analysis

An article can be regarded as information flow that arrive continuously over time [39]. The Kleinberg [40] algorithm formalizes the modeling of burst information flow so that burst information flow can be effectively identified. It can be used to detect a sudden increase in research interest in a particular discipline. CiteSpace uses the Kleinberg algorithm to identify emerging research frontiers [6]. It can reveal the frontiers of research at different stages and predict future research directions.

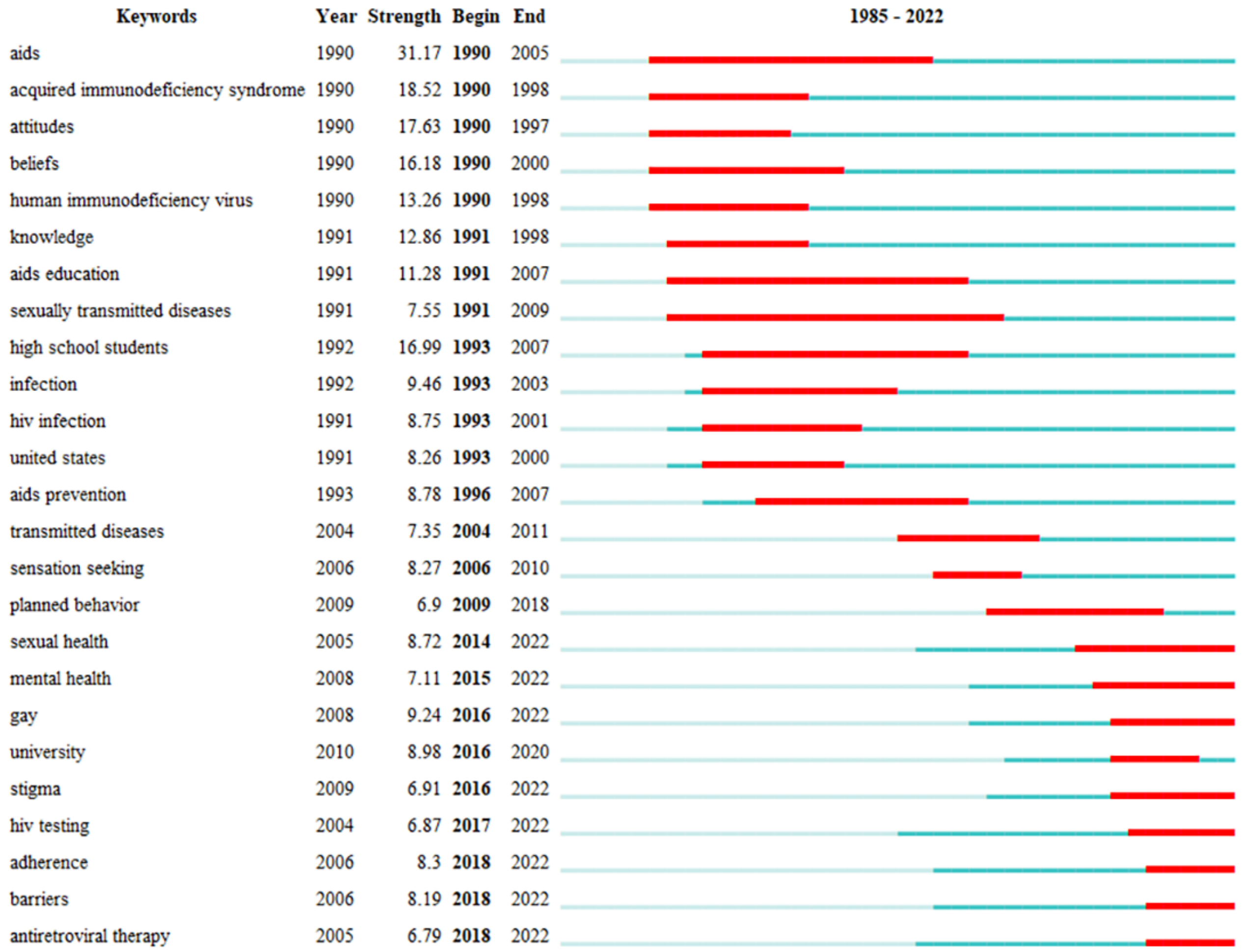
Figure 8 shows the top 25 keywords with the strongest citation bursts. “Year” means the year the keyword first appeared. “Begin” refers to the year in which the keyword’s occurrence frequency increased greatly. “End” represents the year in which the popularity of the keyword declined. The blue line represents the timeline, and the red line represents the time when the keyword burst.

From 1985 to 2002, the study population was mainly high school students. Experts focused on students’ knowledge, attitudes, education, and prevention of HIV/AIDS. During this period, there was a wide range of research content and an increase in AIDS research involving students.

From 2003 to 2017, research on HIV/AIDS among students achieved initial results, and the frontiers were deeper than at the previous stage. Compared with the previous period, the object of research was more specific. The hotspot object changed from “high school student” to “gay.” Second, the scope of hotspots had been narrowed. The research frontier had changed from “attitude,” “knowledge,” “education,” and “prevention” to “sensation seeking,” “sexual health,” and “stigma.” Third, the research methods were richer than before. For example, since 2009, the Theory of Planned Behavior has been widely used by experts to estimate students’ attitudes and behaviors toward HIV/AIDS.

From 2018 to 2022, the emerging area of research was adherence, barriers, and antiretroviral therapy. During this period, “adherence” (burst strength=8.30) was the keyword with the strongest citation burst. It is worth noting that several keywords continue to be popular right now, such as gay (9.24), sexual health (8.72), adherence (8.30), barriers (8.19), mental health (7.11), stigma (6.91), HIV testing (6.87), and antiretroviral therapy (6.79).

Figure 8. Top 25 keywords with the strongest citation bursts.



Discussion

Principal Findings

In this study, bibliometric analysis was used to analyze articles related to HIV/AIDS among students from 1985 to 2022. The United States contributed 47.8% of the total articles, and its betweenness centrality (0.91) was much higher than that of other countries/regions. The United States was the core country for studies on HIV/AIDS among students. South Africa had the second-highest number of articles (10.82%), but its betweenness centrality (0.09) was much lower than that of the United States. In general, the number of articles on HIV/AIDS among students was higher in countries/regions with advanced medical systems and some countries/regions with numerous patients living with HIV/AIDS. An academic collaboration network centered on the United States had been formed, but cooperation among countries/regions needs to be strengthened. The cooperation network of institutions was formed. The University of California system is the core institution in this field. In the future, further collaboration among countries/regions or institutions should be encouraged to promote the flourishing of research on HIV/AIDS among students.

With keyword co-occurrence and cluster analysis, the most important topics and information can be easily figured out [16]. If a keyword appears frequently over a short time period, it can be considered a research hotspot [15]. Accordingly, CiteSpace was used to constantly detect changes in high-frequency keywords to accurately explore the trends in the field. The results showed that hotspots in studies on HIV/AIDS among students were constantly changing. In terms of study objects, the early research objects were school students; in the medium term, studies of medical students, college students, and African American students increased; recently, sexual minority communities, especially men who have sex with men (MSM), have been the focus of research. In terms of study content, in the early years, experts studied students' knowledge and attitudes toward HIV/AIDS and sexual behavior. Then, substance abuse, HIV/AIDS prevention, and education gained a great deal of attention. Recently, research on stigma, HIV testing, and antiretroviral prophylaxis has become increasingly popular. In terms of study methods, earlier studies were conducted mainly through cross-sectional studies and qualitative research. In the medium term, randomized controlled trials were added to the study methods, and the Information-Motivation-Behavioral model was used in research on HIV/AIDS among students. Recently, there has been a tremendous enrichment of research methods. Disorders identification tests, clinical research, implementation science, digital health intervention, and other emerging methods are becoming more widespread in the field. In addition, the application of biopsychosocial model is an emerging practice that has been applied since 2020. The research content was getting deeper and deeper, and the research level kept rising.

The research frontiers have been changing over time since 1985. From 1985 to 2002, the research frontiers were mainly about the initial understanding of HIV/AIDS; from 2003 to 2017, they were mainly focused on the sexual health and stigma of HIV/AIDS; and from 2018 to 2022, they focused on barriers to HIV prevention, HIV testing and treatment adherence, and antiretroviral therapy. Keywords that are still popular today can provide clues for future research, such as gay, sexual health, adherence, barriers, mental health, HIV testing, stigma, and antiretroviral therapy. Risky sexual behavior is popular among college students and has been proven to be a high-risk factor for HIV/AIDS among students [38]. Through strengthening sex health education, students can increase their knowledge of HIV and reduce risky sexual behavior, thereby preventing HIV infection. In addition, the proportion of students who contracted HIV through sexual contact among MSM students is also rising rapidly [41]. However, out of 2726 articles, there were only 343 studies on HIV/AIDS infection in MSM students. It suggests that research on HIV/AIDS among MSM students is still inadequate and should be given more attention. Considering the increasing number of students living with HIV, antiretroviral therapy is particularly important. Experts analyzed the facilitators and barriers to antiretroviral therapy adherence among student with AIDS through qualitative studies [42]. In particular, stigma is a major barrier to students' adherence to HIV testing and antiretroviral therapy. Reducing stigma can help students improve their adherence to HIV testing and antiretroviral therapy [43]. In the future, research can continue on the hotspots in the suggested direction or try to explore the interaction between the hotspots and find their commonalities to obtain new findings.

In bibliometrics analysis, the citing articles constitute the research frontier, and the cited articles constitute the knowledge base [44]. The most frequently cited articles can be considered milestones in a certain field [45]. Through the analysis of cited articles, it was found that a number of experts used questionnaires to assess students' knowledge, attitudes, and behaviors regarding HIV/AIDS. Then, problems were identified and summarized to make suggestions for HIV/AIDS prevention and control among students using the information returned. These cited articles laid the foundation for research on HIV/AIDS among students.

Limitations

There are limitations to this study. First, our study focused on English publications, which inevitably missed some important studies published in other languages. Second, although we used WoSCC in our bibliometric analysis, there may be some articles from other databases (eg, Scopus, MEDLINE, and PubMed) that were not retrieved. However, this study provides a bibliometric analysis of publications on HIV/AIDS among students and is based only on Web of Science data. Through visual analysis software, readers can clearly understand the number of articles, academic cooperation, research hotspots, and research frontiers. It provides hotspots and emerging trends for future research.

Conclusions

The study identified countries/regions and institutions contributing to the research area of HIV/AIDS among students and revealed research hotspots and emerging trends. The field of research on HIV/AIDS among students was growing rapidly. The United States was at the center, and the University of

California system was the core institution. However, academic collaboration should be strengthened. Future research may focus on exploring gay students, sexual health, adherence, barriers, mental health, HIV testing, stigma, and antiretroviral therapy.

Acknowledgments

This work was supported by the Technical Application and Promotion Plan of Guilin City (grant 2020011203-6), National Natural Science Foundation of China (grants 82160636, 12071366, 82160642, and 82260670), Guangxi Natural Science Foundation Project (grant 2020GXNSFAA159020), Guangxi Key Laboratory of AIDS Prevention Control and Translation (grant ZZH2020010), Ministry of Science and Technology of China (grant 2022YFC2305200), Guangxi Bagui Honor Scholarship, and Chinese State Key Laboratory of Infectious Disease Prevention and Control.

Data Availability

All data generated or analyzed during this study are included in this published article and its Multimedia Appendices.

Authors' Contributions

NW, RZ, and ZY analyzed the data and wrote the manuscript. GL and QZ acquired the data. HC, XZ, and ST filtered the data. YR and ML designed the research and revised the manuscript. All authors contributed to the article and approved the submitted version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Research strategies and results.
[ZIP File (Zip Archive), 6010 KB - ijmr_v12i1e46042_app1.zip]

Multimedia Appendix 2
Countries/regions and institutions data.
[DOCX File , 16 KB - ijmr_v12i1e46042_app2.docx]

References

1. In danger: UNAIDS global AIDS update 2022. UNAIDS. 2022 Jul 27. URL: https://www.unaids.org/en/resources/documents/2022/in-danger-global-aids-update [accessed 2022-12-01]
2. Zhang X, Wang N, Vermund SH, Zou H, Li X, Zhang F, et al. Interventions to improve the HIV continuum of care in China. Curr HIV/AIDS Rep 2019 Dec;16(6):448-457. [doi: 10.1007/s11904-019-00469-y] [Medline: 31776975]
3. Cronje JH, Williams M, Steenkamp L, Venter D, Elkonin D. The quality of life of HIV-infected South African university students: experiences with the WHOQOL-HIV-Bref. AIDS Care 2017 May 18;29(5):632-635. [doi: 10.1080/09540121.2016.1234688] [Medline: 27643964]
4. Shiferaw Y, Alemu A, Girma A, Getahun A, Kassa A, Gashaw A, et al. Assessment of knowledge, attitude and risk behaviors towards HIV/AIDS and other sexual transmitted infection among preparatory students of Gondar town, north west Ethiopia. BMC Res Notes 2011 Nov 21;4(1):505 [FREE Full text] [doi: 10.1186/1756-0500-4-505] [Medline: 22099952]
5. Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. How to conduct a bibliometric analysis: an overview and guidelines. J Bus Res 2021 Sep;133:285-296. [doi: 10.1016/j.jbusres.2021.04.070]
6. Chen C. CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature. J Am Soc Inf Sci 2006 Feb 01;57(3):359-377. [doi: 10.1002/asi.20317]
7. Chen C. CiteSpace: A Practical Guide For Mapping Scientific Literature. UK edition. Hauppauge, NY: Nova Science Publishers; 2016.
8. Dai F, Liu H, Zhang X, Li Q. Exploring the emerging trends of spatial epidemiology: a scientometric analysis based on CiteSpace. SAGE Open 2021 Nov 23;11(4). [doi: 10.1177/21582440211058719]
9. Novo MS, Geracitano LA, Henning P. Padrão de relacionamento entre nanociências, saúde e biologia: um levantamento histórico utilizando o programa Citespace. The pattern of relationship between nanosciences, health, and biology: a historical survey using Citespace. Hist Cienc Saude-Manguinhos 2013 Oct;20(4):1657-1670. [doi: 10.1590/s0104-59702013005000008] [Medline: 23827968]
10. Zheng Q, Duan D, Zhu Q, Zhang J, Xiong L, Ullayin S. Hotspots and frontier trends of AIDS research among Chinese and foreign college students. Mod Prev Med 2022;49(8):1506-1513 [FREE Full text] [doi: 10.11656/j.issn.1673-9043.2020.05.15]

11. Singh VK, Singh P, Karmakar M, Leta J, Mayr P. The journal coverage of Web of Science, Scopus and Dimensions: a comparative analysis. Scientometrics 2021 Mar 26;126(6):5113-5142. [doi: 10.1007/s11192-021-03948-5]
12. Chen C. How to Use CiteSpace. Victoria, BC: Leanpub; Sep 12, 2020.
13. Freeman LC. Centrality in social networks conceptual clarification. Soc Networks 1978 Jan;1(3):215-239. [doi: 10.1016/0378-8733(78)90021-7]
14. Bader D, Kintali S, Madduri K, Mihail M. Approximating betweenness centrality. 2007 Presented at: WAW 2007: Algorithms and Models for the Web-Graph; December 11-12, 2007; San Diego, CA p. 124-137. [doi: https://doi.org/10.1007/978-3-540-77004-6_10]
15. Chen H, Zhao G, Xu N. The analysis of research hotspots and fronts of knowledge visualization based on CiteSpace II. 2012 Presented at: ICHL 2012: International Conference on Hybrid Learning; August 13-15, 2012; Guangzhou, China p. 57-68. [doi: 10.1007/978-3-642-32018-7_6]
16. Chen C. Mapping Scientific Frontiers: The Quest for Knowledge Visualization. London, the United Kingdom: Springer; 2013.
17. Liu J, Kong X, Zhou X, Wang L, Zhang D, Lee I, et al. Data mining and information retrieval in the 21st century: a bibliographic review. Comput Sci Rev 2019 Nov;34:100193. [doi: 10.1016/j.cosrev.2019.100193]
18. Price JH, Desmond S, Kukulka G. High school students' perceptions and misperceptions of AIDS. J Sch Health 1985 Mar;55(3):107-109. [doi: 10.1111/j.1746-1561.1985.tb04091.x] [Medline: 3845256]
19. Liu D, Che S, Zhu W. Visualizing the knowledge domain of academic mobility research from 2010 to 2020: a bibliometric analysis using CiteSpace. SAGE Open 2022 Jan 13;12(1). [doi: 10.1177/21582440211068510]
20. Chen Y, Chen C, Hu Z, Wang X. Principles and Applications of Analyzing a Citation Space. Beijing, China: Science Press; 2014.
21. Davis C, Tang CS, Chan SF, Noel B. The development and validation of the International Aids Questionnaire-Chinese version (Iaq-C). Educ Psychol Meas 2016 Jul 02;59(3):481-491. [doi: 10.1177/00131649921969992]
22. Small H. Co-citation in the scientific literature: a new measure of the relationship between two documents. J Am Soc Inf Sci 1973 Jul;24(4):265-269. [doi: 10.1002/asi.4630240406]
23. Chen C, Ibekwe-SanJuan F, Hou J. The structure and dynamics of cocitation clusters: a multiple-perspective cocitation analysis. J Am Soc Inf Sci 2010 Mar 18;61(7):1386-1409. [doi: 10.1002/asi.21309]
24. Chen C. Searching for intellectual turning points: progressive knowledge domain visualization. Proc Natl Acad Sci U S A 2004 Apr 06;101 Suppl 1(Suppl 1):5303-5310 [FREE Full text] [doi: 10.1073/pnas.0307513100] [Medline: 14724295]
25. Zhu Y, Kim MC, Chen C. An investigation of the intellectual structure of opinion mining research. Inf Res 2017 Mar;22(1):739 [FREE Full text]
26. Li G, Jiang Y, Zhang L. HIV upsurge in China's students. Science 2019 May 24;364(6442):711. [doi: 10.1126/science.aay0799] [Medline: 31123112]
27. Hingson R, Strunin L, Berlin B. Acquired immunodeficiency syndrome transmission: changes in knowledge and behaviors among teenagers, Massachusetts statewide surveys, 1986 to 1988. Pediatrics 1990 Jan;85(1):24-29. [Medline: 2296490]
28. Walter HJ, Vaughan RD. AIDS risk reduction among a multiethnic sample of urban high school students. JAMA 1993 Aug 11;270(6):725-730. [Medline: 8336374]
29. Shisana O, Risher K, Celentano DD, Zungu N, Rehle T, Ngcaweni B, et al. Does marital status matter in an HIV hyperendemic country? Findings from the 2012 South African National HIV Prevalence, Incidence and Behaviour Survey. AIDS Care 2016;28(2):234-241 [FREE Full text] [doi: 10.1080/09540121.2015.1080790] [Medline: 26551532]
30. Weinstock H, Berman S, Cates W. Sexually transmitted diseases among American youth: incidence and prevalence estimates, 2000. Perspect Sex Reprod Health 2004 Feb 12;36(1):6-10. [doi: 10.1363/psrh.36.6.04] [Medline: 14982671]
31. Gu D, Li J, Li X, Liang C. Visualizing the knowledge structure and evolution of big data research in healthcare informatics. Int J Med Inform 2017 Feb;98:22-32. [doi: 10.1016/j.ijmedinf.2016.11.006] [Medline: 28034409]
32. Alcocer-Bruno C, Ferrer-Cascales R, Ruiz-Robledillo N, Sánchez-SanSegundo M, Zaragoza-Martí A. Personal and lifestyle determinants of HIV transmission risk in Spanish university students. Int J Environ Res Public Health 2020 Nov 11;17(22):8332 [FREE Full text] [doi: 10.3390/ijerph17228332] [Medline: 33187170]
33. Frain JA. Preparing every nurse to become an HIV nurse. Nurse Educ Today 2017 Jan;48:129-133 [FREE Full text] [doi: 10.1016/j.nedt.2016.10.005] [Medline: 27810630]
34. James TG, Ryan SJ. HIV knowledge mediates the relationship between HIV testing history and stigma in college students. J Am Coll Health 2018 Oct 26;66(7):561-569. [doi: 10.1080/07448481.2018.1432623] [Medline: 29405896]
35. Zhang L, Li X, Mao R, Stanton B, Zhao Q, Wang B. Stigmatizing attitudes towards people living with HIV/AIDS among college students in China: implications for HIV/AIDS education and prevention. Health Educ 2008 Feb 22;108(2):130-144. [doi: 10.1108/09654280810855586]
36. Jahić R, Porobić-Jahić H, Žepić D, Piljić D, Petrović J, Čustović A. Knowledge, attitude and stigma towards HIV patients: a survey among medical students in Tuzla, Bosnia and Herzegovina. J Infect Dev Ctries 2020 Sep 30;14(9):1019-1026 [FREE Full text] [doi: 10.3855/jidc.12526] [Medline: 33031091]
37. Chernyshov PV. Risk factors for HIV transmission and HIV testing among medical students. Acta Dermatovenerol Croat 2020 Jul;28(1):2-8. [Medline: 32650844]

38. Heads AM, Dickson JW, Asby AT. Correlates of HIV risk-taking behaviors among African-American college students: HIV knowledge and ethnic identity. J Health Care Poor Underserved 2017 May;28(2S):155-170. [doi: 10.1353/hpu.2017.0058] [Medline: 28458270]
39. Ahlswede R, Ning Cai, Li S, Yeung R. Network information flow. IEEE Trans Inform Theory 2000 Jul;46(4):1204-1216. [doi: 10.1109/18.850663]
40. Kleinberg J. Bursty and hierarchical structure in streams. 2002 Jul 23 Presented at: KDD '02: the eighth ACM SIGKDD international conference on Knowledge discovery and data mining; July 23-26, 2002; Edmonton, AB. [doi: 10.1145/775047.775061]
41. Kort DN, Samsa GP, McKellar MS. Sexual orientation differences in HIV testing motivation among college men. J Am Coll Health 2017 Apr 06;65(3):223-227. [doi: 10.1080/07448481.2016.1277429] [Medline: 28059669]
42. Mutwa PR, Van Nuil JI, Asiimwe-Kateera B, Kestelyn E, Vyankandondera J, Pool R, et al. Living situation affects adherence to combination antiretroviral therapy in HIV-infected adolescents in Rwanda: a qualitative study. PLoS One 2013;8(4):e60073 [FREE Full text] [doi: 10.1371/journal.pone.0060073] [Medline: 23573232]
43. Johnson-Peretz J, Lebu S, Akatukwasa C, Getahun M, Ruel T, Lee J, et al. "I was still very young": agency, stigma and HIV care strategies at school, baseline results of a qualitative study among youth in rural Kenya and Uganda. J Int AIDS Soc 2022 Jul;25 Suppl 1(Suppl 1):e25919 [FREE Full text] [doi: 10.1002/jia2.25919] [Medline: 35818888]
44. Kotloff KL, Tacket CO, Clemens JD, Wasserman SS, Cowan JE, Bridwell MW, et al. Assessment of the prevalence and risk factors for human immunodeficiency virus type 1 (HIV-1) infection among college students using three survey methods. Am J Epidemiol 1991 Jan;133(1):2-8. [doi: 10.1093/oxfordjournals.aje.a115797] [Medline: 1983894]
45. Chen C, Hu Z, Liu S, Tseng H. Emerging trends in regenerative medicine: a scientometric analysis in CiteSpace. Expert Opin Biol Ther 2012 May;12(5):593-608. [doi: 10.1517/14712598.2012.674507] [Medline: 22443895]

Abbreviations

DCA: documents cocitation analysis
LLR: log-likelihood ratio
MSM: men who have sex with men
WoSCC: Web of Science Core Collection

Edited by T de Azevedo Cardoso, A Mavragani; submitted 30.01.23; peer-reviewed by E Ole, P Dabas, J Jiang; comments to author 09.03.23; revised version received 08.04.23; accepted 24.05.23; published 04.08.23.

Please cite as:
Wang N, Zhang R, Ye Z, Lan G, Zhu Q, Chen H, Zhang X, Tan S, Ruan Y, Lin M
Studies on HIV/AIDS Among Students: Bibliometric Analysis
Interact J Med Res 2023;12:e46042
URL: https://www.i-jmr.org/2023/1/e46042
doi:10.2196/46042
PMID:37540553

Review

Mobile Apps Aimed at Preventing and Handling Unintentional Injuries in Children Aged <7 Years: Systematic Review

Annett Schulze[1*], Dr phil; Ann-Kathrin Lindemann[1*], Dr rer soc; Fabian Brand[1], MA; Johanna Geppert[1], MA; Axel Menning[1], Dr rer pol; Paula Stehr[2], Dr phil; Doreen Reifegerste[3], Prof Dr; Constanze Rossmann[2], Prof Dr

[1]Department of Risk Communication, German Federal Institute for Risk Assessment, Berlin, Germany

[2]Department of Media and Communication, Ludwig-Maximilians-University Munich, Munich, Germany

[3]Department of Prevention and Health Promotion, Bielefeld University, Bielefeld, Germany

[*]these authors contributed equally

Corresponding Author:
Annett Schulze, Dr phil
Department of Risk Communication
German Federal Institute for Risk Assessment
Max-Dohrn-Str 8-10
Berlin, 10589
Germany
Phone: 49 18412 52002
Email: annett.schulze@bfr.bund.de

Abstract

Background: Despite various global health crises, the prevention and handling of unintentional childhood injuries remains an important public health objective. Although several systematic reviews have examined the effectiveness of different child injury prevention measures, these reviews did not address the evaluation of mobile communication intervention tools. Whether and how mobile apps were evaluated provides information on the extent to which communication theories, models, and evidence-based knowledge were considered. Previous studies have shown that the effectiveness of mobile apps increases when theories and evidence are considered during their development.

Objective: This systematic review aimed to identify research on mobile apps dealing with the prevention and handling of unintentional injuries in children and examine the theoretical and methodological approaches thereof. In addition, this review analyzed the different needs of various target groups of the mobile apps described in the articles.

Methods: In total, 8 electronic databases, ranging from interdisciplinary to medical and technical as well as social sciences databases, were searched for original research articles or brief reports in peer-reviewed journals or conference proceedings. Moreover, this review encompassed a systematic scan of articles published in the *BMJ* journal *Injury Prevention*. These steps were followed by a snowball search based on the literature references in the articles identified through the initial screening. The articles had to be written in English or German, published between 2008 and 2021, and evaluate mobile apps dealing with the prevention and handling of unintentional child injuries. The identified 5 studies were analyzed by 5 independent researchers using an inductive approach. Furthermore, the quality of the studies was assessed using the *Mixed Methods Appraisal Tool*.

Results: A total of 5 articles were included and assessed with regard to overall quality of theoretical and methodological foundations, assessed variables, the focal app's architecture, and the needs of the study participants. The overall study quality was moderate, although part of this classification was due to a lack of details reported in the studies. Each study examined 1 mobile app aimed at parents and other caregivers. Each study assessed at least 1 usability- or user experience-related variable, whereas the needs of the included study participants were detailed in only 20% (1/5) of the cases. However, none of the studies referred to theories such as the Technology Acceptance Model during the development of the apps.

Conclusions: The future development and evaluation of apps dealing with the prevention and handling of child injuries should combine insights into existing models on user experience and usability with established theories on mobile information behavior. This theory-based approach will increase the validity of such evaluation studies.

(Interact J Med Res 2023;12:e45258) doi:10.2196/45258

KEYWORDS
mobile health; mHealth; caregiver; parental; prevention; first aid; pediatric; review method; injuries; health app; needs; mobile phone

Introduction

Background

The risk of unintentional injuries in children has repeatedly been revealed as a relevant factor in medical treatment. The second wave of the German Health Interview and Examination Survey for Children and Adolescents (KiGGS Wave 2, 2014-2017) showed that, under the consideration of previous studies [1], unintentional injuries in children remain at a continuously high level, especially in times of changing conditions and associated shifts in risk. For example, at the start of the COVID-19 pandemic, physicians initially predicted a higher risk of unintentional poisoning incidents because of the increased use of cleaning agents and disinfectants [2], which was later reported in studies from several countries (Italy [3], Brazil [4], United States [5], Iran [6], Australia [7], and Morocco [8]). Shifting risks and associated consequences for the well-being of children clearly underline the importance of research and measures to prevent unintentional child injuries.

Unintentional injuries vary by age and encompass various categories, such as traffic injuries (eg, motor vehicle–, bicycle-, and pedestrian-related accidents), burns, fires, falls, drowning, suffocation, and poisonings [9,10]. Most of these injuries are preventable [11]. However, to know how such injuries could be prevented, this information must reach the respective target groups via appropriate communication measures, for example, via mobile apps. Apps may play a decisive role by providing tools, information, and practical guidance to prevent, mitigate, and handle the risk of unintentional injuries [12]. Thus, research on apps proves to be crucial to develop appropriate, informative, and user-friendly communication measures.

In the context of preventing unintentional injuries in children aged <7 years, Stehr et al [13] provided a detailed analysis of intervening communication strategies and their effectiveness. The authors emphasized that the chosen communication was more effective when a theory was applied in the study. Highlighting the potential of digital media, the research also showed that tailoring improves the effectiveness of digital health communication interventions. The researchers concluded that digital adaptation is an important aspect of disseminating knowledge and informing those who can take measures to prevent unintentional child injuries: “caregivers, i.e., parents and other guardians as well as childcare workers and health professionals working with children” [13].

Given these facts, the successful development of an app depends on several factors, which are described in the following section.

Theory- and Evidence-Based Digital Communication Through Mobile Apps

Mobile health (mHealth) technology such as mobile apps can be an effective communication tool to distribute health information [14,15]. Apps offer a number of advantages: they not only are an instrument of mass communication but also enable the provision of (tailored) information on, for example, preventive measures in 1 communication channel [16]. The information can be received anywhere, at any time, and in any situation. To be effective, the information must be evidence-based and evaluated [17]. The term *evidence-based* refers conventionally to evidence-based medicine being defined as “the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients” [18]. As Larson [17] noted, most of the existing mHealth apps lack evidence when it comes to the intervention itself. However, this evidence-based health information is necessary to be able to make evidence-informed decisions [19].

When making decisions based on information provided in an app, several aspects need to be considered: (1) the content of the information; (2) the presentation of the information (eg, design and layout); and (3) from the user’s point of view, how comprehensive and relevant the information is [19] or what the user’s needs are when searching for information (eg, length of text, features, and different modes of communication such as visualization). Thus, the term evidence-based refers not only to proven scientific findings in the respective medical context. It also includes scientific insights into which communication channel, which format, and which message work effectively for which target group [20], as well as data that allow for conclusions to be drawn on the structure and presentation of the app’s content.

In addition, communication strategies that apply a communication theory such as the Theory of Planned Behavior are more effective in producing certain outcomes (eg, knowledge improvement). However, and in relation to prevention, Stehr et al [13] showed that almost two-thirds of their analyzed communication strategies were not theory based. Health communication theories allow for the prediction and explanation of individual behaviors, such as the use of a particular means of communication and the acceptance related to it. Therefore, not only the content itself but also the way it is presented as well as the health information behavior of the target groups play a decisive role in the development and use of apps.

To date, little is known about the usability of mobile apps regarding child injury prevention [21,22] or their effectiveness (for mHealth in general, see the study by Chib and Lin [23], and for child injury prevention, see the studies by Cho et al [24] and Ning et al [25]). After a systematic review of the literature on unintentional injuries in people of different ages, Omaki et al [22] concluded that evaluation data on smartphone apps are lacking. Furthermore, they could not provide any insights into the role that theory plays in the development of these digital intervention tools [22]. Similar to the findings of Stehr et al [13], only a few of the included studies described the theoretical background. Omaki et al [22] suggested that a more detailed description of the theories is necessary to be able to make generalizable statements on the role that theory plays in changing safety behaviors through the use of an app.

Technology Adoption

Changing safety behaviors, disseminating knowledge, and increasing self-efficacy are all important outcome types when communication on prevention is researched. However, the precondition for the effectiveness of an app is its use. Thus, it is not only about adapting protective behavior but also about adapting media use behavior to expedite technology adoption. However, this depends on aspects such as the perceived ease of use or perceived usefulness of the app, its perceived trustworthiness, and perceived risks regarding data protection [26]. In addition, previous research shows that trust in the app publisher and content [25] as well as in the technology itself [27] increases the acceptance of an app.

There are few studies that have developed and evaluated mobile apps for child health promotion (eg, based on the Theory of Planned Behavior [28]) and systematic reviews on educational aspects of mobile injury prevention programs [29]. However, there is to date no systematic review on the theory-related evaluation of mHealth interventions focusing on child injury prevention and handling.

Objectives

This study identified research on mobile apps dealing with the prevention and handling of unintentional injuries in children and sought to gain insights into the different theoretical and methodological approaches of the identified studies. Furthermore, our systematic review tried to understand the different needs of various target groups of these mobile apps and, thus, took a user-centered design approach into account. These objectives are specified in our research questions (RQs): (1) Which apps for the prevention and handling (first aid) of unintentional injuries in children have been studied in scientific literature? Which app architecture characteristics and features can be described? What do they visualize? Who is their target audience? (RQ 1); (2) What theories and methods are used to evaluate these mobile apps? (RQ 2); and (3) Which needs do private caregivers, professional caregivers, and health professionals have when using apps dealing with the prevention and acute (pain) management of unintentional injuries in children aged <7 years? What constraints may prevent the use of the apps? (RQ 3).

Methods

Preparatory Steps

Before we carried out our systematic search, we conducted a cursory review and prereview mapping of articles dealing with mobile apps aiming at the prevention or acute (pain) management of unintentional injuries. This step seemed necessary not only to identify possible keywords for the definition of an appropriate search string but also to gain insights into the field of mobile app research. This prereview also helped specify our RQs and identify relevant databases and journals. Thus, our systematic literature review followed the steps comprehensively described by Xiao and Watson [30] and was conducted in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [31] and checklist (Multimedia Appendix 1). A research protocol was registered in the PROSPERO database (ID omitted for double-blind peer review).

Search Strategy

Regarding academic databases, we included the technical-oriented databases *IEEE Xplore* and *ACM Digital Library*; the multidisciplinary databases *Scopus* and *Web of Science*; and databases from the disciplinary fields of psychology (*PsycINFO*), communication sciences (*Communication Abstracts* and *Communication & Mass Media Complete*), and biomedical and life sciences (*PubMed*) in our search. Using a Boolean approach, the search string combined the terms “app” or “application” with different synonyms for “children” and several unintentional injury-related categories such as “drowning,” “poisoning,” or “accidents” derived from the definitions offered by Sleet [9] and Schnitzer [10] (RQ 1) and the results of our prereview mapping of relevant articles. The search string also encompassed several usability-related categories such as “ease of use” or “user experience” derived from our research interest (RQs 2 and 3). Research has shown that the needs and preferences of users are highly correlated with perceived usability [32], whereas usability-oriented app research might also indicate factors or constraints that limit or prevent the use of an app. Finally, as our search string (Multimedia Appendix 2) aimed at yielding relevant articles on the prevention and acute (pain) management of injuries, we decided to exclude research on “disease self-management” apps.

Furthermore, we decided to conduct a systematic scan of articles published in the *BMJ* journal *Injury Prevention*. The added value of complementary hand searching has repeatedly been marked in scientific discussion [33,34] as it reduces the possibility of missing relevant studies. The separate search for relevant articles published in *Injury Prevention* seemed to promise a useful addition to our search strategy, not only because of the journal’s specific focus on the prevention of unintentional injuries as well as its broad scope with regard to injuries in general but also because our prereview mapping yielded particularly interesting results from the aforementioned journal [22].

Inclusion Criteria

Several inclusion and exclusion criteria were defined to evaluate the identified papers for further consideration in the literature review (*eligibility assessment*). For inclusion, studies had to have been written in English or German, had to elaborate on the elements of the main RQs, and had to have been published between the years 2008 and 2021. We chose this specific time frame as digital apps became increasingly prevalent with the release of the first *iPhone* in 2007 and, more specifically, with the launch of the Apple App Store in 2008 [35]. Following the logic of the research process, the search in the databases could only include articles up to and including March 2021—this was the first point of data collection, whereas the hand search covered all articles up to and including the end of April 2021.

The population of interest was health professionals and private and professional caregivers of children aged <7 years. We chose this age because, first, unintentional injuries in the home and during recreational activities are most common at this age [36].

Second, infants, toddlers, preschoolers, and school beginners depend on the knowledge of caregivers as they need to learn which risks to take and which to avoid [37,38]. If there was no exact age specification for the children and no explicit specification of an age of >6 years, the studies were included. For an overview of all the inclusion criteria, see Multimedia Appendix 3.

Exclusion Criteria

Articles that did not meet the inclusion criteria were excluded. Furthermore, articles describing mobile apps that dealt with the handling and prevention of intentional child injuries (eg, abuse) or tracking and collecting individual health data as part of an ongoing treatment were excluded, as were studies that did not provide empirical data, merely focusing on technical or theoretical aspects. For an overview of all the exclusion criteria, see Multimedia Appendix 3.

Systematic Database Search

A total of 358 papers were retrieved from 8 academic databases. Duplicate removal reduced our sample to 229 papers that went through several assessment and selection stages. The selection process is presented in the PRISMA flowchart (Multimedia Appendix 4) as well as in the overview in Multimedia Appendix 2.

Given the number of identified papers, 5 raters were involved in the eligibility assessment. To ensure a coherent and consistent selection process, we validated our interrater reliability. We randomly selected 100 papers from our sample. The titles and abstracts of these papers were assessed by all 5 raters. The Fleiss κ, an adaptation of the Cohen κ for ≥3 raters [39,40], was used to analyze our interrater agreement. The Fleiss κ accounts for chance agreement between several coders and measures not only accuracy (eg, by aligning the coding with the codebook) but also precision (eg, ensuring that agreement between different coders is not due to chance alone [41]). Measurement is carried out based on a calculated score, whereby a score of 1 indicates perfect agreement and a score of 0 indicates agreement totally owing to chance. After assessing the randomly selected papers, we calculated a Fleiss κ value of *0.63*—implying substantial agreement [42]. In view of the number of raters, we judged the level of agreement to be acceptable to assess the titles and abstracts of the remaining articles by only 1 rater each.

The remaining 129 articles were equally distributed among the 5 researchers involved in the title and abstract check. Each article was assessed by 1 researcher. This procedure left us with 11 papers. To confirm eligibility, these papers were further assessed through a full-text screening by 2 raters each. This step particularly focused on the documented objectives as well as the target users of the mobile apps described in the respective articles. Finally, discrepancies or disagreements concerning a particular eligibility assessment were resolved through discussion and consensus-oriented decision-making by all researchers.

The systematic database search ultimately yielded a total of 4 articles that were considered relevant for our literature review.

Systematic Scan of Injury Prevention (BMJ Journal) Articles

A total of 899 papers were retrieved from volumes 14 to 27 of this journal. The articles were reviewed for eligibility by 2 researchers. Owing to the quantity of articles, they were assessed first based on the title and then on their abstracts. Again, critical cases were discussed within the team, and a final decision was reached by consensus. Initial title analysis led to the exclusion of 95.9% (862/899) of the papers. The following review of the remaining articles' abstracts further excluded 97% (36/37) of the papers. Full-text screening of the remaining paper finally resulted in the inclusion of 1 article.

Reference Tracking

The final step of our literature review included a (backward) snowball search based on the literature references of all 12 papers included in the full-text screening. Following Wohlin [43], snowballing is "particularly useful for extending a systematic literature study," offering options to find additional studies that were not detected in the searched databases. Thus, 330 articles were further reviewed for eligibility by 2 researchers. Title and abstract analysis led to the exclusion of 97.3% (321/330) of the articles, so that 9 articles remained for the subsequent full-text screening. However, none of these articles met the inclusion criteria. As previously, critical cases in this step were discussed within the team.

As a result of these 3 different search strategies, 5 articles were selected for our comprehensive literature review.

Assessing Quality

To critically appraise the quality of the selected studies, we used the *Mixed Methods Appraisal Tool* (*MMAT*), which was developed for the appraisal stage of complex systematic literature reviews [44]. The MMAT specifically allows for the quality assessment of different study designs in systematic mixed study reviews, enabling us to assess the quality of all 5 included studies at once. Conflicting assessments were discussed between 2 researchers involved in this systematic literature review and finally resolved through consensus. The final appraisal and evaluation of the included studies are explained in more detail in the *Results* section.

Extracting, Analyzing, and Synthesizing Data

All 5 articles were coded using the qualitative data analysis tool MAXQDA (VERBI Software GmbH). The code system (Multimedia Appendix 5) was developed using an inductive approach; based on the original articles, we identified dimensions, categories, and subcategories that described the apps as well as their evaluation process and the major findings. The entire article was coded, including graphics and information boxes. A total of 5 coders worked on the material—coded, discussed, and recoded—striving for coherence through an iterative process. The coding process involved 4 stages. First, 2 coders coded the same article on their own. Second, they discussed the marked overlaps and divergences. If an intercoder agreement was not achieved, the whole team reviewed and reflected on the relevant passage ("negotiated agreement" [45]). Third, the codes were refined if necessary. Fourth, coding was repeated. Owing to the small number of studies and their various differences regarding the applied methods and data analysis strategies, we did not carry out any statistical evaluations. In addition, the data were analyzed using Miro (RealtimeBoard Inc dba Miro), a web-based whiteboard that allowed us to group different codings collaboratively.

Results

Characteristics of the Included Studies

Table 1 provides an overview of the app and study characteristics of the included studies.

Table 1. Overview of the included studies.

Study	App characteristics				Study characteristics				
	Name	Objectives	Main focus	Target audience	Research questions or objectives	Consideration of previous research	Participants (sample size)	Data collection	Data analysis
Iskander et al [46]	Dental Trauma first aid mobile app	Deliver information on dental trauma	Handling	Parents	Comparison of the effectiveness and user preference of the app with a poster containing the same information and testing of the app's suitability to inform parents about dental trauma management	Cursory overview of literature on peer-reviewed apps	Caregivers (n=89): any adult accompanying a child to the participating dentist offices	Survey (conducted via a laptop or tablet)	Statistical techniques, for example, cross-tabulations and independent chi-square statistics
Jones et al [47]	GUS[a]	Support parents and carers in reducing unintentional injuries	Prevention	Parents and caregivers	Parents' assessment of the usability of the app (ease of use, usefulness, and esthetics), consideration of multiple perspectives on the app, and identification of strengths and weaknesses of the app as well as potential for improvement	Scoping literature review on unintentional childhood injuries and review of app stores, websites, and apps providing information on unintentional injuries	Mothers (n=15)	3 focus groups	Thematic analysis
Litovitz et al [48]	webPOISONCONTROL app	Increase access to accurate information on poisoning, reduce costs of poison centers and increase their efficiency, and harmonize triage	Handling	The public, parents, and employees of poison centers	Assessment of the feasibility and user acceptance of the app and safety and correctness of triage recommendations	Not mentioned	App users (n=1339 web-based feedback respondents and n=9256 consecutive, public webPOISONCONTROL cases) and a toxicologist (n=1)	Web-based survey, app use data, and reported cases	Statistical techniques and quality assurance review—manual verification of each case
Richmond et al [49]	Active & Safe Central	Provide web-based sport injury and prevention information	Prevention	The public, coaches, parents, athletes, injury researchers, and practitioners	Description of the development of the web-based tool, collection of feedback on the selection of information for the app, and user experience of the app (data visualization and interactivity)	Systematic literature review on sport and recreational activities and their injury risks and review of websites and apps providing information on sports injuries	End users (n=16)	2 focus groups, including app use data	Not mentioned
Roberts et al [50]	Make Safe Happen	Help parents and caregivers make their homes safer by identifying injury risks	Prevention	Parents and caregivers	Awareness of injury prevention and safety behavior at home, user experience of the app, and motivations for injury prevention and challenges in making one's home safer	Cursory overview of literature on the delivery of preventive information	Parents or legal guardians (n=40)	5 focus groups, web-based pretest and posttest survey, and app use data	Thematic analysis and statistical techniques, for example, chi-square test

[a]GUS: Growing up Safely.

Study Quality: MMAT Results

Of the 5 studies included, 2 (40%) exclusively applied qualitative methods such as focus groups, 2 (40%) used quantitative methods such as surveys, and the last study (n=1, 20%) used a mixed methods approach, as documented in the MMAT overview (Multimedia Appendix 6). In total, 60% (3/5) of the studies used usability constructs, such as perceived usefulness or functionality assessment, and let the participants test the app while conducting focus groups or gaining data via web-based surveys [48-50].

Critically appraising the methodological quality of the evaluation studies was challenged by a lack of reported details on data collection, data analysis, and the interpretation of the study results. The mixed methods study by Roberts et al [50] provided the most methodological details. On the basis of this paper, we were able to answer 14 of the 15 MMAT questions on methodological quality. In contrast, the studies by Iskander et al [46] and Richmond et al [49] were the most difficult to assess. In this case, 2 out of 5 questions from the MMAT could not be answered. Therefore, a valid assessment of the overall study quality is hardly possible, which is why we will only address a few specific issues in this section.

Most studies (4/5, 80%) used appropriate measures and methods to address their research interests; however, Litovitz et al [48] did not provide enough information for an assessment. No study used a representative sample of their target population, although they all recruited the study participants from the group of potential app users. Regarding data analysis, Iskander et al [46] and Richmond et al [49] provided insufficient information to assess the validity of the data analysis process and the interpretation of the data. However, as far as can be assessed, Jones et al [47], Litovitz et al [48], and Roberts et al [50] largely evaluated and interpreted the research data adequately.

Overall, the quality appraisal by means of the MMAT resulted in an average overall score of approximately 50%, indicating a rather moderate methodological quality of the included studies. However, the lack of reported details contributed to this low score to a considerable degree. This points to the difficulties related to the adequate assessment of a paper's methodological quality. The MMAT is based on the assumption that publications present all methodological information in an ideal-typical way; however, this is not always the case. Accordingly, the average overall score reported previously should be treated with caution as, in all cases, no definite assessment could be made on at least 1 item.

Answering the RQs

Mobile Apps Dealing With the Prevention and Handling of Unintentional Injuries in Children via Mobile Apps (RQ 1)

Each paper presented the evaluation of a single app. In total, 60% (3/5) of the apps focused on preventing unintentional injuries (*Grow up Safely*, *Active & Safe Central*, and *Make Safe Happen*), whereas the remaining 40% (2/5) offered advice on handling unintentional injuries through appropriate first aid measures (*Dental Trauma* first aid mobile app and *webPOISONCONTROL* app; Table 1).

App Architecture Characteristics and Features

In total, 60% (3/5) of the apps offer features that allow for tailoring the information to the respective user. In both *Grow up Safely* and *Make Safe Happen*, users can select the age of their child and then automatically receive the injury prevention information that is important for this phase of their child's life. The *webPOISONCONTROL* app works in a similar way by providing personalized recommendations on how to react in case of an intoxication based on the child's age and other personal information. To assist users in implementing preventive measures, *Make Safe Happen* and *webPOISONCONTROL* include reminders via calendar notifications or emails to encourage consistent safety behavior. In addition, *Make Safe Happen* provides checklists to assess injury hazards in one's home. A barcode scanner in the *webPOISONCONTROL* app can help identify toxic substance ingredients. *Active & Safe Central*, *Make Safe Happen*, and the *webPOISONCONTROL* app also provide additional information via external links. A download function has been integrated into *Active & Safe Central* allowing resources, evidence synthesis tables, and reports to be saved.

The paper on the *Dental Trauma* app by Iskander et al [46] only provided very rudimentary information on what prevention or handling information was offered and how it was provided by the app (Table 2).

Table 2. Overview of app features and visual elements.

	Features					Visual elements		
	Information provision	Tailored information	Tools for hazard identification	Reminders or follow-up	Data download	Photographs	Infographics	Data tables
Dental Trauma first aid mobile app	✓					✓		
Grow up Safely	✓	✓				✓		
webPOISONCONTROL app	✓	✓	✓	✓		✓		
Active & Safe Central	✓				✓	✓	✓	✓
Make Safe Happen	✓	✓	✓	✓		✓		

Visualization Elements

The use of visualizations was only explained to a very limited extent. Although Roberts et al [50], Iskander et al [46], Litovitz et al [48], and Jones et al [47] included screenshots of the apps in their papers, they did not provide any details on how the pictures in the app were chosen or what purpose they served. On the basis of the screenshots, *Dental Trauma* used photographs to differentiate between injuries, whereas *Grow up Safely* and *Make Safe Happen* used screenshots mainly to illustrate the app and its menu. The *webPOISONCONTROL* screenshots indicated minimal use of visualizations, with the only photo showing 2 pills of a particular drug. *Active & Safe Central* used more visualizations, mentioning infographics and data tables illustrating the individual injury risks associated with different sports and recreational activities (Table 2).

Target Audience

All apps were designed to be used by parents or caregivers (Table 1). Medical staff or physicians were only indicated as target users for *Active & Safe Central*, which also aims to inform the public, mentioning parents, coaches, athletes, and injury researchers. Roberts et al [50] and Iskander et al [46] collected demographic information such as age, level of education, and ethnicity of the app users. However, only Iskander et al [46] included this information in their analysis of individual preferences. They found that age did not have an effect on the frequency of internet use or on the preference for health information being delivered via posters or apps. The authors noted that this might be due to the small age range of the caregivers represented in the study. Differences were found in the interest in using an app to inform oneself about (dental) health information being higher for people with a lower level of education. In this case, a lower level of education was associated with a higher interest in information. The authors suggested that this might be due to a higher baseline health knowledge in people with higher education levels [46].

Theoretical and Methodological Foundations of the Evaluative Study Designs (RQ 2)

Theoretical Foundations and Evidence Research

Although there are a number of established approaches to researching the acceptability and use of an app (eg, the Technology Acceptance Model [51] and Unified Theory of Acceptance and Use of Technology model [52]), these were not used in any of the studies. Furthermore, no existing theories were applied in the development of the apps. However, 20% (1/5) of the papers mentioned *health literacy* as relevant to (visual) language [47] in the discussion part, whereas another research team used “an integrated knowledge translation approach” [49] but did not reference previous research on this approach or describe how they operationalized it. Therefore, 80% (4/5) of the studies did not follow a theory-based research approach, whereas the fifth paper did not provide sufficient information on how the chosen approach was translated into methodology.

Existing evidence-based knowledge on how best to communicate prevention or handling information to caregivers via mHealth apps (eg, simple language, visuals, and examples) was not systematically considered in any of the 5 included articles (Table 1). However, Richmond et al [49] and Jones et al [47] thematically reviewed the literature on unintentional injuries in children focusing on the type of injury and the percentage of accidents.

Assessed Variables and Methodological Foundations

Despite the lack of a common theoretical background, all studies (5/5, 100%) evaluated at least 1 variable related to usability or user experience, although this was evaluated in varying degrees of detail. Jones et al [47] and Richmond et al [49] also collected data on the evaluation of the apps’ look and appearance. User-related variables were also included in all studies (5/5, 100%). Table 3 provides an overview of the variables that were not only mentioned but also analyzed.

Table 3. Overview of variables.

Study	App-related variables					User-related variables		
	Usability or user experience	Look or appearance	Functionality	Interactivity or navigation	Issues or limitations	Awareness or knowledge	Attitudes or motivations	Demographics
Iskander et al [46]	✓					✓	✓	✓
Jones et al [47]	✓	✓				✓		
Litovitz et al [48]	✓		✓				✓	
Richmond et al [49]	✓	✓	✓	✓	✓		✓	
Roberts et al [50]	✓					✓	✓	

The variety of the assessed variables is also reflected in the different methods used in the studies, ranging from simple to rather complex designs. The most commonly used qualitative method was focus groups to evaluate the app either during or after development [47,49,50]. Jones et al [47] and Richmond et al [49] examined the interaction of users with the app, whereas Roberts et al [50] focused more on the participants’ attitudes. The most common quantitative method used in the analyzed studies was surveys. These were also used at different points in time within the study design—Iskander et al [46] only surveyed participants after they had used the app, whereas Roberts et al [50] used a pre- and posttest design to assess the extent to which the app influenced safety behavior. This study also included an analysis of the app usage data regarding frequency of use and retrieved information. In all studies (5/5, 100%), participants were recruited from the target group of app users, mainly parents or caregivers of young children.

Needs of Private Caregivers, Professional Caregivers, and Health Professionals and Constraints Preventing the Use of the App (RQ 3)

Overview

As mentioned previously, all studies (5/5, 100%) focused on private caregivers (eg, parents and legal guardians), although Litovitz et al [48] offered no detailed description of the study participants. No study explicitly involved professional caregivers or health professionals in the evaluation of the app except for Richmond et al [49], who included "injury researchers and practitioners" as well as end users [49], and Litovitz et al [48], who asked a toxicologist to analyze the safety of the triage recommendations. Jones et al [47] mentioned parents and "carers" but did not explain who was considered a carer and only included mothers in their sample.

Only 40% (2/5) of the papers offered results on the needs of the study participants. Jones et al [47] provided a detailed report on participants' needs based on several focus group discussions, whereas Litovitz et al [48] only briefly summarized the most liked or preferred features that were collected through a web-based survey via open comment fields. Both papers mentioned trust as a relevant factor. However, without further elaboration, Litovitz et al [48] indicated the need for trusted and accurate information sources, whereas Jones et al [47] found trust to be an influencing factor for app use, with participants not only demanding links to other reputable sources in the app but also feeling "more likely to use the app if it had been recommended by a trusted source such as a health visitor or midwife" [47]. Data security and privacy were only reported in the study by Litovitz et al [48].

Considering the contents of the app, the participants expressed their need for simple, easy-to-understand texts that offered more than trivial or "common sense" information [47]. This need for simplicity was also emphasized in the study by Litovitz et al [48] but not discussed further. Instead, Litovitz et al [48] listed several other needs—such as helpfulness, convenience, or a step-by-step approach with regard to their app—without further elucidating them. Visuals were generally evaluated as relevant in the study by Jones et al [47]; however, the participants expressed a dislike of images of wounds or other injuries, stating that they might distract users from the text or context.

Constraints of App Use

Jones et al [47] and Roberts et al [50] also identified several aspects that might prevent either app use or the implementation of new prevention behaviors. The main constraint for app use is a lack of awareness of the app (eg, as the users forget to access the app regularly after downloading it). Push notifications that promote regular engagement with the app were seen as an adequate tool to combat this constraint [47]. However, even if users engage with the app and receive relevant information through it, this is no guarantee that they will also follow the app's recommendations. This might be due to being overwhelmed by a large number of recommended actions, a lack of resources to implement new safety measures, or disagreements with other family members regarding their appropriateness [50].

Discussion

Principal Findings

This systematic review is the first to focus exclusively on mHealth interventions aimed at the prevention *and* handling of unintentional injuries in children aged <7 years as well as the methodology applied in these evaluation studies. The results show that a theory- and method-driven evaluation of app usability is rarely applied. This hampers the comparability of evaluations of different apps with regard to user preferences and target group-specific adaptation.

Previous reviews have assessed the effectiveness of preventive interventions regarding child injuries in general [11,13], examined the state of research in relation to low- and middle-income countries [53] or disadvantaged groups [54], or focused on technology-based interventions without specifically addressing private and professional caregivers [22]. Although unintentional injuries cause pain, disabilities, and death in children and, therefore, pose a major public health challenge [53,55,56], not many apps deal with the prevention or handling of unintentional child injuries. Accordingly, research on these apps is limited. We identified 5 papers that did analyze existing apps on this issue and evaluated them. However, the excluded conference abstracts [22,57] and study protocols [58,59] indicate that more research is being conducted. Cooray et al [57] proved this with their article published in 2021 on a behavioral theory-based app for parents to prevent falls in children. They emphasized that their methodological approach was the first of its kind, driven by the same rationale on which our review is based. That study shows that a theory-based, evidence-based, and user-centered approach to digital app development is a useful way to intentionally affect the behavior of the target group. However, as their article was published after our data collection period, it was not included in our review.

In addition, we identified studies on a theoretical framework for developing such an app; however, they did not include the development of the mobile app itself [25,60].

All 5 included studies explored factors of app usability; however, none of them drew on findings from evaluation research on mHealth technologies. In addition, the study designs focused on an exploratory or inductive approach instead of relying on existing knowledge that explains use or health information–seeking behavior. Even though data on individual variables were collected, these were considered independently of health behavior theories or models as a research framework.

This isolated examination of single variables reduces the validity of the studies. By drawing on proven theories and models, it would have been possible to systematically test central basic assumptions whose scope exceeds the specific study context. In contrast, without reference to existing theories and models, it is not clear whether the findings of a study represent an isolated phenomenon or whether they point to generalizable findings.

Moreover, even before the communicative content is developed, different theories can provide clues as to which variables, such as the users' attitudes, subjective norms, or perceived self-efficacy, influence their behavior. This knowledge can help develop more effective communication messages for different target groups.

As previous usability studies in this specific research field have not been linked to mobile information behavior models or constructs such as health literacy, the applied research on mHealth interventions with regard to unintentional injuries in children is rather rudimentary (for similar findings on mobile apps in general, see the study by Chib and Lin [23]).

An explanatory factor could be that most researchers on this topic work in medical fields and, therefore, might not be familiar with information behavior models from a communication science perspective. An interdisciplinary look at the development and evaluation of such apps could advance research in this area.

Another point concerns the publishing strategies for scientific publications. In the study by Richmond et al [49], we can see that another publication by this research team [60] describes in much more detail how their framework for possible intervention strategies was developed "emphasizing four types of research evidence [...]: (1) epidemiologic evidence describing the burden and cause of injury, (2) evidence concerning the effectiveness of interventions, (3) evidence on effective methods for implementing promising interventions at a population level, and (4) evidence and theory from the behavioral sciences" [60]. This shows that knowledge from different disciplines did indeed play a role in the development of the mobile apps. However, their 2015 paper had to be excluded as it did not meet our inclusion criteria given that there was no mobile app developed or evaluated.

However, the included 2019 paper only mentions this information in passing without explaining how the different types of evidence influenced the implementation and transformation of scientific knowledge on sports injuries and prevention. Our review would have needed both papers to provide a complete picture of the current state of the art (see the *Strengths and Limitations* section). Such publication strategies need to be considered for future reviews.

Although the papers we analyzed did not have much in common, two categories could be identified among the examined variables: (1) app-related variables describing, for example, ease of use, and (2) user-related variables such as knowledge or risk perception. However, these papers did not provide information on how these variables were operationalized in the surveys [46,48,50]. This is a general challenge when researchers have to describe their study in a very limited format. Therefore, the methodological section should become a more important chapter in journal articles so that the methods used to obtain the data can be reflected upon. Only then will it be possible to evaluate the quality of a study and check the possibilities of reproducing it. This not only helps promote good scientific practice but can also strengthen trust in science.

Current Gaps

Usability is not just about the presentation of knowledge but also deals with ergonomic design, user-friendly settings, and accessibility (eg, clear layout, icons and their comprehensibility, and understandable error messages [61]) as well as technical setting options. Although some of the studies (2/5, 40%) evaluated aspects such as the layout of the apps, icon selection, and comprehensibility, they did not use a comprehensive concept of software ergonomics [62].

Regarding visualizations, the studies showed somewhat conflicting results [49]. On the one hand, the data proved that images help users navigate and search within apps. In contrast, especially in the area of injury management, images of injuries can be stressful. This fine line between the positive and negative effects of using images in health apps needs to be handled with great care.

Health literacy was implicitly addressed in the work by Jones et al [47]. In this case, the study participants discussed whether the texts in the app were easy and quick to understand. Nevertheless, the content should still be interesting enough to hold the readers' attention. For app developers, it could be a challenge to address a wider range of health literacy (see the *Constraints of App Use* section).

Linked to this are the needs of the target users—although some apps (eg, *Make Safe Happen* [50]) distinguished very precisely between different groups of parents (eg, based on the age of the child), others had rather vague target groups [49]. Targeted communication is essential if the mHealth intervention intends to make the information relevant to the specific target groups—considering data on their knowledge, (risk) perceptions, and motivations [63]. Furthermore, contrary to our expectations, only private caregivers were examined in more detail. Even though health professionals were not addressed as target users—except in the study by Richmond et al [49]—their views and reflections might offer relevant insights, especially as they are an effective intermediary [13] in the prevention and handling of unintentional child injuries.

Interestingly enough, gender did not play an analytical role, even though research shows that especially mothers of young children, expectant mothers [64,65], and mothers with low health literacy [66] use mHealth communication.

At least one study was able to show that users pay attention to (1) sources that recommend the app or (2) whether the app quotes, mentions, or links to reputable sources (eg, well-known people such as midwives [47]). This means that it is advisable for app creators not only to obtain the information in the app from evidence-based or familiar sources and relevant institutions but also to clearly mark the sources in the app.

Constraints of App Use

We have already mentioned how difficult it is for app content developers to consider different health literacy levels and demands. This includes not only the step of accessing information but also of understanding, appraising, and applying it [67]. One way to give people access to information via mobile apps could be push notifications. In particular, injury prevention apps should be designed to remind the user to consult the app in situations of uncertainty. In this case, too, the needs of the target groups are decisive for the design of the app, be it the ergonomics or the content. The digital possibilities of mHealth interventions to support behavior change or informed decision-making are present [68], but so are barriers. The amount of information may be sufficient for one user but too much for another—even in the same target group. These aspects cannot be easily generalized for every app. It depends on the topic, the individual user, whether the amount of information is appropriate, and many other aspects. In short, theory-based evaluations are crucial. This also applies to the question of whether and how app use leads to changes in behavior or has an impact on prevention or first aid behavior [23]. None of the studies evaluated the effects of app use. This highlights the importance of continuous app evaluation to match the intended use with the actual app use.

Strengths and Limitations

One of the assumed strengths of the review later turned out to be a weakness as well: the scope of this review was quite narrow, which led to a very limited number of results.

However, this particular limitation may also indicate one of its strengths. The function of a literature review may lie not only in its identification, analysis, or synthesis of relevant research literature but also in its ability to identify desiderata or research gaps in a specific research area. Therefore, the study not only contributes to the investigation of existing studies' deficits regarding their generalizability but also links the results to extant theories or models. The literature review might also help uncover previously unconsidered aspects or variables that could benefit future studies in the field of mobile apps dealing with the prevention and handling of unintentional injuries in children. At the same time, a desiderata-oriented form of literature review might also help in identifying common deficits in publication and writing processes that so far have often been overlooked.

Concerning the search method of our systematic review, a further limitation should be noted. In its third pillar, the developed search string only comprises generalized categories of unintentional injuries, which were identified by means of a prereview mapping as well as by recourse to the definition of unintentional injuries offered by Sleet [9] and Schnitzer [10]. Thus, this operationalization of unintentional injuries only covers categories of unintentional injuries (such as burns or poisonings) but not the concrete or possible causes of such accidents (such as stoves and ovens or a specific chemical) as, otherwise, the search string would have had to define an all-encompassing, expansive list of possible causes covering every conceivable scenario (eg, poisoning through medications, alcohol, and hydrocarbons). Therefore, the developed search string cannot fully ensure that all the relevant papers were actually identified. In at least 1 known case [69], a relevant article was not identified because it did not refer to specific categories of unintentional injuries in the title, abstract, or keywords.

Owing to the number of articles, the titles and abstracts of 129 articles were checked by only 1 rater each. To keep potential bias to a minimum, the interrater reliability was examined beforehand by calculating the Fleiss κ. In case of uncertainty, the inclusion or exclusion of an article was discussed among the researchers.

As we excluded all papers that did not explicitly report an empirical study on the evaluation of an app, we were not able to include papers that solely discussed the theoretical frameworks of app development. However, when designing the review, we did not consider that researchers might publish the results of their work in several articles (eg, 1 focusing on theoretical [60] and 1 on empirical [49] aspects). This turned out to limit the validity of our findings, especially our conclusions, as we were not able to include all published information on a particular research project.

Conclusions and Directions for Future Research

The aim of this systematic review was to summarize the state of the art on the evaluation of mHealth apps in the context of the handling and prevention of unintentional injuries in children aged <7 years. The findings of this review highlight two objectives that play an important role in developing and evaluating these apps: (1) the use of tailored (visualized) information being a part of (2) knowledge transfer and transformation. Both of these objectives play a role when it comes to app- and user-related variables.

Information tailoring in mHealth contexts contains more than the message itself; the app allows for personalization features such as the age of the child that can make the provided content more useful for the user [47]. Future research should combine insights into user experience and usability with existing theories on mobile information behavior and (visual) literacy constructs [14]. Thus, use and health information-seeking behavior variables may deliver data on how to tailor information to the needs of different target groups, including professionals. As previous research shows, gender should also be considered as a relevant predictor of mHealth app use [70]. When all this is considered, the efficiency of an mHealth intervention can be increased.

On the basis of this review, it is concluded that the focus of evaluative usability studies on the prevention and handling of unintentional injuries in children should be shifted from technical developments and first-phase studies examining singular variables to collecting evaluation data derived from theories and models to raise the validity of the foundational premises measuring health behavior.

Acknowledgments

The authors would like to thank their colleague Alena Biegert for reviewing this paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 checklist.
[PDF File (Adobe PDF File), 760 KB - ijmr_v12i1e45258_app1.pdf]

Multimedia Appendix 2
Research questions and literature search process.
[PDF File (Adobe PDF File), 498 KB - ijmr_v12i1e45258_app2.pdf]

Multimedia Appendix 3
Inclusion and exclusion criteria.
[PDF File (Adobe PDF File), 496 KB - ijmr_v12i1e45258_app3.pdf]

Multimedia Appendix 4
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 flowchart.
[PDF File (Adobe PDF File), 389 KB - ijmr_v12i1e45258_app4.pdf]

Multimedia Appendix 5
Code system.
[PDF File (Adobe PDF File), 553 KB - ijmr_v12i1e45258_app5.pdf]

Multimedia Appendix 6
Results of the Mixed Methods Appraisal Tool.
[PDF File (Adobe PDF File), 757 KB - ijmr_v12i1e45258_app6.pdf]

References

1. Saß AC, Kuhnert R, Gutsche J. [Unintentional injuries in childhood and adolescence-prevalence, locations, and mechanisms: results from KiGGS Wave 2 and trends]. Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz 2019 Oct 16;62(10):1174-1183. [doi: 10.1007/s00103-019-03013-w] [Medline: 31529185]
2. Le Roux G, Sinno-Tellier S, French Poison Control Centre members, Descatha A. COVID-19: home poisoning throughout the containment period. Lancet Public Health 2020 Jun;5(6):e314 [FREE Full text] [doi: 10.1016/S2468-2667(20)30095-5] [Medline: 32339480]
3. Bressan S, Gallo E, Tirelli F, Gregori D, Da Dalt L. Lockdown: more domestic accidents than COVID-19 in children. Arch Dis Child 2021 Feb 02;106(2):e3. [doi: 10.1136/archdischild-2020-319547] [Medline: 32487724]
4. Ferrer IL, Baptistella MK, Oliveira FN, Souza AG, Cunha LC, Magalhães AF. Poisoning in children and adolescents assisted during the COVID-19 pandemic at the Toxicological Information and Assistance Center in Federal District, Brazil (CIATOX -DF): descriptive, cross-sectional, and analytical study with 1.037 patients. Res Soc Dev 2021 Nov 15;10(15):e25101521960. [doi: 10.33448/rsd-v10i15.21960]
5. Gielen AC, Bachman G, Badaki-Makun O, Johnson RM, McDonald E, Omaki E, et al. National survey of home injuries during the time of COVID-19: who is at risk? Inj Epidemiol 2020 Nov 11;7(1):63 [FREE Full text] [doi: 10.1186/s40621-020-00291-w] [Medline: 33176881]
6. Mahdavi SA, Kolahi AA, Akhgari M, Gheshlaghi F, Gholami N, Moshiri M, et al. COVID-19 pandemic and methanol poisoning outbreak in Iranian children and adolescents: a data linkage study. Alcohol Clin Exp Res 2021 Sep 06;45(9):1853-1863 [FREE Full text] [doi: 10.1111/acer.14680] [Medline: 34487368]
7. Palmer CS, Teague WJ. Childhood injury and injury prevention during COVID-19 lockdown - stay home, stay safe? Injury 2021 May;52(5):1105-1107 [FREE Full text] [doi: 10.1016/j.injury.2021.04.032] [Medline: 33962722]
8. Sahar AH, Hasnae H, Hajar M, Mohammed C, Sana C, Sanae A, et al. Children's poisoning profile during the Covid-19 pandemic – experience of Hassan II University Hospital in Fez, MOROCCO. In: Proceedings of the International Congress on Health Vigilance (VIGISAN 2021). 2021 Presented at: VIGISAN 2021; July 15-17, 2021; Kenitra, Morocco. [doi: 10.1051/e3sconf/202131901077]
9. Sleet DA. The global challenge of child injury prevention. Int J Environ Res Public Health 2018 Sep 04;15(9):1921 [FREE Full text] [doi: 10.3390/ijerph15091921] [Medline: 30181462]
10. Schnitzer PG. Prevention of unintentional childhood injuries. Am Fam Physician 2006 Dec 01;74(11):1864-1869 [FREE Full text] [Medline: 17168342]
11. Jullien S. Prevention of unintentional injuries in children under five years. BMC Pediatr 2021 Sep 08;21(Suppl 1):311 [FREE Full text] [doi: 10.1186/s12887-021-02517-2] [Medline: 34496772]
12. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]

13. Stehr P, Reifegerste D, Rossmann C, Caspar K, Schulze A, Lindemann AK. Effective communication with caregivers to prevent unintentional injuries in children under seven years. A systematic review. Patient Educ Couns 2022 Aug;105(8):2721-2730 [FREE Full text] [doi: 10.1016/j.pec.2022.04.015] [Medline: 35537900]
14. Kreps GL. The relevance of health literacy to mHealth. Stud Health Technol Inform 2017;240:347-355. [Medline: 28972527]
15. Nacinovich M. Defining mHealth. J Commun Healthc 2013 Jul 18;4(1):1-3. [doi: 10.1179/175380611x12950033990296]
16. Kreps GL. Online information and communication systems to enhance health outcomes through communication convergence. Hum Commun Res 2017 Jun 30;43(4):518-530. [doi: 10.1111/hcre.12117]
17. Larson RS. A path to better-quality mHealth apps. JMIR Mhealth Uhealth 2018 Jul 30;6(7):e10414 [FREE Full text] [doi: 10.2196/10414] [Medline: 30061091]
18. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996 Jan 13;312(7023):71-72 [FREE Full text] [doi: 10.1136/bmj.312.7023.71] [Medline: 8555924]
19. Hirschberg I, Seidel G, Strech D, Bastian H, Dierks ML. Evidence-based health information from the users' perspective – a qualitative analysis. BMC Health Serv Res 2013 Oct 10;13(1):405 [FREE Full text] [doi: 10.1186/1472-6963-13-405] [Medline: 24112403]
20. Stehr P, Heinemeier D, Rossmann C. Evidenzbasierte | evidenzinformierte Gesundheitskommunikation. Baden-Baden, Germany: Nomos; 2018.
21. Lupton D. Towards critical digital health studies: reflections on two decades of research in health and the way forward. Health (London) 2016 Jan 20;20(1):49-61. [doi: 10.1177/1363459315611940] [Medline: 26487686]
22. Omaki E, Rizzutti N, Shields W, Zhu J, McDonald E, Stevens MW, et al. A systematic review of technology-based interventions for unintentional injury prevention education and behaviour change. Inj Prev 2017 Apr 19;23(2):138-146. [doi: 10.1136/injuryprev-2015-041740] [Medline: 26787740]
23. Chib A, Lin SH. Theoretical advancements in mHealth: a systematic review of mobile apps. J Health Commun 2018 Nov 19;23(10-11):909-955. [doi: 10.1080/10810730.2018.1544676] [Medline: 30449261]
24. Cho YM, Lee S, Islam SM, Kim SY. Theories applied to m-health interventions for behavior change in low- and middle-income countries: a systematic review. Telemed J E Health 2018 Oct;24(10):727-741 [FREE Full text] [doi: 10.1089/tmj.2017.0249] [Medline: 29437546]
25. Ning P, Gao D, Cheng P, Schwebel DC, Wei X, Tan L, et al. Needs analysis for a parenting app to prevent unintentional injury in newborn babies and toddlers: focus group and survey study among Chinese caregivers. JMIR Mhealth Uhealth 2019 Apr 30;7(4):e11957 [FREE Full text] [doi: 10.2196/11957] [Medline: 31038465]
26. Schnall R, Higgins T, Brown W, Carballo-Dieguez A, Bakken S. Trust, perceived risk, perceived ease of use and perceived usefulness as factors related to mHealth technology use. Stud Health Technol Inform 2015;216:467-471 [FREE Full text] [Medline: 26262094]
27. Sowon K, Chigona W. Trust in mHealth: how do maternal health clients accept and use mHealth interventions? In: Proceedings of the Conference of the South African Institute of Computer Scientists and Information Technologists 2020. 2020 Presented at: SAICSIT '20; September 14-16, 2020; Cape Town, South Africa. [doi: 10.1145/3410886.3410895]
28. Nolen SL, Giblin-Scanlon LJ, Boyd LD, Rainchuso L. Development and testing of a smartphone application prototype for oral health promotion. J Dent Hyg 2018 Apr;92(2):6-14. [Medline: 29739842]
29. Carroll AL, Christian R, Palokas M. Mobile injury prevention programs for children: a scoping review protocol. JBI Evid Synth 2022 Jun 01;20(6):1601-1607. [doi: 10.11124/JBIES-21-00150] [Medline: 35249994]
30. Xiao Y, Watson M. Guidance on conducting a systematic literature review. J Plan Educ Res 2017 Aug 28;39(1):93-112. [doi: 10.1177/0739456x17723971]
31. Page M, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Syst Rev 2021 Mar 29;10(1):89 [FREE Full text] [doi: 10.1186/s13643-021-01626-4] [Medline: 33781348]
32. Lee S, Koubek RJ. Understanding user preferences based on usability and aesthetics before and after actual use. Interact Comput 2010 Nov;22(6):530-543. [doi: 10.1016/j.intcom.2010.05.002]
33. Armstrong R, Jackson N, Doyle J, Waters E, Howes F. It's in your hands: the value of handsearching in conducting systematic reviews of public health interventions. J Public Health (Oxf) 2005 Dec 01;27(4):388-391. [doi: 10.1093/pubmed/fdi056] [Medline: 16311247]
34. Richards D. Handsearching still a valuable element of the systematic review. Evid Based Dent 2008 Sep 24;9(3):85. [doi: 10.1038/sj.ebd.6400602] [Medline: 18927572]
35. iPhone App Store downloads top 10 million in first weekend. Apple. 2008 Jul 14. URL: https://www.apple.com/newsroom/2008/07/14iPhone-App-Store-Downloads-Top-10-Million-in-First-Weekend/#:~:text=CUPERTINO%2C%20California%E2%80%94July%2014%2C,its%20launch%20late%20last%20week [accessed 2020-03-11]
36. Ellsäßer G. Unfälle, Gewalt, selbst verletzung bei Kindern und Jugendlichen Ergebnisse der amtlichen Statistik zum Verletzungsgeschehen 2014. Statistisches Bundesamt. 2017. URL: https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Gesundheit/Gesundheitszustand-Relevantes-Verhalten/Publikationen/Downloads-Gesundheitszustand/unfaelle-gewalt-kinder-5230001149004.pdf?__blob=publicationFile [accessed 2021-05-12]

37. Brussoni M, Brunelle S, Pike I, Sandseter EB, Herrington S, Turner H, et al. Can child injury prevention include healthy risk promotion? Inj Prev 2015 Oct 22;21(5):344-347 [FREE Full text] [doi: 10.1136/injuryprev-2014-041241] [Medline: 25535208]
38. Ellsäßer G, Trost-Brinkhues G, Albrecht M. [Injury prevention in young children]. Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz 2014 Jun 27;57(6):681-686. [doi: 10.1007/s00103-014-1971-y] [Medline: 24863710]
39. Gisev N, Bell JS, Chen TF. Interrater agreement and interrater reliability: key concepts, approaches, and applications. Res Social Adm Pharm 2013 May;9(3):330-338. [doi: 10.1016/j.sapharm.2012.04.004] [Medline: 22695215]
40. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med 2012 Oct 15;22(3):276-282. [doi: 10.11613/bm.2012.031]
41. Belur J, Tompson L, Thornton A, Simon M. Interrater reliability in systematic review methodology: exploring variation in coder decision-making. Sociol Methods Res 2018 Sep 24;50(2):837-865. [doi: 10.1177/0049124118799372]
42. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [doi: 10.2307/2529310]
43. Wohlin C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering. 2014 May Presented at: EASE '14: 18th International Conference on Evaluation and Assessment in Software Engineering; May 13-14, 2014; London, UK. [doi: 10.1145/2601248.2601268]
44. Hong QN, Pluye P, Fàbregues S, Bartlett G, Boardman F, Cargo M, et al. Improving the content validity of the mixed methods appraisal tool: a modified e-Delphi study. J Clin Epidemiol 2019 Jul;111:49-59.e1 [FREE Full text] [doi: 10.1016/j.jclinepi.2019.03.008] [Medline: 30905698]
45. Campbell JL, Quincy C, Osserman J, Pedersen OK. Coding in-depth semistructured interviews: problems of unitization and intercoder reliability and agreement. Sociol Methods Res 2013 Aug 21;42(3):294-320. [doi: 10.1177/0049124113500475]
46. Iskander M, Lou J, Wells M, Scarbecz M. A poster and a mobile healthcare application as information tools for dental trauma management. Dent Traumatol 2016 Dec 03;32(6):457-463. [doi: 10.1111/edt.12278] [Medline: 27140068]
47. Jones F, Whitehouse A, Dopson A, Palaghias N, Aldiss S, Gibson F, et al. Reducing unintentional injuries in under fives: development and testing of a mobile phone app. Child Care Health Dev 2020 Mar 11;46(2):203-212. [doi: 10.1111/cch.12729] [Medline: 31782175]
48. Litovitz T, Benson BE, Smolinske S. webPOISONCONTROL: can poison control be automated? Am J Emerg Med 2016 Aug;34(8):1614-1619 [FREE Full text] [doi: 10.1016/j.ajem.2016.06.018] [Medline: 27321939]
49. Richmond SA, Black AM, Jacob J, Babul S, Pike I. 'Active & Safe Central': development of an online resource for the prevention of injury in sport and recreational activity. Inj Prev 2019 Dec 14;25(6):546-551. [doi: 10.1136/injuryprev-2019-043164] [Medline: 31088897]
50. Roberts KJ, McAdams RJ, Kristel OV, Szymanski AM, McKenzie LB. Qualitative and quantitative evaluation of the make safe happen app: mobile technology-based safety behavior change intervention for parents. JMIR Pediatr Parent 2019 Mar 14;2(1):e12022 [FREE Full text] [doi: 10.2196/12022] [Medline: 31518322]
51. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
52. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]
53. Tupetz A, Friedman K, Zhao D, Liao H, Isenburg MV, Keating EM, et al. Prevention of childhood unintentional injuries in low- and middle-income countries: a systematic review. PLoS One 2020 Dec 29;15(12):e0243464. [doi: 10.1371/journal.pone.0243464] [Medline: 33373371]
54. Möller H, Falster K, Ivers R, Jorm L. Inequalities in unintentional injuries between indigenous and non-indigenous children: a systematic review. Inj Prev 2015 Apr 28;21(e1):e144-e152. [doi: 10.1136/injuryprev-2013-041133] [Medline: 24871959]
55. Burns. World Health Organization. 2018 Mar 6. URL: https://www.who.int/news-room/fact-sheets/detail/burns [accessed 2022-01-18]
56. Drowning. World Health Organization. 2022 Jan 18. URL: https://www.who.int/news-room/fact-sheets/detail/drowning [accessed 2023-08-01]
57. Cooray N, Sun S, Adams S, Elkington J, Ho C, Keay L, et al. Developing behavioural theory grounded and user-centred mobile app to prevent infant falls. Injury Prevent 2021;27:A41. [doi: 10.1136/injuryprev-2021-safety.124]
58. McKenzie LB, Roberts KJ, Clark R, McAdams R, Abdel-Rasoul M, Klein EG, et al. A randomized controlled trial to evaluate the Make Safe Happen® app-a mobile technology-based safety behavior change intervention for increasing parents' safety knowledge and actions. Inj Epidemiol 2018 Mar 12;5(1):5 [FREE Full text] [doi: 10.1186/s40621-018-0133-3] [Medline: 29527644]
59. Omaki E, Shields WC, McDonald E, Aitken ME, Bishai D, Case J, et al. Evaluating a smartphone application to improve child passenger safety and fire safety knowledge and behaviour. Inj Prev 2017 Feb 05;23(1):58. [doi: 10.1136/injuryprev-2016-042161] [Medline: 27597399]

60. Chambers A, Richmond SA, Logan L, Macarthur C, Mustard CA. The development of a framework to integrate evidence into a national injury prevention strategy. J Public Health (Oxf) 2015 Dec 01;37(4):671-677. [doi: 10.1093/pubmed/fdu069] [Medline: 26607757]
61. Yong TS, Perialathan K, Ahmad M, Juatan N, Abdul Majid L, Johari MZ. Perceptions and acceptability of a smartphone app intervention (ChildSafe) in Malaysia: qualitative exploratory study. JMIR Pediatr Parent 2021 Jun 01;4(2):e24156 [FREE Full text] [doi: 10.2196/24156] [Medline: 34061039]
62. Bevan N, Carter J, Harker S. Iso 9241-11 revised: what have we learnt about usability since 1998? In: Kurosu M, editor. Human-Computer Interaction: Design and Evaluation. Cham, Switzerland: Springer; 2015.
63. Elling JM, De Vries H. Influence of animation- versus text-based delivery of a web-based computer-tailored smoking cessation intervention on user perceptions. Eur J Health Commun 2021 Jul 28;2(3):1-23. [doi: 10.47368/ejhc.2021.301]
64. Hiebert B, Hall J, Donelle L, Facca D, Jackson K, Stoyanovich E. "Let me know when I'm needed": exploring the gendered nature of digital technology use for health information seeking during the transition to parenting. Digit Health 2021 Oct 18;7:20552076211048638 [FREE Full text] [doi: 10.1177/20552076211048638] [Medline: 34691754]
65. Lupton D, Pedersen S. An Australian survey of women's use of pregnancy and parenting apps. Women Birth 2016 Aug;29(4):368-375. [doi: 10.1016/j.wombi.2016.01.008] [Medline: 26874938]
66. Manganello JA, Falisi AL, Roberts KJ, Smith KC, McKenzie LB. Pediatric injury information seeking for mothers with young children: the role of health literacy and ehealth literacy. J Commun Healthc 2016 Jun 16;9(3):223-231 [FREE Full text] [doi: 10.1080/17538068.2016.1192757] [Medline: 29051785]
67. Kaphingst KA, Kreuter MW, Casey C, Leme L, Thompson T, Cheng MR, et al. Health Literacy INDEX: development, reliability, and validity of a new tool for evaluating the health literacy demands of health information materials. J Health Commun 2012 Oct;17 Suppl 3(sup3):203-221. [doi: 10.1080/10810730.2012.712612] [Medline: 23030571]
68. Sartori F, Savi M, Talpini J. Tailoring mHealth apps on users to support behavior change interventions: conceptual and computational considerations. Appl Sci 2022 Apr 08;12(8):3782. [doi: 10.3390/app12083782]
69. Gielen AC, Bishai DM, Omaki E, Shields WC, McDonald EM, Rizzutti NC, et al. Results of an RCT in two pediatric emergency departments to evaluate the efficacy of an m-health educational app on car seat use. Am J Prev Med 2018 Jun;54(6):746-755. [doi: 10.1016/j.amepre.2018.01.042] [Medline: 29656914]
70. Bol N, Helberger N, Weert JC. Differences in mobile health app use: a source of new digital inequalities? Inform Soc 2018 Apr 26;34(3):183-193. [doi: 10.1080/01972243.2018.1438550]

Abbreviations

mHealth: mobile health
MMAT: Mixed Methods Appraisal Tool
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RQ: research question

Edited by T de Azevedo Cardoso; submitted 02.01.23; peer-reviewed by E Omaki, P Dabas, R Eckhoff; comments to author 29.05.23; revised version received 15.06.23; accepted 20.06.23; published 06.09.23.

Please cite as:
Schulze A, Lindemann AK, Brand F, Geppert J, Menning A, Stehr P, Reifegerste D, Rossmann C
Mobile Apps Aimed at Preventing and Handling Unintentional Injuries in Children Aged <7 Years: Systematic Review
Interact J Med Res 2023;12:e45258
URL: https://www.i-jmr.org/2023/1/e45258
doi:10.2196/45258
PMID:37672312

Review

Nonface-to-Face Visitation to Restrict Patient Visits for Infection Control: Integrative Review

Hyunwoo Jeong[1,2], RN, MSN; Yonsu Choi[3], RN, MSN; Heejung Kim[1,4], GNP-BC, RN, PhD

[1]College of Nursing, Yonsei University, Seoul, Republic of Korea

[2]Department of Internal Medicine Nursing, Seoul National University Hospital, Seoul, Republic of Korea

[3]Department of Surgical Nursing, Seoul National University Hospital, Seoul, Republic of Korea

[4]Mo-Im Kim Nursing Research Institute, Yonsei University, Seoul, Republic of Korea

Corresponding Author:
Heejung Kim, GNP-BC, RN, PhD
College of Nursing
Yonsei University
50-1 Yonsei-ro, Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 02 2228 3273
Fax: 82 02 2227 8303
Email: hkim80@yuhs.ac

Abstract

Background: In the COVID-19 pandemic, a visit restriction policy for patients has been implemented in medical institutions worldwide and visits are being made using alternative communication technologies. This shift has also required the use of platforms to prevent negative consequences of these restrictions.

Objective: The purpose of this review was to comprehensively explore nonface-to-face visits as an alternative during infection prevention and to synthesize the scientific evidence of their benefits and disadvantages.

Methods: A comprehensive search was conducted via the PubMed, Embase, CINAHL, Cochrane, and Web of Science electronic databases; unpublished trials in the clinical trials register ClinicalTrials.gov; and Virginia Henderson International Nursing Library up to September 10, 2021. The search query was developed according to the guidelines of the Peer Review of Electronic Search Strategies and included keywords on the topics of telemedicine and visitation restrictions. The inclusion criteria were a nonface-to-face modality using telemedicine with family in a hospital setting, experimental and observational studies, and articles written in English. The exclusion criteria were inaccessible in full text, not related to patient or family involvement, mainly focused on the study protocol, or only discussing the pros and cons of telemedicine.

Results: Overall, patients' families experienced emotional distress due to restrictions on face-to-face visits. Nonface-to-face virtual visits compensating for these restrictions had a positive effect on reducing the risk of infection to the patient and the family. This further encouraged psychological and physical recovery and decreased psychological distress. However, nonface-to-face virtual technology could not replace the existence of actual families, and technical problems with networks and devices are reported as limitations.

Conclusions: Ensuring the availability of technology and educating on the same in alignment with the characteristics of patients and their families, nonface-to-face virtual visits need to show more potential as an effective patient-centered treatment strategy based on more research and advanced practice.

(Interact J Med Res 2023;12:e43572) doi:10.2196/43572

KEYWORDS

nonface-to-face visitation; visit restriction; infection control; patient; family

Introduction

Visiting family members in the hospital provides a chance for interaction and emotional stability to patients. Previous studies reported that open and flexible family visitation prevents patients' delirium; reduces hospital days in the intensive care unit (ICU) [1,2]; and decreases anxiety, depression, loneliness, and distress levels [3,4]. In the case of newborns, parents have limited visits to the neonatal intensive care unit (NICU), which may increase health inequalities related to poor parental bond and postpartum depression [5,6]. Moreover, it is difficult for family members to receive family-centered care when visitation is restricted [7]. For example, restricted visitation inhibits communication with health care providers and can cause emotional distress for family members [8]. Thus, many studies have supported open patient visitation to meet family needs, prevent emotional distress, and improve the satisfaction of care [9,10].

However, an inverse policy has recently been implemented for family visitation. For example, Korea, which had the second largest number of confirmed Middle East respiratory syndrome (MERS) cases in the world after Saudi Arabia in 2015 [11], established strict regulations for infection control in medical institutions, including strong restrictions on family visitation at the hospital. The cultural customs of patient visitation and family caregiving are recognized as the main reasons why Korea initially failed to control the MERS outbreak [12]. Another example of such a policy shift relates to the COVID-19 pandemic. In particular, many countries have implemented administrative orders and quarantine guidelines to maintain social distancing and intercity travel restrictions for infection control. Medical institutions restricted people from visiting patients at hospitals to prevent the spread of COVID-19. According to a UK national survey [13], 117 (100%) hospitals reported that the family face-to-face visit policy in ICUs changed during the surge of COVID-19 cases; 19 (16%) hospitals reported no face-to-face family visits under any circumstances and 63% of hospitals indicated allowing family presence in certain circumstances such as at the end of life. Many hospitals still restrict family member visits to nonface-to-face visits [14]; consequently, several patients died without seeing their loved ones due to continuous social distancing and border closures [15,16].

Given a visit restriction policy implemented for patients in medical institutions worldwide, additional efforts are needed to prevent the negative consequences of these restrictions [6]. To reduce the negative impact, medical institutions have been using alternative communication technologies and platforms to conduct telephone calls and teleconferencing [5,6,8,17]. However, very few studies have explored whether such nonface-to-face visits have effects similar to those of face-to-face visits. Therefore, it is necessary to comprehensively review studies on nonface-to-face family visits through a systematic approach for generating evidence [18]. The purpose of this review was therefore to comprehensively explore nonface-to-face visits as an alternative during infection prevention and to synthesize the scientific evidence.

Methods

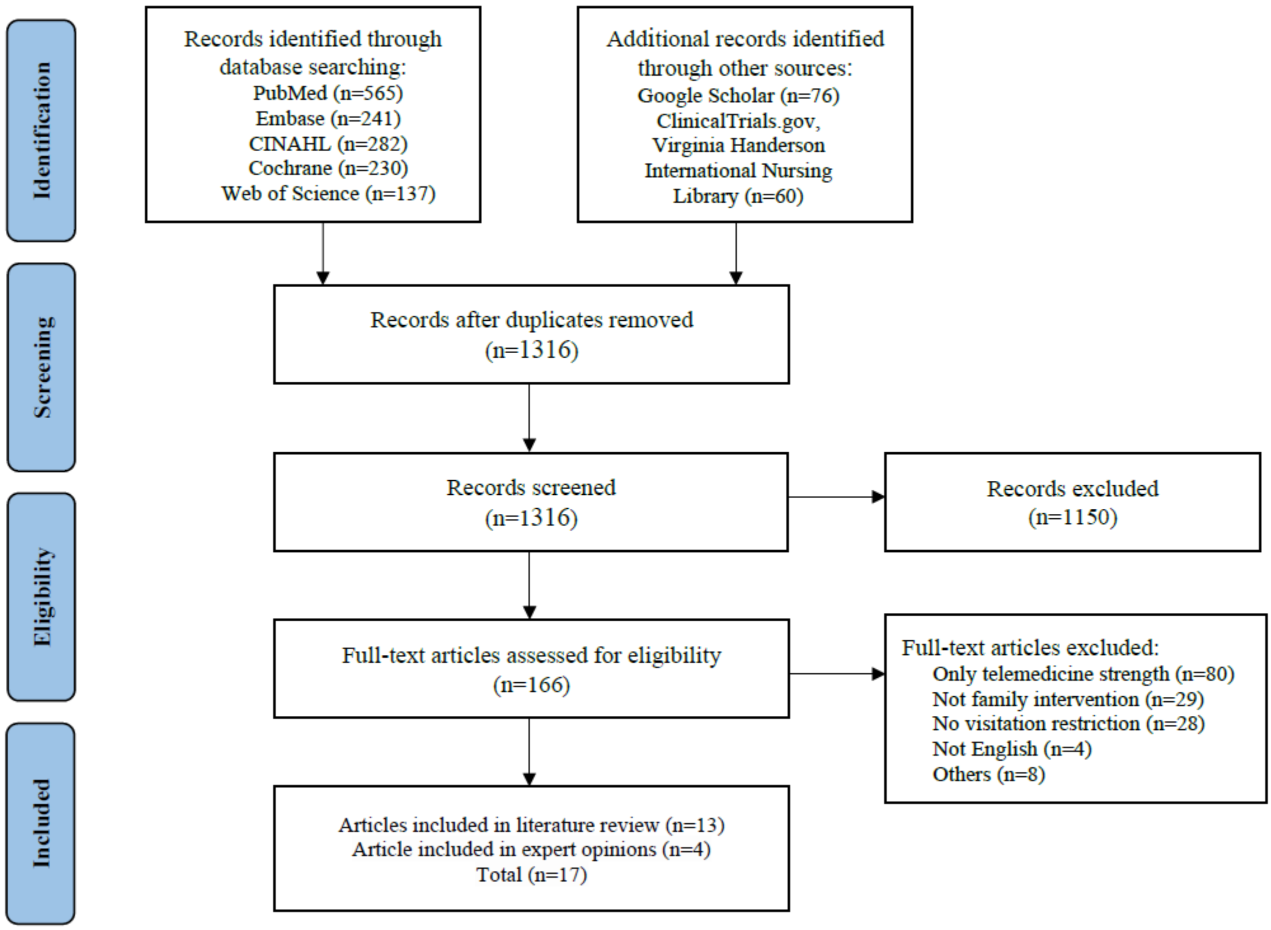
Search Strategy

A comprehensive search was conducted via the following five electronic databases up to September 10, 2021: PubMed, Embase, CINAHL, Cochrane, and Web of Science. In addition, we searched the clinical trials register ClinicalTrials.gov and Virginia Henderson International Nursing Library for unpublished trials up to the same date. The search query was developed according to the guidelines of the Peer Review of Electronic Search Strategies [19] and included keywords on the topics of telemedicine and visitation restrictions. The following search terms were used: telemedicine OR mobile health OR mhealth OR telehealth OR ehealth AND family AND visit. The search had no restrictions with respect to publication date or research design. Manual searches were performed via Google Scholar based on a reference list compiled from the articles retrieved from the above search strategy and databases for cross-referencing.

Eligibility Criteria

The inclusion criteria were as follows: (1) nonface-to-face modality using telemedicine with family in a hospital setting (ICU or non-ICU ward), (2) experimental and observational studies, and (3) articles written in English. Exclusion criteria were as follows: (1) studies without access to the full text, (2) studies not including patient or family involvement, and (3) the results included information on the protocol or only a discussion of pros and cons without associated data. Finally, we identified 17 studies as eligible according to the inclusion/exclusion criteria (see Figure 1).

Figure 1. Flowchart of the data selection process. Others: No access to full-text article or other formats, such as editorials or reports.



Selection Process

All three authors independently reviewed the retrieved studies throughout the selection process. The screening process was conducted by two authors (JH and CY) who independently extracted and cross-checked the literature using search queries. Data were extracted in the Covidence program (Melbourne, Australia), which is a web-based software that specializes in systematic reviews. Covidence allows researchers to import and screen citations and full-text articles, resolve conflicts, extract data using customizable forms, and export results in standardized formats. Using Covidence in the process of reviewing the study, the criteria for inclusion and exclusion can be continuously developed and shared with the research team. Moreover, random arrangement methods can be designated, including listing the articles in author order or in the most recent order, which can help to avoid systemic bias in reviewing studies.

The search results were exported from Endnote into Covidence for screening. Duplicates were then automatically identified and removed using the same software. Two authors (JH and CY) independently screened titles, abstracts, and the full text by applying the potential eligibility criteria using Covidence. Thus, the full text of 166 articles was reviewed and the results were discussed until there was acceptable interreliability between the reviewers (κ=0.9). The two authors fully reviewed the selected articles after developing definite eligibility criteria and showed over 95% agreement regarding the final selection of the articles. The third author (KH) served as the external validator when the analyses were conflicting and helped to reach the final agreement of the selection. At this stage, any ambiguous aspects were discussed until a consensus was reached.

Data Extraction and Analysis

The synthesis of evidence focused on the outcomes of nonface-to-face visits implemented during patient visit restrictions reported in each article. Because the review included various types of studies such as randomized controlled trials (RCTs), qualitative studies, and quasiexperimental studies, the following categories were used to conduct an integrative review: (1) characteristics of selected studies, (2) participants, (3) types of telemedicine, (4) benefits, and (5) limitations. Data extraction was independently performed by two authors (JH and CY) and any discrepancies were resolved through discussion.

Verification was conducted by comparing the results of the data analysis with the original articles. After data extraction and evaluation of study quality, summary tables were constructed regarding the study aims (see Multimedia Appendix 1). A detailed description of the data screening process is shown in Figure 1.

Risk of Bias and Quality Assessment

Among the 17 studies included for review, four were based on expert opinions and commentary; hence, quality assessment was not applicable to these studies. The remaining 13 studies were critically assessed for methodological quality using the Joanna Briggs Institute critical appraisal checklists [20] depending on study design such as RCTs, qualitative, cross-sectional, quasiexperimental, and cohort studies. The quality summary of each study was determined by integrating the contents. The quality of the 13 studies was scored according to 1 point for "yes" and 0 points for "no," "not applicable," and "not reported" for each item on the checklist. The quality of each study was assessed independently by two authors (JH and CY) and discrepancies were discussed and resolved by the third author (KH). A summary of the quality assessments for the 13 included studies is presented in Multimedia Appendix 2.

Results

Characteristics of Selected Studies

Among the 17 studies selected, there were five qualitative studies, six survey studies, two experimental studies, and four expert opinions. The majority of the studies were conducted in the United States (n=9) and published in 2020 and 2021 (n=15). In 13 studies, excluding the four studies based on expert opinions, the duration of the study period varied from 2 weeks to 2 years of follow-up. The number of study participants ranged between 20 and 367. Most studies included family members of patients in the ICU (n=10). Some studies (n=2) included families of patients in palliative care. Most studies were limited to parents or family members (11/13, 85%), and the remaining studies included some nurses or health care providers (2/13, 154%). A summary of the study objectives, study designs, study participants, and their characteristics is provided in Multimedia Appendix 1.

Content and Quality Assessment According to Study Type

Since various research methods were used in the included studies, the contents were analyzed according to the research design. The most common methods of data collection were interviews (n=5) and cross-sectional surveys (n=4). All five qualitative studies were conducted in the form of interviews, with research questions focused on describing clinical virtual pathways for visitation and communication [14] and exploring users' evaluation of telemedicine for patients or family members [15,16,21,22]. The six survey studies consisted of five cross-sectional studies and one prospective cohort study including patients and families belonging to a wide range of ages. The nonface-to-face virtual visits involved families of patients in the ICU [13,21-25]; pediatric patients, including newborns in the NICU [14,26-28]; and older adults in hospitals and long-term care facilities [29,30]. With the advent of the global COVID-19 pandemic in 2020, virtual visitation and nonface-to-face visitation were generally introduced, and most of these studies adopted a qualitative research and survey design.

There were two experimental studies in total: one RCT evaluating the role of online video visitations [24] and one study on changes in patients and their families through intervention with virtual visit programs [31]. The quality assessment revealed that the quality of cohort, RCT, and qualitative research studies was generally acceptable. Although the quality of one of the cross-sectional studies was poor, the overall quality of these studies was high. However, a study using a quasiexperimental research method [31] had a mid-level quality standard because it did not clearly report whether participants were included in similar comparisons, whether group differences were adequately explained in follow-up measures and measured in the same way, and the statistical analysis method used for measuring the results.

Types of Nonface-to-Face Visitation

Only 13 studies were reviewed in terms of the types of nonface-to-face visitation methods analyzed, excluding the four studies based on expert opinions. These 13 studies were related to new applications or programs and existing platforms such as Webcam, FaceTime (Apple), Zoom, Skype, Cisco Webex, and Microsoft Teams [13,15,27,31,32]. Since patients in the ICU who are generally sedated and intubated do not have independent access to these nonface-to-face telemedicine technologies, the medical staff is required to be at the patient's side to provide mobile devices; thus, a platform was developed to meet the needs of these patients, their families, and medical staff [22]. New virtual care platforms were developed targeting specific groups, including Chez NICU Home [14] for newborns, Sickbay [22,23] for patients in the ICU, and Family-Link for children and their parents [28]. These new programs or applications are summarized in Table 1. Some of the telemedicine platforms only involve online video visitation [21,24,27,29,30]. Other platforms were also featured in the included studies, such as WhatsApp [16], TouchAway [13], and HowRU [25].

Table 1. Characteristics of new programs or applications for nonface-to-face visits.

Platform	Purpose	Function or features	Target user
Chez NICU home	A secure web-based platform that provides a NICU[a] family with the training and resources needed to actively participate in baby care	Allows NICU patients and families, regardless of location, to connect with health care providers, families, and community health professionals and participate in customized interactive parental education and treatment	NICU patients and their families
Sickbay	Supports virtual care and remote monitoring workflows	Physiological data monitoring, including vitality sign monitors such as ventilators and virtual rounding, and mentoring of health care providers	Patients in the ICU[b]
TouchAway	Connects patients, clinicians, and families	Provides patient-centric care using virtual communication, treatment pathways, remote patient monitoring, care plan management, and many other tools	Clinicians, older adults, hospital patients, hospital administrators
HowRU	Open, flexible virtual visits tailored to patient-family–centric care	Ensures the privacy, dignity, and security of patients and families; facilitates an open and flexible communication line that adapts to the needs of each patient and family	Patients, families, and residents
Family-Link	Family child protection services that allow parents to adjust their children's devices	Various content restrictions, device usage time management, GPS phone search, educational app download	Parents, children, and adolescents

[a]NICU: neonatal intensive care unit.

[b]ICU: intensive care unit.

Benefits of Nonface-to-Face Visitation

The commonly reported benefits of nonface-to-face visits were promoting the psychological and physical recovery of patients [13,24,28-30] and reducing the psychological distress of family members by connecting them to their loved ones [13,16,22,23,25-27,31,32]. These nonface-to-face visits allow family members to meet their loved ones and make informed decisions about follow-up or provide end-of-life care before the patient's death [31]. Another advantage reported was increased collective interaction when patients and their families met over video calls, which could include a group of people at the same time rather than one-on-one phone calls, allowing them to experience more social group dynamics [32]. In addition, nonface-to-face visits can enhance communication between family and health care teams through these virtual technology platforms, allowing family members to participate in the patient's treatment [14,15,21-23]. In particular, video calls may be superior to phone calls to convey a general impression of the patient's condition [15,21,30].

Although no study directly confirmed infection control as a primary outcome, almost all of the included studies stated some beneficial aspects of infection control [13-16,21-25,28-33]. Restricting face-to-face visits can prevent the spread of the virus, protect vulnerable patients from infectious diseases, and reduce the potential impact of infections on organizational environments [13,14,16,21,22,25,28,29,31-33]. As needed, nurses in the hospital rooms were required to perform nonface-to-face visits during patient treatment–related tasks to reduce the exposure of infection [15,22]. A study in Iran conducted before COVID-19 [24] reported that face-to-face visits were restricted due to concerns such as infection risk, delayed patient rest, invasion of patient privacy, and obstruction of nursing care.

Nonface-to-face visits also affected health care providers. Nonface-to-face visits can reduce the burden on health care providers and improve employee work ethic [13,32]. However, an employee survey revealed that the physical presence of physicians in the virtual ICU decreased [32]. In addition, a new virtual treatment platform and guidelines for the standardization of treatment have been established to provide continuous education even after nonface-to-face visits to enhance patient-centered treatment [13,14]. Using existing communication methods such as phone calls and email can also further promote positive emotions [21,29].

It is also necessary to develop a new platform because medical staff use personal protective equipment (PPE) and are exposed to unnecessary risk of infection [22]. The nonface-to-face visits can reduce unnecessary exposure and the overuse of PPE by the medical staff when caring for patients with infectious diseases [15,16,22,23,25,31]. In addition, the established virtual ICU allows nurses to request a quick visual review from a doctor, who is not required to wear PPE outside the patient room [32]. These nonface-to-face visits also provide additional benefits to obtain a convenient consultation from other health care professionals such as extracorporeal membrane oxygenation specialists, cardiologists, and other specialists, without physical contact [23].

Limitations of Nonface-to-Face Visitation

Since face-to-face visitation restrictions come with barriers to effectively understanding and making decisions about the seriousness of the patient's disease, it is always necessary for health care providers to use statements that can empathize with the family's potential shock at seeing the potentially deteriorated state of the patient and warn the families accordingly before starting a nonface-to-face visit [15,32]. Most of the families who participated in nonface-to-face visitation shared positive emotions. However, some family members felt negative emotions, as they were sad to see the patient's critical condition; however, these should not be interpreted as negative emotions about the nonface-to-face visit technology specifically [22,32]. Owing to the unpredictability of bedside nursing, it is impossible to reserve a nonface-to-face visit in advance. Hence, it is necessary to explain to the families in advance that the reservation may be canceled or stopped if urgent patient treatment is required since the family may not always be familiar with these aspects [22].

Media-specific difficulties of nonface-to-face telemedicine technology were also reported in the reviewed studies. Common barriers to nonface-to-face visits include technical problems such as network connection problems [13,22], access to appropriate devices [13,22,29,30,32], lack of staff time [13,22], potential for increased workload [13,14], and insufficient visiting time [31]. There are also problems associated with the effective use of nonverbal communication, including silence, limited facial expression, and difficulties in important discussions [21,22]. Specific concerns were raised about the lack of physical contact for patients in the pediatric ward who are separated from their parents [26-28]; however, no studies evaluated the effect of a lack of physical contact, and only one study reported a greater reduction in stress for children and parents who used nonface-to-face visits compared to those who did not [28].

Application of the latest technology is emphasized to protect medical staff and patients' families from being in contact with infected patients as a part of patient-centered care in the pandemic era [23]. Although Facetime and Zoom can be used in everyday environments, patients in the ICU who are intubated and sedated cannot independently access these systems. Thus, medical staff must provide and operate mobile equipment to connect these patients to their families. If the patients do not have direct access to a platform for communication, there are privacy concerns about the sharing of personal information in the process of seeking external help [25]. Privacy protection may be difficult because of unintentional exposure of the medical staff to personal conversations between patients and their families in the ICU [32].

Discussion

Principal Results

This integrative literature review provides a timely understanding of virtual nonface-to-face visits for inpatient care. Our study summarizes the advantages and disadvantages of nonface-to-face virtual visits when face-to-face visits of families are impossible due to the risk of infection. The key findings from the included studies highlight the emotional distress experienced by patients and their families, such as isolation and loneliness, due to restrictions on face-to-face visits. Virtual nonface-to-face visits have been conducted using a variety of systems and platforms, with effects on patients, their families, and health care providers. Considering the characteristics of users and available technology, virtual nonface-to-face visits have become an important communication alternative with both advantages and disadvantages.

Comparison With Prior Work

Since the outbreak of the COVID-19 pandemic, a nonface-to-face approach has been adopted worldwide in many areas such as education, health care, and business. Previous studies have reported that nonface-to-face family visits protect health care providers and reduce PPE use while providing treatment [34]. Similar to the 2009 H1N1 influenza pandemic and the 2014 MERS epidemic, hospital transmission should be reduced through the use of protective equipment for visitors and staff, hand hygiene, and proper precautions [35,36]. Restricting patient visits has become an axiom-based public health policy to maximize the benefits for the community [5]. Strict isolation and visitation restrictions protect vulnerable patients from infection and reduce the risk of infection to their families and health care providers who care for infected patients [37,38]. Specifically, infection of health care providers causes a workforce shortage, increasing the burden on remaining health care providers in a pandemic [39].

Nonface-to-face visits have been proposed as an alternative to support patients and their families, most significantly since the outbreak of the COVID-19 pandemic with advances in associated technology. Virtual visitation aids in meeting the patient and communicating with the medical team without physical contact. To reduce the negative consequences of patient isolation, access to various social technologies has been widely proposed and the use of telemedicine services has increased [21,30]. However, it has been difficult to cope with the unexpected demise of loved ones [15] and some report a struggle with unfamiliar communication methods [16,29]. For example, the studies on video technology–based interventions such as FaceTime and Skype have shown parental appreciation for being able to see their baby when the NICU is inaccessible [40]. However, a few parents felt guilty for not being able to stay with their children when they met them virtually [41]. Similarly, families of severely ill patients felt negative emotions such as sadness when they observed the serious situation on video [15,32]. In addition, the type of support or demand received after the video session is unknown [15], and families indicated that it is confusing to ask medical providers about the health status or treatment process of the patient due to concerns related to the unpredictable nature of a disease course [22]. Families should be able to visit patients whenever they wish; however, the importance of communication is emphasized over visits because of the priority accorded to patients [42].

There are several potential ways to expand communication with the medical team that continuously informs family members about the patient's current state or to help with emotional conversations and provide support to the patients. Health care providers are also grateful for being able to emotionally help patients in difficult-to-face situations and acknowledged that these were crucial interactions [43]. In a situation where face-to-face visits were completely banned, ICU medical teams were encouraged to form a bond with the patients' families and act as mediators between the patients and their families [44]. However, there was a difference in perspective between the family and the medical team: while families wanted to communicate with medical staff more regularly and frequently, clinicians were responsible for managing nonface-to-face virtual visits, which was perceived as a heavy emotional burden [21,45].

Furthermore, prohibiting or restricting visitors has raised the ethical issues of exercising the right to freedom and not being able to see family members. Some studies expressed concern about the impact that such restrictions have on the bond between pediatric patients and their families, suggesting that the potential risk of infection should be weighed against the adverse effects of visit restrictions [46]. There is a conflict between the demands of the patient's family when they want to see the patient and the medical staff who wish to manage the patient's treatment. Finally, the health care provider who is close to a patient can inflict moral damage by invading their privacy or eavesdropping on the intimate and emotional conversations between the patient and family in the process of helping with nonface-to-face visits [32].

Implications for Research

This review provides a comprehensive understanding of the effects of nonface-to-face virtual visits for patients, their families, and medical staff; however, only one study was conducted based on a rigorous research design such as an RCT. Thus, more studies are required to examine the effectiveness of nonface-to-face virtual visits on patient outcomes and care team performance in diverse care settings with larger samples. In this review, the data were synthesized and presented descriptively along with context from other studies, including articles based on expert opinions. This is a methodological limitation, and further research will require more rigorous methodologies such as using probability sampling, controlling for confounding variables, and focusing on a narrow range of subjects such as patients in critical conditions and the NICU to increase the likelihood of the generalization of results.

Most studies included in this literature review were conducted in developed countries such as the United States, the United Kingdom, and Australia. This delineates the limitation of differences in accessibility to virtual visits and telemedicine in different countries. Since nonface-to-face and virtual visits are not limited to geographic areas, it is considered necessary to apply various research methodologies and to conduct more follow-up research on various patient groups and families by country and culture. A wide range of large-scale studies is needed, including various countries and institutions and patients of varied ages and with different diseases. Different countries have varied accessibility to virtual nonface-to-face systems, which may lead to a cultural difference in their understanding of visits; thus, our results may not be appropriate for generalization.

Experienced multidisciplinary medical staff can strengthen positive attitudes toward patients and their families to overcome the shortcomings of nonface-to-face virtual visits. This offers the potential for integrated development of a research-education-theory pathway, as the initial idea of research has begun in practice. It is further necessary to develop appropriate programs with the help of user-friendly technologies for nonface-to-face virtual visits. To supplement these virtual visits, ensuring the availability of technology is necessary. Future recommendations are proposed to expand the number of electronic devices and employees, as well as to simplify communication technologies to improve programs and platforms and extend allocated nonface-to-face visits [30,31].

A theory-based study is needed to explain health-related behaviors using nonface-to-face virtual smartphone apps or digital electronic devices. In this review, the complexity of communication technology and difficulty in accessing devices were reported as disadvantages of nonface-to-face visits; thus, efforts to simplify them are needed. As it is more difficult for older people to use new technologies and devices than young people, a systematic theoretical framework is needed to consider digital literacy and develop reliable and effective tools so as not to distort the actual meaning of the results.

We suggest measures to ensure the sustainability of these digital solutions. There is a need to expand facilities that allow digital solutions such as teleconferencing, telemedicine, and nonface-to-face visits to be implemented without direct contact. In addition, training on digital technology and equipment is required, and systematic education is also needed for patients, caregivers, and families. To continue using these digital solutions, privacy regulations and policies for personal information protection should be established so that the technology can be used without invading personal privacy. In fact, the biggest critique is that these digital solutions can become "fun and expensive toys." Therefore, continuous verification is required to ensure that the platforms are truly therapeutic and to optimize by whom, when, how, and how long they should be used to achieve an appropriate treatment effect; only then can a clinical protocol be presented. This can be followed by an economic evaluation such as the cost of the interface or decrease in the prescription of anxiolytics.

Implications for Education and Practice

The family's role as a gatekeeper is important for nonface-to-face virtual visits [25]. It is important to find a family-centered approach, guide the family to maximize virtual visits, and improve their understanding of the family experience through telephone and video communication [13,21]. The family can have direct access to the patient, improve the technical function of the system to use technology that simulates the family's face-to-face visit experience, improve the visiting process, and allow for more frequent communication between families and health care providers [21,22]. In particular, proficient use of platforms or mobile apps has a great influence on the results of interventions using virtual and telehealth technologies. To overcome the shortcomings of nonface-to-face virtual visits and enhance their strengths, educational opportunities to learn about multimedia devices, technologies, and mobile phones should be provided in hospitals. Education on these technologies is indispensable and can be further expanded to experience in other fields in the future.

Health care providers expressed that remote nonface-to-face visits were not a completely new communication strategy but rather a modification of existing communication strategies [21]. A study related to communication types and emotional experiences emphasized the synchronicity of communication, reporting that more frequent calls were associated with less negative emotional experiences and more positive perceived experiences [30]. In addition, a high level of satisfaction was associated with both video and telephone calls and the satisfaction level of video calls increased even more with the help of technology. Thus, it is necessary to select a communication type that suits the characteristics of the participants. Only 30% of medical institutions provide employee education for family communication and virtual visits [13]. Since virtual visits are widespread during the pandemic, it is necessary to expand employee training.

Some of the extracted literature included discussion of efforts toward protecting the private data of the patients and families during virtual visits [13,21,23,27,32]. Specific methods have been implemented by controlling the spread of webcam passwords [27], using “clean” iPads [32], or using proprietary software limited by concerns about the security of the system [13,32]. Other activities to secure privacy include (1) one-way calls only initiated by the medical team, (2) using secure cloud-based storage, (3) setting up two-step authentication for virtual visits, and (4) avoiding the use of personal devices by the medical team [13]. In addition, researchers also emphasized the expansion of video communication software [32], administrative and institutional support [21], and customization to comply with the Health Insurance Portability and Accountability Act [21,23]. Thus, the approach to security should take into account ease of use so that it works for users [47]. Therefore, medical team practitioners should value data security and software developers need to develop new technologies for user-friendly security systems.

Strengths and Limitations

The advantages of this review were that each process of the research—keyword identification, extracted studies, quality assessment, and analysis—ensured independence and consistency while using the Covidence program, which helped to standardized the review. Moreover, this study is closely related to practices utilizing research team clinical experience in the ICU and other special units in the hospital.

There are also several limitations to this study. First, there are few experimental studies to ensure the effectiveness of virtual visitation on diverse patient outcomes. Second, many studies on nonface-to-face and virtual interventions, telemedicine, and smartphone app development have been reported in a short period due to visit restrictions and social distancing recommendations in the COVID-19 era. As research results related to nonface-to-face medicine, telemedicine, and digital therapeutics are reported explosively during this period, many studies that meet the criteria for this review but were reported after the search period may have been excluded. Third, there were many potentially eligible studies in the grey literature, some of which were reported only as conference abstracts in posters. In addition, it is impossible to measure effect size or perform statistical pooling in qualitative studies at this point and there is a lack of information to conduct a systemic review and meta-analysis.

Conclusions

Despite its limitations, this study provides important information about patients, their families, and medical staff for nonface-to-face virtual visits. Face-to-face visit limitations caused emotional distress for the families of hospitalized patients and nonface-to-face virtual visits that made up for these restrictions helped to lower the risk of infection for the patient and family. Virtual visits also facilitated the interaction between patients and their families and helped families participate in the patient’s care by communicating with medical staff. This promoted recovery on both psychological and physical levels, while decreasing psychological distress. However, technical issues with networks and devices are reported as limitations and nonface-to-face virtual technology could not replace the actual presence of families. Ensuring the availability of technology and educating on the same in alignment with the characteristics of patients and their families, nonface-to-face virtual visits need to show more potential as an effective patient-centered treatment strategy based on more research and advanced practice.

Acknowledgments

This research was supported by the Professor-Student Academic Research Project through the Yonsei University College of Nursing (grant 6-2021-0052). This work was also supported by the Interdisciplinary Joint Research Fund from Yonsei University College of Nursing and College of Engineering (grant 6-2023-0165). The funders had no role in the study design, literature searches, data collection, analysis, or decision to submit the paper for publication. We did not utilize ChatGPT or other generative language models for this study or manuscript preparation.

Data Availability

All data analyzed during this study are included in Multimedia Appendix 1.

Authors' Contributions

JH designed the study, searched and selected the articles, performed the quality assessment, analyzed and interpreted the articles, and drafted and revised the manuscript. CH searched and selected the articles, performed quality assessments, analyzed and interpreted the articles, and drafted the manuscript. KH designed the study, collected the articles, analyzed and interpreted the articles, and revised the manuscript for important intellectual content. All authors made substantial contributions to the study and approved the submitted version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary of characteristics of articles and methodological approaches from 17 studies.
[DOCX File , 27 KB - ijmr_v12i1e43572_app1.docx]

Multimedia Appendix 2
Methodological evaluation of study quality.
[DOCX File , 32 KB - ijmr_v12i1e43572_app2.docx]

Multimedia Appendix 3
PRISMA checklist.
[PDF File (Adobe PDF File), 130 KB - ijmr_v12i1e43572_app3.pdf]

References

1. Rosa RG, Tonietto TF, da Silva DB, Gutierres FA, Ascoli AM, Madeira LC, ICU Visits Study Group Investigators. Effectiveness and safety of an extended ICU visitation model for delirium prevention: a before and after study. Crit Care Med 2017 Oct;45(10):1660-1667. [doi: 10.1097/CCM.0000000000002588] [Medline: 28671901]
2. Schwanda M, Gruber R. Extended visitation policy may lower risk for delirium in the intensive care unit. Evid Based Nurs 2018 Jul;21(3):80 [FREE Full text] [doi: 10.1136/eb-2018-102884] [Medline: 29592860]
3. Rosa R, Sganzerla D, Falavigna M, Kochhann R, da Silva DB, Teixeira C. Satisfaction and involvement in patient care influence the interaction between flexible ICU visiting hours and symptoms of anxiety and depression among family members of critically ill patients: a sub-analysis of a cluster-randomized crossover trial. 2019 Presented at: American Thoracic Society International Conference: Enhancing Patient-Centered Care in Critical Care; May 17-22, 2019; Dallas, TX p. A5570. [doi: 10.1164/ajrccm-conference.2019.199.1_MeetingAbstracts.A5570]
4. Zeh RD, Santry HP, Monsour C, Sumski AA, Bridges JF, Tsung A, et al. Impact of visitor restriction rules on the postoperative experience of COVID-19 negative patients undergoing surgery. Surgery 2020 Nov;168(5):770-776 [FREE Full text] [doi: 10.1016/j.surg.2020.08.010] [Medline: 32943203]
5. Murray PD, Swanson JR. Visitation restrictions: is it right and how do we support families in the NICU during COVID-19? J Perinatol 2020 Oct;40(10):1576-1581 [FREE Full text] [doi: 10.1038/s41372-020-00781-1] [Medline: 32772051]
6. Raphael JL, Kessel W, Patel M. Unintended consequences of restrictive visitation policies during the COVID-19 pandemic: implications for hospitalized children. Pediatr Res 2021 May;89(6):1333-1335 [FREE Full text] [doi: 10.1038/s41390-021-01439-0] [Medline: 33654277]
7. Goldschmidt K, Mele C. Disruption of patient and family centered care through the COVID-19 pandemic. J Pediatr Nurs 2021;58:102-103 [FREE Full text] [doi: 10.1016/j.pedn.2021.03.001] [Medline: 33781631]
8. Valley TS, Schutz A, Nagle MT, Miles LJ, Lipman K, Ketcham SW, et al. Changes to visitation policies and communication practices in Michigan ICUs during the COVID-19 pandemic. Am J Respir Crit Care Med 2020 Sep 15;202(6):883-885 [FREE Full text] [doi: 10.1164/rccm.202005-1706LE] [Medline: 32687720]
9. Nassar Junior AP, Besen BAMP, Robinson CC, Falavigna M, Teixeira C, Rosa RG. Flexible versus restrictive visiting policies in ICUs: a systematic review and meta-analysis. Crit Care Med 2018 Jul;46(7):1175-1180. [doi: 10.1097/CCM.0000000000003155] [Medline: 29642108]
10. Dan SY, Park SH, Lee S, Park HY, Yi YH. Experience of patients and families about flexible visiting. J Korean Crit Care Nurs 2017;10(1):51-62 [FREE Full text]
11. Cowling BJ, Park M, Fang VJ, Wu P, Leung GM, Wu JT. Preliminary epidemiological assessment of MERS-CoV outbreak in South Korea, May to June 2015. Euro Surveill 2015 Jun 25;20(25):1-7 [FREE Full text] [doi: 10.2807/1560-7917.es2015.20.25.21163] [Medline: 26132767]
12. Jack A. Why the panic? South Korea's MERS response questioned. BMJ 2015 Jun 24;350(Jun24):h3403. [doi: 10.1136/bmj.h3403] [Medline: 26108610]
13. Rose L, Yu L, Casey J, Cook A, Metaxa V, Pattison N, et al. Communication and virtual visiting for families of patients in intensive care during the COVID-19 pandemic: A UK National Survey. Ann Am Thorac Soc 2021 Oct;18(10):1685-1692 [FREE Full text] [doi: 10.1513/AnnalsATS.202012-1500OC] [Medline: 33617747]
14. Campbell-Yeo M, Dol J, Richardson B, McCulloch H, Hundert A, Foye S, et al. A co-design of clinical virtual care pathways to engage and support families requiring neonatal intensive care in response to the COVID-19 pandemic (COVES study). J Neonatal Nurs 2021 Dec;27(6):463-470 [FREE Full text] [doi: 10.1016/j.jnn.2021.06.010] [Medline: 34220279]
15. Kuntz JG, Kavalieratos D, Esper GJ, Ogbu N, Mitchell J, Ellis CM, et al. Feasibility and acceptability of inpatient palliative care e-family meetings during COVID-19 pandemic. J Pain Symptom Manage 2020 Sep;60(3):e28-e32 [FREE Full text] [doi: 10.1016/j.jpainsymman.2020.06.001] [Medline: 32505643]

16. Mercadante S, Adile C, Ferrera P, Giuliana F, Terruso L, Piccione T. Palliative care in the time of COVID-19. J Pain Symptom Manage 2020 Aug;60(2):e79-e80 [FREE Full text] [doi: 10.1016/j.jpainsymman.2020.04.025] [Medline: 32376263]
17. Pfefferbaum B, North CS. Mental health and the Covid-19 pandemic. N Engl J Med 2020 Aug 06;383(6):510-512. [doi: 10.1056/NEJMp2008017] [Medline: 32283003]
18. Rosenthal R, DiMatteo MR. Meta-analysis: recent developments in quantitative methods for literature reviews. Annu Rev Psychol 2001;52:59-82. [doi: 10.1146/annurev.psych.52.1.59] [Medline: 11148299]
19. McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. J Clin Epidemiol 2016 Jul;75:40-46 [FREE Full text] [doi: 10.1016/j.jclinepi.2016.01.021] [Medline: 27005575]
20. Critical Appraisal Tools. Joanna Briggs Institute. 2021. URL: https://jbi.global/critical-appraisal-tools [accessed 2022-02-01]
21. Kennedy NR, Steinberg A, Arnold RM, Doshi AA, White DB, DeLair W, et al. Perspectives on telephone and video communication in the intensive care unit during COVID-19. Ann Am Thorac Soc 2021 May;18(5):838-847 [FREE Full text] [doi: 10.1513/AnnalsATS.202006-729OC] [Medline: 33181033]
22. Sasangohar F, Dhala A, Zheng F, Ahmadi N, Kash B, Masud F. Use of telecritical care for family visitation to ICU during the COVID-19 pandemic: an interview study and sentiment analysis. BMJ Qual Saf 2021 Sep;30(9):715-721 [FREE Full text] [doi: 10.1136/bmjqs-2020-011604] [Medline: 33028659]
23. Dhala A, Sasangohar F, Kash B, Ahmadi N, Masud F. Rapid implementation and innovative applications of a virtual intensive care unit during the COVID-19 pandemic: case study. J Med Internet Res 2020 Sep 03;22(9):e20143 [FREE Full text] [doi: 10.2196/20143] [Medline: 32795997]
24. Shahdosti H, Mazlom SR, Vaghee S, Amini S. Evaluating the effect of planned online video visitations on anxiety and depression of patients at open heart intensive care unit: a randomized controlled trial. Iran Red Crescent Med J 2020 Aug 04;22(7):e102578 [FREE Full text] [doi: 10.5812/ircmj.102578]
25. Thomas KAS, O'Brien BF, Fryday AT, Robinson EC, Hales MJL, Karipidis S, et al. Developing an innovative system of open and flexible, patient-family-centered, virtual visiting in ICU during the COVID-19 pandemic: a collaboration of staff, patients, families, and technology companies. J Intensive Care Med 2021 Oct;36(10):1130-1140 [FREE Full text] [doi: 10.1177/08850666211030845] [Medline: 34291683]
26. Chheda S, Leiner M. Virtual neonatal visitation in an intensive care unit: advantages of technology during a pandemic. Pediatrics 2021;147(3_MeetingAbstract):455 [FREE Full text] [doi: 10.1542/peds.147.3MA5.455a]
27. Thibeau S, Ricouard D, Gilcrease C. Innovative technology offers virtual visitation for families. J Contin Educ Nurs 2012 Oct;43(10):439-440. [doi: 10.3928/00220124-20120925-94] [Medline: 23051777]
28. Yang NH, Dharmar M, Hojman NM, Sadorra CK, Sundberg D, Wold GL, et al. Videoconferencing to reduce stress among hospitalized children. Pediatrics 2014 Jul;134(1):e169-e175. [doi: 10.1542/peds.2013-3912] [Medline: 24982102]
29. Monin JK, Ali T, Syed S, Piechota A, Lepore M, Mourgues C, et al. Family communication in long-term care during a pandemic: lessons for enhancing emotional experiences. Am J Geriatr Psychiatry 2020 Dec;28(12):1299-1307 [FREE Full text] [doi: 10.1016/j.jagp.2020.09.008] [Medline: 33004262]
30. Sacco G, Lléonart S, Simon R, Noublanche F, Annweiler C, TOVID Study Group. Communication technology preferences of hospitalized and institutionalized frail older adults during COVID-19 confinement: cross-sectional survey study. JMIR Mhealth Uhealth 2020 Sep 18;8(9):e21845 [FREE Full text] [doi: 10.2196/21845] [Medline: 32896832]
31. Mendiola B, Gomez C, Furst C, Rasmussen-Winkler J. Facilitating virtual visitation in critical care units during a pandemic. Holist Nurs Pract 2021;35(2):60-64. [doi: 10.1097/HNP.0000000000000432] [Medline: 33555718]
32. Webb H, Parson M, Hodgson LE, Daswani K. Virtual visiting and other technological adaptations for critical care. Future Healthc J 2020 Oct;7(3):e93-e95 [FREE Full text] [doi: 10.7861/fhj.2020-0088] [Medline: 33094267]
33. Voo TC, Senguttuvan M, Tam CC. Family presence for patients and separated relatives during COVID-19: physical, virtual, and surrogate. J Bioeth Inq 2020 Dec;17(4):767-772 [FREE Full text] [doi: 10.1007/s11673-020-10009-8] [Medline: 32840835]
34. Volcy J, Smith W, Mills K, Peterson A, Kene-Ewulu I, McNair M, et al. Assessment of patient and provider satisfaction with the change to telehealth from in-person visits at an academic safety net institution during the COVID-19 pandemic. J Am Board Fam Med 2021 Feb;34(Suppl):S71-S76 [FREE Full text] [doi: 10.3122/jabfm.2021.S1.200393] [Medline: 33622821]
35. Poalillo FE, Geiling J, Jimenez EJ. Healthcare personnel and nosocomial transmission of pandemic 2009 influenza. Crit Care Med 2010 Apr;38(4 Suppl):e98-e102. [doi: 10.1097/CCM.0b013e3181d41d45] [Medline: 20154603]
36. Zumla A, Hui DS. Infection control and MERS-CoV in health-care workers. Lancet 2014 May 31;383(9932):1869-1871 [FREE Full text] [doi: 10.1016/S0140-6736(14)60852-7] [Medline: 24857701]
37. Magoon V. Operationalizing virtual visits during a public health emergency. Fam Pract Manag 2020;27(3):5-12 [FREE Full text] [Medline: 32393013]
38. Lipsitz LA, Lujan AM, Dufour A, Abrahams G, Magliozzi H, Herndon L, et al. Stemming the tide of COVID-19 infections in Massachusetts nursing homes. J Am Geriatr Soc 2020 Nov;68(11):2447-2453 [FREE Full text] [doi: 10.1111/jgs.16832] [Medline: 32930389]

39. Jansky M, Schade F, Rieder N, Lohrmann D, Gebel C, Kloppenburg L, PallPan Study Group. 'It felt like a black hole, great uncertainty, but we have to take care for our patients'-Qualitative findings on the effects of the COVID-19 pandemic on specialist palliative home care. PLoS One 2021;16(12):e0260767 [FREE Full text] [doi: 10.1371/journal.pone.0260767] [Medline: 34910741]
40. Epstein EG, Arechiga J, Dancy M, Simon J, Wilson D, Alhusen JL. Integrative review of technology to support communication with parents of infants in the NICU. J Obstet Gynecol Neonatal Nurs 2017;46(3):357-366 [FREE Full text] [doi: 10.1016/j.jogn.2016.11.019] [Medline: 28263727]
41. Rhoads SJ, Green A, Mitchell A, Lynch CE. Neuroprotective Core Measure 2: Partnering with Families - Exploratory study on web-camera viewing of hospitalized infants and the effect on parental stress, anxiety, and bonding. Newborn Infant Nurs Rev 2015 Sep;15(3):104-110. [doi: 10.1053/j.nainr.2015.06.011]
42. Mitchell M, Aitken L. Flexible visiting in ICU is highly rated by patients, family members and staff. Australian Critical Care 2012 May;25(2):136. [doi: 10.1016/j.aucc.2011.12.040]
43. Thomas G, Rajadurai R, Browne C, Wiggins N, Elias T. 14 The development and evaluation of a virtual visiting programme facilitating regular friends and family contact for hospital inpatients during the COVID-19 pandemic. BMJ Support Palliat Care 2021;11(Suppl 1):A13 [FREE Full text] [doi: 10.1136/spcare-2021-PCC.32]
44. Kentish-Barnes N, Cohen-Solal Z, Morin L, Souppart V, Pochard F, Azoulay E. Lived experiences of family members of patients with severe COVID-19 who died in intensive care units in France. JAMA Netw Open 2021 Jun 01;4(6):e2113355 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.13355] [Medline: 34152418]
45. Ashana DC, Cox CE. Providing family-centered intensive care unit care without family presence-human connection in the time of COVID-19. JAMA Netw Open 2021 Jun 01;4(6):e2113452 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.13452] [Medline: 34152421]
46. Banach DB, Bearman GM, Morgan DJ, Munoz-Price LS. Infection control precautions for visitors to healthcare facilities. Expert Rev Anti Infect Ther 2015;13(9):1047-1050. [doi: 10.1586/14787210.2015.1068119] [Medline: 26175335]
47. Howe A, Ray I, Roberts M, Urbanska M, Byrne Z. The psychology of security for the home computer user. 2012 Presented at: IEEE Symposium on Security and Privacy; May 20-23, 2012; San Francisco, CA. [doi: 10.1109/sp.2012.23]

Abbreviations

ICU: intensive care unit
MERS: Middle East respiratory syndrome
NICU: neonatal intensive care unit
PPE: personal protective equipment
RCT: randomized controlled trial

Edited by T de Azevedo Cardoso; submitted 16.10.22; peer-reviewed by K Le Du, H Alshawaf; comments to author 14.09.23; revised version received 07.11.23; accepted 08.11.23; published 28.11.23.

Please cite as:
Jeong H, Choi Y, Kim H
Nonface-to-Face Visitation to Restrict Patient Visits for Infection Control: Integrative Review
Interact J Med Res 2023;12:e43572
URL: https://www.i-jmr.org/2023/1/e43572
doi:10.2196/43572
PMID:38015595

Review

Phone-Based Text Therapy for Youth Mental Health: Rapid Review

Varun Karnik[1,2], BSc (Hons); Hamish Henderson[2], BSc; Urooj Raza Khan[1], PhD; James Boyd[1], PhD

[1]La Trobe University, Melbourne, Australia

[2]Griffith University, Gold Coast, Australia

Corresponding Author:
Varun Karnik, BSc (Hons)
La Trobe University
360 Collins Street
Melbourne, 3086
Australia
Phone: 61 3 9479 6000
Email: varun.karnik@griffithuni.edu.au

Abstract

Background: Mental illness has become a prevalent issue impacting adolescents worldwide. Many barriers, including stigma and poor health literacy, prevent this population group from accessing reliable mental health care services. Synchronous text–therapy counseling is an underused therapeutic approach in combating adolescent mental illness. Phone-based text therapy is uniquely placed to offer personalized counseling to adolescents through a familiar and engaging treatment modality.

Objective: This rapid review aims to understand the clinical effectiveness, usability, and accessibility of phone-based text therapy for youth mental health.

Methods: Cochrane CENTRAL, Embase, PubMed, and PsycINFO were used to search for suitable literature. Five groups of keywords were used: those related to (1) "therapy," (2) "text," (3) "phone," (4) "youth," and (5) "mental health." Eligibility criteria were formed through the PICO (Population, Intervention, Control, and Outcome) framework. Studies were included if a synchronous phone-based text therapy intervention was used in an adolescent population, with an age range of 12-24 years. Only literature available in full-text, English, and a peer-reviewed journal was considered. Furthermore, a date limit of 5 years was set to reflect the recent development of digital interventions for mental health. Pertinent information from each study was tabulated, and a narrative synthesis was used to assess, describe, and organize the included studies comprehensively and concisely.

Results: Of the 771 studies dual screened, 7 studies were included in this rapid review. Most of the exclusions occurred due to the use of the wrong intervention, such as asynchronous messaging. The selected studies had a low risk of bias and were suitable for the review. All interventional trials demonstrated reductions in mental health symptoms, primarily depression and anxiety. Most studies displayed high usability among participants, while data were unclear regarding accessibility.

Conclusions: This review reveals the high potential of phone-based text therapy as an intervention for adolescents experiencing mental illness. We hope that this review promotes further refinement of text-based phone therapies and encourages future research on this subject matter.

(Interact J Med Res 2023;12:e47250) doi:10.2196/47250

KEYWORDS

text therapy; mHealth; adolescent health; distance counseling; mental illness; mobile health intervention; adolescent; health promotion; digital mental health intervention

Introduction

Rationale

The rates of mental illness have increased rapidly worldwide, with 29% of individuals experiencing mental health issues at some point in their lifetime [1]. Mental disorders significantly impact the quality of life and are predicted to cost the global economy over US $16 USD trillion between 2011 and 2030 [2]. The mental health of our youth is of particular concern, as 50% of adolescents are currently experiencing mental health problems and 70% of mental disorder diagnoses occur before the age of 25 years [3,4].

Despite the numerous services offered, a large proportion of affected youth remain unsupported. Often termed the "missing-middle," this population is not reached by mental health services due to a variety of barriers that contribute to a lower access rate. Barriers include a shortage of resources; prolonged waiting lists; poor health literacy; and concerns related to trust, stigma, and confidentiality [5-7]. While primary care–based models, such as *Headspace* in Australia, can help in the initial stage of mental health problems, evidence shows that adolescents often disengage from these services much earlier than desirable [8,9]. Furthermore, only a minority of those who remain engaged experience positive outcomes [10]. It is evident that additional services are required to provide more substantial and sustained care for our youth.

A growing field of interest is digital mental health, which encompasses any application of digital technology to the assessment, prevention, or treatment of mental health issues [11]. In a digital world, where most communication is performed on smartphones, the use of phone-based text therapy has emerged as a promising option. This has the potential to connect youth with therapists and mentors, improve accessibility to resources, and manage youth mental health through an engaging medium. Text therapy has been investigated in the literature, with a recent scoping review from Dwyer et al [12] analyzing 70 studies and concluding that text-based services can deliver effective mental health support to the population. Interestingly, the authors asserted that the assumed limitations of text therapy may act as an advantage, with increased anonymity effectively assuaging patient concerns about confidentiality and privacy [12].

While the general use of text therapy to deliver mental health support has been explored, there is limited evidence focusing on the delivery of these interventions through a phone to the adolescent population. Consequently, we performed a rapid review to collate information regarding phone-based text therapy for youth mental health, informing the design and delivery of future digital mental health interventions.

Objectives

This rapid review aimed to collate and summarize current evidence relating to phone-based text therapy options for youth mental health. The following research questions were asked: What is the effectiveness of phone-based synchronous text therapy options for youth mental health in terms of (1) clinical outcomes and (2) usability and accessibility?

Methods

Eligibility Criteria

The eligibility criteria for this review were developed using the PICO (Population, Intervention, Control, and Outcome) framework. In this study, the "population" encompassed youth experiencing mental health symptoms. Youth was defined using the Australian Institute of Welfare and Health definition as individuals aged 12-24 years [13]. All mental health symptoms were included, with particular emphasis on the most prevalent symptoms of depression and anxiety [13]. The "intervention" was any phone-based text mental health therapy intervention. This intervention had to be synchronous, provided by a human therapist or support staff with adequate training, and delivered through an app or service available on a phone. Combination interventions were included if text therapy was 1 of the main components of the treatment. Where applicable, the "control" was a lack of treatment or any comparison treatment. This included individuals on a waitlist, self-help therapy, face-to-face therapy, and telehealth services. Finally, the "outcome" was the effectiveness of treatment in terms of (1) reducing clinical outcomes, such as mental health symptoms, and (2) usability and accessibility.

According to the *Cochrane Rapid Reviews Interim Guidance*, only literature available in full text, English, and a peer-reviewed journal was considered [14]. The literature was date limited to 5 years, in order to reflect the recent development of digital interventions for mental health and to ensure the services explored in the review were contemporary and comparable. Observational and experimental studies were included. Studies were excluded if they described automated, nonsynchronous, scheduled, or unidirectional messaging. Similarly, any articles that used text therapy to manage medication compliance, reminders regarding health care, or treatment adherence were also excluded.

Information Sources

The following electronic databases were searched: Cochrane CENTRAL, Embase, PubMed, and PsycINFO. The reference lists of each article selected for full-text review were also manually searched.

Search Strategy

Advanced search functionality was used, with searches including keyword truncations, Medical Subject Headings terms, and subject headings. An information specialist from the Griffith University assisted with the development of the strategy. The searches were conducted in December 2022, using five groups of keywords: those related to (1) "therapy," (2) "text," (3) "phone," (4) "youth," and (5) "mental health." Full search strategies for each database are included in Multimedia Appendix 1.

Selection Process

Literature retrieved from searches was exported into EndNote X9 (Clarivate). These were then transferred into Covidence (Veritas Health Innovation Ltd), a digital data extraction and screening tool. Duplicates were removed, and items were screened in accordance with the eligibility criteria by screening the title or abstract (1) and then by full text (2). One reviewer screened all titles or abstracts, while the full-text screening was undertaken by 2 independent reviewers (VK and HH). Any disagreements were discussed extensively, and a third reviewer (JB) was brought in if needed. The *Cochrane Rapid Reviews Interim Guidance* from the *Cochrane Rapid Reviews Methods Group* was used [14], and the selection process was recorded via a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) chart [15] (Multimedia Appendix 2).

Data Collection Process

Data were extracted from the literature into an Excel (Microsoft Corp) table. The extraction was completed by 1 researcher and then reviewed by another (VK and HH). If inconsistencies arose, a third researcher (JB) was involved to achieve agreement. The information extracted included (1) information about the study, including the authors, publication date, study type, and location; (2) information about the study participants, including the sample size, age, and sex; (3) for research question 1, information about the effect on clinical outcomes, primarily depression and anxiety, or any other mental health symptoms; and (4) for research question 2, information about accessibility and usability.

Study Risk of Bias Assessment

The appropriate tool was used to evaluate the literature included in the review:

- Randomized controlled trial: risk of bias in randomized trials tool [16]
- Uncontrolled studies: a quality assessment tool for quantitative studies by the Effective Public Health Practice Project [17]
- Cohort studies: Risk of Bias in Non-Randomized Studies of Interventions (ROBINS-1) tool [18]
- Cross-sectional studies: Appraisal Tool for Cross-Sectional Studies (AXIS) [19]
- Case control studies: Critical Appraisal Programme Skills [20]
- Qualitative studies: Joanna Briggs Institute Critical Appraisal Checklist for Qualitative Research [21]
- Ecological studies: Modification of Strengthening the Reporting of Observational Studies in Epidemiology. [22]

Synthesis Methods

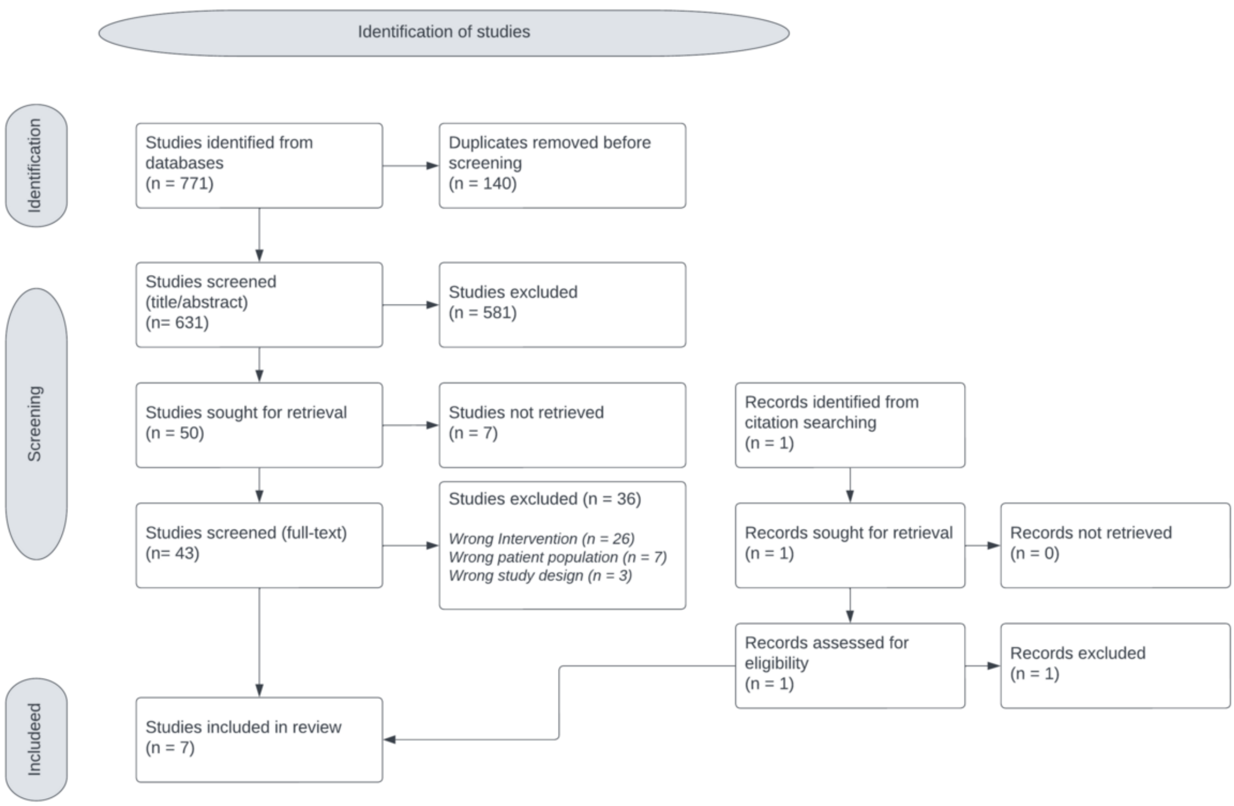
A narrative synthesis was used. First, the pertinent information from each study was tabulated. Following this, the results of each included study were assessed, described, and organized in a comprehensive and concise format [23].

Results

Study Selection

Studies were selected using the processes shown in the PRISMA diagram (Figure 1). Searches were executed on December 20, 2022, returning 771 citations. A total of 140 duplicate studies were eliminated, resulting in 631 studies available for initial screening. Subsequently, 581 studies were excluded based on title and abstract assessment. Additionally, 7 studies proved inaccessible for retrieval, and 31 more were eliminated after a thorough examination of their full texts. Finally, 7 studies were included in the rapid review. Studies were excluded primarily due to the implementation of the wrong intervention such as the use of automated or asynchronous messaging. Dual screening was performed for the full-text screening, with any disagreements being extensively discussed. Furthermore, the reference lists for articles selected for full-text screening were manually searched, with 1 study being assessed for eligibility, but ultimately excluded.

Figure 1. Untitled.



Study Characteristics

Information on each study is shown in Table 1. The studies were conducted in a variety of settings: Australia [24], Europe [25,26], Asia [27], and North America [28-30]. Several trials contained both quantitative and qualitative components [24,27,28], and some were purely quantitative [25,26,29,30]. The majority (4/7, 57%) were interventional trials, with 3 uncontrolled trials [24,25,28] and 1 randomized controlled trial [26]. The remaining trials (3/7, 43%) consisted of a cross-sectional study [30], an ecological study [29], and a retrospective audit study [27].

Table 1. Results of interventional trials.

Reference and study characteristics	Intervention	Participants	Clinical effectiveness	Usability
• Alvarez-Jimenez et al [24] (2020) • Study type: uncontrolled single-group trial • Setting: Australia (Oceania)	• Intervention: MOST+[a] for 9 weeks with a qualified therapist • Time accessible: 4 PM and 12 AM • Control: N/A[b]	• N=157 • Mean age 19.1 (SD 2.3) years • 121 (77%) female	• Statistical improvements in psychological distress, perceived stress, well-being, depression, loneliness, social support, autonomy, and self-competence after MOST+ use • 82% of participants reported that intervention helped them feel better • 86% of participants reported more social connectedness • Correlation between self-reported benefits and web-based messages with clinicians	• 98% of participants reported positive experience • 86% of participants reported easy to use • 92% of participants would recommend it to others
• Lindqvist et al [26] (2020) • Study type: randomized controlled trial • Setting: Sweden (Europe)	• Intervention: IPDT[c] for 8 weeks with a qualified therapist • Time accessible: whenever needed • Control: supportive contact over the internet	• N=76 • Mean age 16.6 years • 80% female	• Patients in the intervention group (IPDT) saw a significantly faster decline in depression symptoms compared to those in the control group improvements • Intervention reduced anxiety • Intervention increased emotional regulation and self-compassion • Results regarding depression and anxiety symptoms were stable at follow-up after 6 months	• 12% of participants dropped out • 3% of participants complained of loneliness when using the text-based intervention
• Goldin et al [25] (2019) • Study type: uncontrolled single-group trial • Setting: Finland (Europe)	• Intervention: ascend program for 8 weeks with a qualified therapist • Time accessible: whenever needed • Control: N/A	• N=22 • Mean age 23.2 years • 100% female	• Significant reduction in depression • Correlation between the number of days of practice and reduction in depressive symptoms • Correlation between the number of weeks of group chat use and reduction in depressive symptoms	• 27% of participants dropped out
• Chyzzy et al [28] (2020) • Study type: uncontrolled single group trial + qualitative component • Setting: Canada (North America)	• Intervention: MPPS[d] for 7 months with a trained mentor • Time accessible: whenever needed • Control: N/A	• N=21 • Mean age 21.3 years • 100% female	• 65.1% of participants confirmed improvements in stress and coping • Correlation between the number of contacts and participant satisfaction • 43.4% of participants confirmed that improvements in stress or coping were due to intervention • 59.7% of participants confirmed that improvements in social integration were due to intervention	• 78.2% of participants reported convenience of support • 79.2% of participants reported they enjoyed the access to support • 37.5% of participants enjoyed receiving peer support by phone, particularly messaging

[a]MOST+: Moderated Online Social Therapy+.

[b]N/A: not available.

[c]IPDT: internet-based psychodynamic therapy.

[d]MPPS: mobile phone–based peer support.

Participants

The interventional studies had a wide range of sample sizes (from n=21 [28] to n=157 [24]). The participants in the interventional trials were predominantly female, with studies by Goldin et al [25] and Chyzzy et al [28] consisting of a 100% female population, and Alvarez-Jimenez et al [24] and Lindqvist et al [26] were made up of 77% and 80% female participants, respectively. These trials all had participants of similar ages, with the mean age ranging from 16.6 [26] to 23.2 years [25]. The noninterventional trials comprised various observational studies and had larger sample sizes (n=662 [30] to n=849,483 [29]). Yip et al [27] and Thompson et al [29] lacked information about age and sex, while Toscos et al [30] slightly favored female participants, with a mean age of 20.91 years.

Intervention Characteristics

Phone-Based Text Therapy

Yip et al [27] and Thompson et al [29] collated data from an adolescent population that solely used a text therapy service available on phones. These services, known as Crisis Text Line and OpenUp, are free text counseling support systems that enable adolescents to connect with social workers and volunteer counselors.

Combination Therapies

The other interventions largely consisted of combination digital health treatments, with synchronous text therapy comprising a significant component. Goldin et al [25] and Alvarez-Jimenez et al [24] used a combination of digital interventions named Moderated Online Social Therapy+ and the Ascend Program, respectively. These interventions consisted of 3 separate components, including a self-learning component (modules and daily exercises), a group connection (peer social network [24] and group chat [25]), and private phone–based text therapy with a qualified therapist [24,25]. Toscos et al [30] examined youth perceptions of “tele-mental health services” in their cross-sectional study, which was described as self-help materials, web-based therapy, and a phone-based text therapy service. Finally, Chyzzy et al [28] and Lindqvist et al [26] used a combination of digital interventions with 2 components named mobile phone–based peer support and internet-based psychodynamic therapy, respectively. One study used a call service and phone messaging support [28], and another used modules and phone-based text therapy [26].

Support Staff and Time Used

The support staff connected to participants consisting of qualified therapists [24-26] and volunteer counselors or mentors [27-29]. Several interventional trials (n=3) consisted of a use period of approximately 2 months [24-26]. Chyzzy et al [28] allowed support for 7 months, as this study focused on postpartum depression and included support before and after giving birth. The majority of the interventional trials allowed phone-based text support whenever required [25,26,28]; however, 1 study only provided access from 4 PM to 12 AM [24]. As for the noninterventional studies, Yip et al [27] and Thompson et al [29] included information from 4 and 5 years of data, respectively [27,29].

Results of Syntheses

The results from the collated studies are outlined in Table 2. Due to the heterogeneity of the primary interventions used across the included studies, a meta-analysis was not performed.

Table 2. Results of included noninterventional trials.

Reference	Data studied	Participants	Accessibility and Usability
• Yip et al [27] (2021) • Study type: retrospective audit • Setting: Hong Kong (Asia)	A survey from users of OpenUp chat services between 2018 and 2021	• N=29,400 • Mean age unknown • Sex: unknown	• From October 2018 to June 2021: OpenUp served 81,654 total sessions • 13,244 (81.5%) participants found the service helpful • 12,688 (85.4%) participants were likely to seek help in the future • 5015 (79%) participants found the community information helpful • 10,143 (34.5%) participants used OpenUp services multiple times • 11,917 (45.5%) users had never sought help from others about the issue they discussed in OpenUp. • 2471 (85.3%) high risk and crisis cases, in terms of suicide risk levels, were lowered by the intervention
• Toscos et al [30] (2018) • Study type: cross-sectional study • Setting: United States (North America)	Survey of college students about telemental health	• N=662 • Mean age 20.91 (SD 1.69) years • 438 (66%) female	• 88 (13.8%) participants had used text therapy before • 452 (68.3%) students would prefer to talk to someone, face to face, about their stress or problems in person • 111 (16.8%) students preferred text or web-based chat, 61 (9.2%) students preferred a phone call, 12 (1.8%) students preferred a video chat, 5 (0.8%) students preferred social media, and 8 (1.2%) students preferred something else • Independent of depression, sex, or stress level, participants preferred face-to-face versus web-based therapy • Students reported the most interest in using self-help resources (n=258, 40.1%), followed by a web-based therapist (n=184, 28.8%), followed by anonymous web-based chats with trained nonprofessionals (n=158, 24.6%)
• Thompson et al [29] (2018) • Study type: ecological study • Setting: United States (North America)	Use of crisis text line from 2013 to 2017 in different areas	• N=849,483 • Mean age unknown • Sex: unknown	• The biggest predictor of CTL nonuse was rural communities • CTL use was also associated with higher mean household incomes, higher divorce rates, and lower residential stability • Did not eliminate socioeconomic disparities in service use • Use not dependent upon on-ground mental health services

Clinical Effectiveness

All 4 interventional trials demonstrated a reduction in mental health symptoms. Improvements in depression, anxiety, psychological distress, and emotional regulation were confirmed [24-26,28]. Other reported outcomes included an improvement in social connectedness, social integration, and coping mechanisms [24,28]. Correlations between text-based messaging and mental health outcomes were reported by 2 studies [24,25]. Only 1 randomized controlled trial was included, with Lindqvist et al [26] demonstrating that patients in the intervention group had a significantly faster improvement in mental health symptoms when compared to a control group.

Usability

In terms of usability, most participants reported a positive experience with phone-based text therapy interventions. Users reported that text therapy interventions were easy to use (86% from Alvarez-Jimenez et al [24]), highly convenient (78.2% from Chyzzy et al [28]), reusable (34.5% from Alvarez-Jimenez et al [24]), and helpful (79% from Alvarez-Jimenez et al [24]). Notably, studies described increases in feelings of autonomy, self-competence, and self-compassion [24,26]. Conversely, in 1 study, 27% of participants dropped out [25], while in another study, 3% of participants made complaints regarding loneliness when using the intervention [26].

Accessibility

Regarding accessibility, most of the information provided was from noninterventional studies. Chyzzy et al [28] reported that 79.2% of users enjoyed the accessibility of the interventions. Toscos et al [30] indicated participants are more comfortable engaging in face-to-face therapy compared to web-based or phone-based therapy. When comparing all forms of digital mental health support, it was evident that 24.6% of participants preferred using text or web-based chats. Thompson et al [29] indicated that there was a higher use and accessibility of text therapy in communities with higher mean household incomes, higher divorce rates, and lower residential stability. Most notably, text-line use was not dependent upon existing physical and mental health services, and rurality was the biggest predictor of nonuse.

Quality Assessment

All studies were subject to quality assessment with the appropriate tool, and the results are shown in Multimedia Appendix 3. All studies had a low risk of bias and were suitable to include in the review. The cross-sectional study, uncontrolled trials, and ecological study were associated with some bias due to the nature of their study types, and this was considered during interpretation.

Discussion

Principal Findings

This rapid review aimed to collate information regarding the clinical effectiveness, usability, and accessibility of phone-based text therapy interventions for youth mental health. This review identified 7 studies that used phone-based text therapy as a primary component in a digital mental health intervention for adolescents. Improvements in clinical outcomes were seen across all studies, particularly in symptoms of depression and anxiety. Furthermore, combination interventions using phone-based text therapy were seen to be generally usable, while evidence regarding accessibility was mixed.

Our study corroborates results from a previous review by Dwyer et al [12], which explored text-based interventions in the general population. This signifies that similar clinical benefits can be seen when applying text therapy in a younger cohort. Specifically, our review presented improvements in a wide array of mental health symptoms, including distress, well-being, depression, and anxiety. While the extensive use of combination digital health treatments limits our examination of phone-based text therapy as a sole intervention, general inferences can still be made. Notably, an included study from Alvarez-Jimenez et al [24] reported a correlation between the number of text-based messages exchanged with a clinician, and the self-reported benefit in well-being. This demonstrates that the text therapy component of the combination digital intervention was a major contributing factor to the reported clinical benefit. Goldin et al [25] presented similar results, indicating that the reduction in depressive symptoms was predicted by the use of the group chat function. While Goldin et al [25] emphasized the use of the group chat function as opposed to the private text therapy component, these results still suggest that great benefit comes from connecting patients to support via phone. In addition to improving well-being through direct counseling, phone-based text therapy enhances social connectedness to therapists, mentors, and peers [31]. This fulfills the need for belonging and security that have been shown to be extremely important in developing adolescents [31,32].

Theoretically, digital mental health interventions can alleviate stress by providing "received support," including advice, emotional comfort, and empathy [33,34]. Interestingly, research has shown that "perceived support," which is merely the trust that support will be readily available if needed, may have a stronger influence on mental health outcomes [33,34]. In most of the interventional trials, phone-based text therapy was available whenever required, and this "perceived support" undoubtedly had a strong therapeutic influence on the mental health of participants. The "perceived support" afforded to adolescents had multiple other benefits, with our results noting improvements in autonomy, self-competence, and self-compassion [24,26]. This suggests that the self-directedness, flexibility, and freedom of phone-based text therapy allow adolescents to be independent in the pursuit of their recovery. This could improve their self-esteem and maturity while concurrently providing the mental health therapy required.

The usability of digital interventions involving phone-based text therapy was consistently described as easy to use, helpful, reusable, and convenient. This could be attributed to the high digital literacy levels of younger people, specifically their familiarity with, and reliance on, phone-based communication [35]. A known limitation of text therapy is the difficulty in forming meaningful interpersonal relationships [36]. One study did reflect this, noting that a very small number of participants complained of loneliness despite text support being available 24/7 [26]. Evidently, while some adolescents enjoy the advantages of text therapy, others may require a therapeutic relationship characterized by face-to-face counseling. This signifies the nuances in mental health therapy and emphasizes the importance of an individualistic and bespoke approach to optimize treatment.

Our review presents varied results in terms of accessibility, with 1 study asserting that adolescents enjoyed accessing digital text interventions, while another study indicated that young college students were more comfortable engaging in face-to-face therapy. This contrasts with the literature, as multiple studies have concluded that text-based digital counseling options are more accessible to youth, due to their flexibility, anonymity, and affordability [12,37,38]. Toscos et al [30] attributed the conflicting preference for face-to-face care in their study to lack of knowledge and previous experiences. While the literature might indicate that phone-based text therapy is widely accessible for the youth population, Thompson et al [29] demonstrated that text therapy is still not well used in rural areas, even in areas with a strong established presence of physical and mental health services. This suggests that while text therapy might be more accessible to some adolescents, there are still unique barriers in rural areas that require further research [39,40].

Limitations and Future Directions

This rapid review has numerous limitations which must be considered. First, to provide more timely information the literature search was limited in various ways: (1) the "mental health" component of the search strategy was limited to primarily depression and anxiety, (2) only 4 databases were searched, and (3) a 5-year date limit was implemented. Furthermore, there is a lack of standard nomenclature in this space and a subsequent heterogeneity of terms used to describe text-based synchronous interventions. All these factors may have contributed to some relevant studies being missed in the literature search. Additionally, the studies included were highly heterogeneous, with most using text therapy as a component of a combined intervention. This reduces the strength of conclusions made in this review, as it was difficult to elucidate the impact of text therapy as a sole treatment option. Furthermore, the interventional studies included consisted primarily of uncontrolled trials, which limits our ability to confirm clinical effectiveness. Nevertheless, the positive results in these studies do establish that there is a clinical effect worth further investigation.

Due to their cost-effectiveness and ease of implementation, we expect an increase in the use of text therapy in youth mental health. While feasibility studies show promising results, the clinical viability of this form of therapy requires further investigation. Additional research should be performed using phone-based text therapy as the sole mental health intervention in the youth population. While combination interventions have advantages, it is important to establish the effectiveness of each component to optimize these therapies. Research focusing on the perceptions of the rural youth population on phone-based text interventions must be emphasized, as these areas experience the most significant shortages in mental health resources. An improvement in our understanding of this will greatly improve future interventions.

Conclusions

In conclusion, this rapid review provides a collation of recent research into phone-based text therapy options for youth mental health. Our review provided mixed evidence regarding accessibility but strongly indicated that phone-based text therapy improves mental health symptoms and is generally usable. We hope this review informs the design and delivery of phone-based text therapy interventions for adolescents struggling with their mental health and instigates further research in this area.

Acknowledgments

This research was supported by Digital Health Cooperative Research Center Limited. Digital Health CRC Limited is funded under the Commonwealth Government's Cooperative Research Centres Program.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search Strategies.
[DOCX File , 17 KB - ijmr_v12i1e47250_app1.docx]

Multimedia Appendix 2
PRISMA Checklist.
[DOCX File , 21 KB - ijmr_v12i1e47250_app2.docx]

Multimedia Appendix 3
Quality Assessments.
[DOCX File , 26 KB - ijmr_v12i1e47250_app3.docx]

References

1. Steel Z, Marnane C, Iranpour C, Chey T, Jackson JW, Patel V, et al. The global prevalence of common mental disorders: a systematic review and meta-analysis 1980-2013. Int J Epidemiol 2014;43(2):476-493 [FREE Full text] [doi: 10.1093/ije/dyu038] [Medline: 24648481]
2. Anthes E. Mental health: there's an app for that. Nature 2016;532(7597):20-23 [FREE Full text] [doi: 10.1038/532020a] [Medline: 27078548]
3. Caspi A, Houts RM, Ambler A, Danese A, Elliott ML, Hariri A, et al. Longitudinal assessment of mental health disorders and comorbidities across 4 decades among participants in the dunedin birth cohort study. JAMA Netw Open 2020;3(4):e203221 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.3221] [Medline: 32315069]
4. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. Arch Gen Psychiatry 2005;62(6):593-602 [FREE Full text] [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
5. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
6. Rickwood DJ, Deane FP, Wilson CJ. When and how do young people seek professional help for mental health problems? Med J Aust 2007;187(S7):S35-S39. [doi: 10.5694/j.1326-5377.2007.tb01334.x] [Medline: 17908023]
7. Lawrence D, Hafekost J, Johnson S, Boterhoven de Haan K, Sawyer M, Ainley J, et al. The Mental Health of Children and Adolescents. Report on the second Australian Child and Adolescent Survey of Mental Health and Wellbeing. Department of Health. Canberra; 2015. URL: https://research.acer.edu.au/well_being/1/ [accessed 2023-10-31]
8. Booth ML, Bernard D, Quine S, Kang MS, Usherwood T, Alperstein G, et al. Access to health care among Australian adolescents young people's perspectives and their sociodemographic distribution. J Adolesc Health 2004;34(1):97-103. [doi: 10.1016/j.jadohealth.2003.06.011] [Medline: 14706412]

9. Hoagwood K, Burns BJ, Kiser L, Ringeisen H, Schoenwald SK. Evidence-based practice in child and adolescent mental health services. Psychiatr Serv 2001;52(9):1179-1189 [FREE Full text] [doi: 10.1176/appi.ps.52.9.1179] [Medline: 11533391]
10. Iorfino F, Carpenter JS, Cross SP, Crouse J, Davenport TA, Hermens DF, et al. Social and occupational outcomes for young people who attend early intervention mental health services: a longitudinal study. Med J Aust 2022;216(2):87-93 [FREE Full text] [doi: 10.5694/mja2.51308] [Medline: 34664282]
11. Wies B, Landers C, Ienca M. Digital mental health for young people: a scoping review of ethical promises and challenges. Front Digit Health 2021;3:697072 [FREE Full text] [doi: 10.3389/fdgth.2021.697072] [Medline: 34713173]
12. Dwyer A, de Almeida Neto A, Estival D, Li W, Lam-Cassettari C, Antoniou M. Suitability of text-based communications for the delivery of psychological therapeutic services to rural and remote communities: scoping review. JMIR Ment Health 2021;8(2):e19478 [FREE Full text] [doi: 10.2196/19478] [Medline: 33625373]
13. Australia's youth. Australian Institute of Health and Welfare. 2021. URL: https://www.aihw.gov.au/reports/children-youth/australias-youth/contents/summary [accessed 2023-10-31]
14. Garritty C, Gartlehner G, Nussbaumer-Streit B, King VJ, Hamel C, Kamel C, et al. Cochrane rapid reviews methods group offers evidence-informed guidance to conduct rapid reviews. J Clin Epidemiol 2021;130:13-22 [FREE Full text] [doi: 10.1016/j.jclinepi.2020.10.007] [Medline: 33068715]
15. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
16. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
17. Thomas BH, Ciliska D, Dobbins M, Micucci S. A process for systematically reviewing the literature: providing the research evidence for public health nursing interventions. Worldviews Evid Based Nurs 2004;1(3):176-184. [doi: 10.1111/j.1524-475X.2004.04006.x] [Medline: 17163895]
18. Risk of bias in non-randomised studies of interventions (ROBINS-I). Cochrane Methods. URL: https://methods.cochrane.org/bias/risk-bias-non-randomized-studies-interventions [accessed 2023-10-31]
19. Downes MJ, Brennan ML, Williams HC, Dean RS. Development of a critical appraisal tool to assess the quality of cross-sectional studies (AXIS). BMJ Open 2016;6(12):e011458 [FREE Full text] [doi: 10.1136/bmjopen-2016-011458] [Medline: 27932337]
20. CASP cohort study checklist. Critical Appraisal Skills Programme. 2018. URL: https://casp-uk.net/wp-content/uploads/2018/01/CASP-Cohort-Study-Checklist_2018.pdf [accessed 2023-10-31]
21. Appendix 2.1: JBI critical appraisal checklist for qualitative research. Joanna Briggs Institute. 2019. URL: https://jbi.global/sites/default/files/2019-05/JBI_Critical_Appraisal-Checklist_for_Qualitative_Research2017_0.pdf [accessed 2023-10-31]
22. Dufault B, Klar N. The quality of modern cross-sectional ecologic studies: a bibliometric review. Am J Epidemiol 2011;174(10):1101-1107 [FREE Full text] [doi: 10.1093/aje/kwr241] [Medline: 21940800]
23. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis Without Meta-Analysis (SWiM) in systematic reviews: reporting guideline. BMJ 2020;368:l6890 [FREE Full text] [doi: 10.1136/bmj.l6890] [Medline: 31948937]
24. Alvarez-Jimenez M, Rice S, D'Alfonso S, Leicester S, Bendall S, Pryor I, et al. A novel multimodal digital service (moderated online social therapy+) for help-seeking young people experiencing mental ill-health: pilot evaluation within a national youth e-mental health service. J Med Internet Res 2020;22(8):e17155 [FREE Full text] [doi: 10.2196/17155] [Medline: 32788151]
25. Goldin PR, Lindholm R, Ranta K, Hilgert O, Helteenvuori T, Raevuori A. Feasibility of a therapist-supported, mobile phone-delivered online intervention for depression: longitudinal observational study. JMIR Form Res 2019;3(1):e11509 [FREE Full text] [doi: 10.2196/11509] [Medline: 30682726]
26. Lindqvist K, Mechler J, Carlbring P, Lilliengren P, Falkenström F, Andersson G, et al. Affect-focused psychodynamic internet-based therapy for adolescent depression: randomized controlled trial. J Med Internet Res 2020;22(3):e18047 [FREE Full text] [doi: 10.2196/18047] [Medline: 32224489]
27. Yip PSF, Chan WL, Chan CS, He L, Xu Y, Chan E, et al. The opportunities and challenges of the first three years of open up, an online text-based counselling service for youth and young adults. Int J Environ Res Public Health 2021;18(24):13194 [FREE Full text] [doi: 10.3390/ijerph182413194] [Medline: 34948802]
28. Chyzzy B, Nelson LE, Stinson J, Vigod S, Dennis CL. Adolescent mothers' perceptions of a mobile phone-based peer support intervention. Can J Nurs Res 2020;52(2):129-138. [doi: 10.1177/0844562120904591] [Medline: 32036679]
29. Thompson LK, Sugg MM, Runkle JR. Adolescents in crisis: a geographic exploration of help-seeking behavior using data from crisis text line. Soc Sci Med 2018;215:69-79 [FREE Full text] [doi: 10.1016/j.socscimed.2018.08.025] [Medline: 30216891]
30. Toscos T, Carpenter M, Drouin M, Roebuck A, Kerrigan C, Mirro M. College students' experiences with, and willingness to use, different types of telemental health resources: do gender, depression/anxiety, or stress levels matter? Telemed J E Health 2018;24(12):998-1005. [doi: 10.1089/tmj.2017.0243] [Medline: 29658826]

31. Lamblin M, Murawski C, Whittle S, Fornito A. Social connectedness, mental health and the adolescent brain. Neurosci Biobehav Rev 2017;80:57-68. [doi: 10.1016/j.neubiorev.2017.05.010] [Medline: 28506925]
32. Weziak-Bialowolska D, Bialowolski P, Lee MT, Chen Y, VanderWeele TJ, McNeely E. Prospective associations between social connectedness and mental health. Evidence from a longitudinal survey and health insurance claims data. Int J Public Health 2022;67:1604710 [FREE Full text] [doi: 10.3389/ijph.2022.1604710] [Medline: 35755953]
33. Lakey B, Orehek E. Relational regulation theory: a new approach to explain the link between perceived social support and mental health. Psychol Rev 2011;118(3):482-495. [doi: 10.1037/a0023477] [Medline: 21534704]
34. Wills TA, Shinar O. Measuring perceived and received social support. In: Social Support Measurement and Intervention: A Guide for Health and Social Scientists. New York, NY, US: Oxford University Press; 2000:86-135.
35. Kumpulainen K, Sairanen H, Nordström A. Young children's digital literacy practices in the sociocultural contexts of their homes. J Early Childhood Literacy 2020;20(3):472-499 [FREE Full text] [doi: 10.1177/1468798420925116]
36. Sucala M, Schnur JB, Constantino MJ, Miller SJ, Brackman EH, Montgomery GH. The therapeutic relationship in e-therapy for mental health: a systematic review. J Med Internet Res 2012;14(4):e110 [FREE Full text] [doi: 10.2196/jmir.2084] [Medline: 22858538]
37. Hoermann S, McCabe KL, Milne DN, Calvo RA. Application of synchronous text-based dialogue systems in mental health interventions: systematic review. J Med Internet Res 2017;19(8):e267 [FREE Full text] [doi: 10.2196/jmir.7023] [Medline: 28784594]
38. Navarro P, Bambling M, Sheffield J, Edirippulige S. Exploring young people's perceptions of the effectiveness of text-based online counseling: mixed methods pilot study. JMIR Ment Health 2019;6(7):e13152 [FREE Full text] [doi: 10.2196/13152] [Medline: 31271149]
39. Barry MM, Doherty A, Hope A, Sixsmith J, Kelleher CC. A community needs assessment for rural mental health promotion. Health Educ Res 2000;15(3):293-304 [FREE Full text] [doi: 10.1093/her/15.3.293] [Medline: 10977377]
40. Hirsch JK. A review of the literature on rural suicide: risk and protective factors, incidence, and prevention. Crisis 2006;27(4):189-199. [doi: 10.1027/0227-5910.27.4.189] [Medline: 17219751]

Abbreviations

AXIS: Appraisal Tool for Cross-Sectional Studies
PICO: Population, Intervention, Control, and Outcome
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
ROBINS-1: Risk of Bias in Non-Randomized Studies of Interventions

Edited by T de Azevedo Cardoso; submitted 26.03.23; peer-reviewed by J Cornelius, G Puinean; comments to author 30.05.23; revised version received 14.06.23; accepted 10.07.23; published 14.12.23.

Please cite as:
Karnik V, Henderson H, Khan UR, Boyd J
Phone-Based Text Therapy for Youth Mental Health: Rapid Review
Interact J Med Res 2023;12:e47250
URL: https://www.i-jmr.org/2023/1/e47250
doi:10.2196/47250
PMID:38096012

Short Paper

Appropriateness and Comprehensiveness of Using ChatGPT for Perioperative Patient Education in Thoracic Surgery in Different Language Contexts: Survey Study

Chen-ye Shao[1*], MD; Hui Li[2*], MD; Xiao-long Liu[3*], PhD; Chang Li[1], PhD; Li-qin Yang[1], PhD; Yue-juan Zhang[1], MD; Jing Luo[1], MD; Jun Zhao[1], PhD

[1]Department of Thoracic Surgery, The First Affiliated Hospital of Soochow University, Suzhou, China

[2]Department of Obstetrics and Gynecology, The First Affiliated Hospital of Soochow University, Suzhou, China

[3]Department of Cardiothoracic Surgery, Jinling Hospital, Medical School of Nanjing University, Nanjing, China

[*]these authors contributed equally

Corresponding Author:
Jun Zhao, PhD
Department of Thoracic Surgery
The First Affiliated Hospital of Soochow University
899 Pinghai Road
Gusu District
Suzhou, 215006
China
Phone: 86 15250965957
Email: zhaojia0327@126.com

Abstract

Background: ChatGPT, a dialogue-based artificial intelligence language model, has shown promise in assisting clinical workflows and patient-clinician communication. However, there is a lack of feasibility assessments regarding its use for perioperative patient education in thoracic surgery.

Objective: This study aimed to assess the appropriateness and comprehensiveness of using ChatGPT for perioperative patient education in thoracic surgery in both English and Chinese contexts.

Methods: This pilot study was conducted in February 2023. A total of 37 questions focused on perioperative patient education in thoracic surgery were created based on guidelines and clinical experience. Two sets of inquiries were made to ChatGPT for each question, one in English and the other in Chinese. The responses generated by ChatGPT were evaluated separately by experienced thoracic surgical clinicians for appropriateness and comprehensiveness based on a hypothetical draft response to a patient's question on the electronic information platform. For a response to be qualified, it required at least 80% of reviewers to deem it appropriate and 50% to deem it comprehensive. Statistical analyses were performed using the unpaired chi-square test or Fisher exact test, with a significance level set at $P<.05$.

Results: The set of 37 commonly asked questions covered topics such as disease information, diagnostic procedures, perioperative complications, treatment measures, disease prevention, and perioperative care considerations. In both the English and Chinese contexts, 34 (92%) out of 37 responses were qualified in terms of both appropriateness and comprehensiveness. The remaining 3 (8%) responses were unqualified in these 2 contexts. The unqualified responses primarily involved the diagnosis of disease symptoms and surgical-related complications symptoms. The reasons for determining the responses as unqualified were similar in both contexts. There was no statistically significant difference (34/37, 92% vs 34/37, 92%; P=.99) in the qualification rate between the 2 language sets.

Conclusions: This pilot study demonstrates the potential feasibility of using ChatGPT for perioperative patient education in thoracic surgery in both English and Chinese contexts. ChatGPT is expected to enhance patient satisfaction, reduce anxiety, and improve compliance during the perioperative period. In the future, there will be remarkable potential application for using artificial intelligence, in conjunction with human review, for patient education and health consultation after patients have provided their informed consent.

(Interact J Med Res 2023;12:e46900) doi:10.2196/46900

KEYWORDS

patient education; ChatGPT; Generative Pre-trained Transformer; thoracic surgery; evaluation; patient; education; surgery; thoracic; language; language model; clinical workflow; artificial intelligence; AI; workflow; communication; feasibility

Introduction

The release of a dialogue-based artificial intelligence (AI) language model called ChatGPT (OpenAI) [1] has garnered global attention. ChatGPT is an advanced language model developed by OpenAI for generating human-like text responses and engaging in interactive conversations. It has been trained on a large corpus of internet text and has extensive applications in natural language understanding, question answering, language generation, and interactive dialogue. Several studies have documented the utilization of ChatGPT in the medical field, such as clinical decision assistance [2,3], medical document generation [4,5], and medical question answering [6-8]. ChatGPT demonstrates substantial potential in assisting health care professionals with real-time, web-based health consultations by providing patients with disease- or treatment-related knowledge and education. For example, Yeo et al [7] assessed the accuracy and reproducibility of ChatGPT in answering questions about cirrhosis and hepatocellular carcinoma and found that ChatGPT displayed extensive knowledge on cirrhosis (79.1% correct) and hepatocellular carcinoma (74% correct). Responses generated by ChatGPT regarding cardiovascular disease prevention queries were also graded as appropriate (21/25, 84%) in an exploratory study [8], demonstrating the potential of interactive AI to assist clinical workflows by augmenting patient education and patient-clinician communication.

Perioperative patient education is acknowledged as a critical component of thoracic surgical recovery. Enhancing patients' understanding of the general information of their disease, treatment plans, and recovery process has been shown to increase patient satisfaction, reduce undue anxiety, and increase their involvement in surgical recovery [9]. Until now, limited research has evaluated the use of ChatGPT for perioperative patient education in thoracic surgery. Moreover, most studies assessing the use of ChatGPT in the medical field have been conducted in English contexts. Considering that Chinese is also one of the most widely spoken languages worldwide, this study aimed to assess the appropriateness and comprehensiveness of using ChatGPT in perioperative patient education in both English and Chinese contexts.

Methods

This pilot study was conducted in February 2023. Following guideline-based topics [10] and clinical experience, 37 questions (Table 1) focused on perioperative thoracic surgery patient education were created. For each question, 2 inquiries were made to ChatGPT, one in English and the other in Chinese, and all responses were documented. The 2 sets of responses were evaluated separately in the following 2 aspects by thoracic surgical clinicians: appropriateness and comprehensiveness. The reviewers were composed of relevant practitioners with various years of experience in the field (Table 2). To ensure the reliability of the evaluation process, each response was independently assessed by multiple individuals. For appropriateness, a response was deemed "Y" (yes) if a hypothetical draft response would be considered appropriate when a patient asked the same question to a clinician on the electronic information platform, or "N" (no) if it was inappropriate. For comprehensiveness, a response was deemed "Y" (yes) if a hypothetical draft response would be considered comprehensive when a patient asked the same question to a clinician on the electronic information platform, or "N" (no) if it was incomprehensive. To be qualified, a response needed at least 80% of reviewers to deem it appropriate and 50% to deem it comprehensive. The response qualification criteria were established based on a consensus among clinical experts involved in the evaluation process. The reason for setting this criterion is that a qualified response requires a relatively higher level of appropriateness, as an inappropriate response can pose harm to patients. The unpaired chi-square test or Fisher exact test was used to assess differences in distributions between the categorical variables studied. All statistical analyses were performed using SPSS for Windows (version 23.0; IBM Corp). A 2-sided P value <.05 was considered significant. As the data collection process exclusively involved voluntary participation and did not involve any interventions, patient data, or sensitive personal information, ethics board approval was not applicable.

Table 1. Evaluation of the appropriateness and comprehensiveness of using ChatGPT for perioperative patient education in thoracic surgery in different language contexts (English and Chinese).

Question	Appropriateness, Y[a]			Comprehensiveness, Y		
	English (n=24), n (%)	Chinese (n=35), n (%)	*P* value	English (n=24), n (%)	Chinese (n=35), n (%)	*P* value
Q1: What is lung cancer?	24 (100)	32 (91)	.26	15 (62)	25 (71)	.47
Q2: What are the causes of lung cancer?	23 (96)	32 (91)	.64	22 (92)	30 (86)	.69
Q3: How can I prevent lung cancer?	21 (88)	32 (91)	.68	21 (88)	28 (80)	.51
Q4: What are the symptoms of lung cancer?	23 (96)	33 (94)	.99	21 (88)	31 (89)	.99
Q5: Why do some lung cancer patients develop hoarse voice as a symptom?	15 (62)	22 (63)	.99	11 (46)	19 (54)	.60
Q6: What diagnostic tests should be performed to diagnose lung cancer?	24 (100)	31 (89)	.14	19 (79)	33 (94)	.11
Q7: How can I determine if a lung nodule is benign or malignant?	23 (96)	33 (94)	.99	22 (92)	32 (91)	.99
Q8: What precautions should be taken prior to lung cancer surgery?	22 (92)	31 (89)	.99	19 (79)	31 (89)	.46
Q9: What are the complications that may arise from lung cancer surgery?	24 (100)	31 (89)	.14	22 (92)	28 (80)	.29
Q10: What is Mobocertinib?	24 (100)	32 (91)	.26	19 (79)	32 (91)	.25
Q11: What is Amivantamab-vmjw?	24 (100)	34 (97)	.99	23 (96)	33 (94)	.99
Q12: What is Adagrasib?	23 (96)	33 (94)	.99	21 (88)	32 (91)	.68
Q13: Do EGFR[b]-positive lung cancer patients who have received adjuvant chemotherapy also require adjuvant targeted therapy?	24 (100)	34 (97)	.99	21 (88)	32 (91)	.68
Q14: Is local treatment necessary for oligometastatic lung cancer?	24 (100)	33 (94)	.51	19 (79)	33 (94)	.11
Q15: Can Osimertinib be considered for EGFR-positive lung cancer patients with brain metastasis but without T790m mutation?	23 (96)	32 (91)	.64	22 (92)	32 (91)	.99
Q16: Why is lung cancer gene mutation testing necessary and who should undergo this testing?	23 (96)	31 (89)	.64	21 (88)	31 (89)	.99
Q17: What should I do if my CEA[c] level is found to be abnormal after 1 year of lung cancer surgery?	24 (100)	32 (91)	.26	24 (100)	33 (94)	.51
Q18: What is the cause of subcutaneous emphysema after lung cancer surgery and how can it be treated?	24 (100)	34 (97)	.99	21 (88)	31 (89)	.99
Q19: How can lung infections be prevented after lung cancer surgery?	21 (88)	30 (86)	.99	16 (67)	30 (86)	.11
Q20: How can the development of deep vein thrombosis be prevented after surgery?	23 (96)	29 (83)	.22	18 (75)	28 (80)	.75
Q21: What is the cause of an unpleasant odor from the surgical wound and how can it be treated?	24 (100)	34 (97)	.99	22 (92)	33 (94)	.56
Q22: What is a closed thoracic drainage tube and what precautions should be taken?	23 (96)	31 (89)	.64	19 (79)	30 (86)	.73
Q23: How often should lung cancer patients undergo follow-up exams and what tests should be performed?	24 (100)	34 (97)	.99	16 (67)	28 (80)	.36
Q24: What is esophageal cancer and its definition?	24 (100)	34 (97)	.99	22 (92)	33 (94)	.99
Q25: What are the causes of esophageal cancer?	23 (96)	31 (89)	.64	22 (92)	32 (91)	.99
Q26: How can one prevent the onset of esophageal cancer?	24 (100)	34 (97)	.99	24 (100)	33 (94)	.51
Q27: What are the symptoms of esophageal cancer?	24 (100)	29 (83)	.07	21 (88)	29 (83)	.73
Q28: What diagnostic tests should be performed to diagnose esophageal cancer?	23 (96)	32 (91)	.64	21 (88)	31 (89)	.99
Q29: What are the potential complications of esophageal cancer surgery?	22 (92)	32 (91)	.99	13 (54)	21 (60)	.79

Question	Appropriateness, Y[a]			Comprehensiveness, Y		
	English (n=24), n (%)	Chinese (n=35), n (%)	*P* value	English (n=24), n (%)	Chinese (n=35), n (%)	*P* value
Q30: How can a patient determine if targeted therapy is necessary for their esophageal cancer?	24 (100)	31 (89)	.14	18 (75)	31 (89)	.29
Q31: I experience hoarseness after surgery for esophageal cancer, what should I do?	12 (50)	20 (57)	.59	11 (46)	19 (54)	.60
Q32: If pleural fluid turns milky white in the chest tube following esophageal cancer surgery, what could be the cause and what steps should be taken?	9 (38)	15 (43)	.68	4 (17)	9 (26)	.53
Q33: What is anastomotic leak after esophageal cancer and how can it be managed?	22 (92)	32 (91)	.99	18 (75)	29 (83)	.46
Q34: What is Nivolumab?	23 (96)	31 (89)	.64	18 (75)	30 (86)	.33
Q35: Why is a jejunostomy tube used after esophageal cancer surgery and what precautions should be taken?	24 (100)	31 (89)	.14	18 (75)	30 (86)	.33
Q36: How can lung infection be prevented after esophageal cancer surgery	22 (92)	32 (91)	.99	15 (62)	30 (86)	.06
Q37: How often should individuals undergo follow-up after esophageal cancer surgery and what tests should be performed?	24 (100)	31 (89)	.14	24 (100)	31 (89)	.14

[a]Y: yes.

[b]EGFR: epidermal growth factor receptor.

[c]CEA: carcinoembryonic antigen.

Table 2. The thoracic surgery experience of the reviewers who assessed the responses generated by ChatGPT in English and Chinese language contexts.

Years of experience	English (n=24), n (%)	Chinese (n=35), n (%)
5-10	8 (33)	10 (29)
10-20	9 (38)	17 (48)
≥20	7 (29)	8 (23)

Results

A total of 35 reviewers participated in this study; 24 of these reviewers assessed the English responses, and all reviewers assessed the Chinese responses (Table 2). As shown in Table 1, of the 37 responses, 34 (92%) were qualified both in English and Chinese contexts, whereas the remaining 3 (8%) responses were unqualified in both contexts. The unqualified responses primarily focused on diagnosing disease symptoms and symptoms related to surgical complications. For example, in the case of hoarseness (Q5) in patients with lung cancer, there was a lack of consideration for the possibility of tumor or metastatic lymph node involvement of the recurrent laryngeal nerve. Similarly, responses about hoarseness after esophageal cancer surgery (Q31) failed to mention surgery-related recurrent laryngeal nerve injury, a common complication of the procedure. Additionally, responses regarding postoperative milky white pleural effusion after esophageal cancer surgery (Q32) omitted the description of surgery-related thoracic duct injury, which can lead to chyle leak. The reasons for determining the responses as unqualified in English and Chinese contexts were similar. Detailed information is listed in Multimedia Appendix 1. There was no statistically significant difference (34/37, 92% vs 34/37, 92%; P=.99) in the qualification rate between the 2 sets, indicating that ChatGPT has the potential to provide comparable quality of responses in English and Chinese contexts. Moreover, we ensured the reliability of the evaluation process by having all qualified and unqualified responses reevaluated and confirmed by 7 clinicians with over 20 years of experience in the field of thoracic surgery.

Discussion

ChatGPT achieved a satisfactory qualification rate (92%) in generating responses related to disease, diagnostic procedures, perioperative complications, treatment measures, disease prevention, and perioperative care considerations in both language contexts. This opens new avenues for enhancing patient education through AI-driven applications. ChatGPT is a versatile tool that might improve patient satisfaction, alleviate anxiety, increase compliance, and enhance the quality of clinical service in this setting. From a 24/7 availability standpoint, it is a convenient tool for users to obtain medical information at any time, thus reducing the communication costs between health care professionals and patients. These costs include time and, in certain cases, monetary expenses. By providing immediate access to information, ChatGPT saves time for both health care providers and patients and can potentially reduce expenses associated with traditional consultations or repetitive inquiries.

Our study also indicates a small portion responses generated by ChatGPT were unqualified (3/37, 8%). Consequently, the manual scrutiny of health care professionals remains necessary, particularly in instances involving the diagnosis and treatment of diseases or perioperative complications. Consistent with existing literature [11-13], our findings suggest the importance of considering the benefits and risks of using ChatGPT in the medical field. Additionally, evaluating ChatGPT in various language contexts provides valuable insights into its performance across diverse cultural and linguistic backgrounds. The comparable qualification rates demonstrate that ChatGPT is effective in supporting perioperative patient education for both English- and Chinese-speaking populations. This ensures that individuals who prefer or are more comfortable with either language can equally benefit from the AI-generated responses. In the future, there will be substantial prospects for the application of AI, combined with human review, in patient education and health consulting following the patients' signing of relevant informed consent documents.

Notably, the global prevalence of Chinese and English necessitates the testing of ChatGPT in less commonly spoken languages. In addition, perioperative patient education in thoracic surgery is a broad topic, and the 37 queries addressed in this research constitute only a fraction of it. The inclusion of reviewers with diverse working experience inevitably leads to heterogeneity in their opinions. However, by considering different perspectives, the evaluation process becomes more objective and less susceptible to personal preferences or preconceived notions. This reduces the potential for bias. Lastly, the study did not assess the concurrence between multiple responses given by ChatGPT for a single query.

In summary, the evaluation of clinicians on the generated responses from ChatGPT demonstrated the potential feasibility of using ChatGPT in both Chinese and English contexts to assist in patient education during the perioperative period of thoracic surgery. This study is expected to stimulate further dialogue and collaboration among patients, clinicians, and scholars, aiming to improve health care services while ensuring safety.

Acknowledgments

We express our gratitude to all the reviewers who participated in this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Detailed reasons for determining the responses as unqualified, and the questions and responses generated by ChatGPT in English.
[PDF File (Adobe PDF File), 1615 KB - ijmr_v12i1e46900_app1.pdf]

References

1. ChatGPT: optimizing language models for dialogue. OpenAI. URL: https://openai.com/blog/ChatGPT/ [accessed 2023-08-03]
2. Hirosawa T, Harada Y, Yokose M, Sakamoto T, Kawamura R, Shimizu T. Diagnostic accuracy of differential-diagnosis lists generated by Generative Pretrained Transformer 3 chatbot for clinical vignettes with common chief complaints: a pilot study. Int J Environ Res Public Health 2023 Feb 15;20(4):3378 [FREE Full text] [doi: 10.3390/ijerph20043378] [Medline: 36834073]
3. Liu S, Wright A, Patterson B, Wanderer J, Turer R, Nelson S, et al. Assessing the value of ChatGPT for clinical decision support optimization. medRxiv. Preprint posted on online on February 23, 2023. [FREE Full text] [doi: 10.1101/2023.02.21.23286254] [Medline: 36865144]
4. Ali SR, Dobbs TD, Hutchings HA, Whitaker IS. Using ChatGPT to write patient clinic letters. Lancet Digit Health 2023 Apr;5(4):e179-e181 [FREE Full text] [doi: 10.1016/S2589-7500(23)00048-1] [Medline: 36894409]
5. Cascella M, Montomoli J, Bellini V, Bignami E. Evaluating the feasibility of ChatGPT in healthcare: an analysis of multiple clinical and research scenarios. J Med Syst 2023 Mar 04;47(1):33 [FREE Full text] [doi: 10.1007/s10916-023-01925-4] [Medline: 36869927]
6. Grünebaum A, Chervenak J, Pollet SL, Katz A, Chervenak FA. The exciting potential for ChatGPT in obstetrics and gynecology. Am J Obstet Gynecol 2023 Jun;228(6):696-705. [doi: 10.1016/j.ajog.2023.03.009] [Medline: 36924907]
7. Yeo YH, Samaan JS, Ng WH. Correspondence on letter 2 regarding "assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma". Clin Mol Hepatol 2023 Jul;29(3):823-824 [FREE Full text] [doi: 10.3350/cmh.2023.0182] [Medline: 37254485]
8. Sarraju A, Bruemmer D, Van Iterson E, Cho L, Rodriguez F, Laffin L. Appropriateness of cardiovascular disease prevention recommendations obtained from a popular online chat-based artificial intelligence model. JAMA 2023 Mar 14;329(10):842-844. [doi: 10.1001/jama.2023.1044] [Medline: 36735264]
9. Whyte RI, Grant PD. Preoperative patient education in thoracic surgery. Thorac Surg Clin 2005 May;15(2):195-201. [doi: 10.1016/j.thorsurg.2005.02.002] [Medline: 15999517]

10. Ettinger DS, Wood DE, Aisner DL, Akerley W, Bauman JR, Bharat A, et al. NCCN Guidelines® Insights: Non-Small Cell Lung Cancer, Version 2.2023. J Natl Compr Canc Netw 2023 Apr;21(4):340-350. [doi: 10.6004/jnccn.2023.0020] [Medline: 37015337]
11. Liu J, Wang C, Liu S. Utility of ChatGPT in clinical practice. J Med Internet Res 2023 Jun 28;25:e48568 [FREE Full text] [doi: 10.2196/48568] [Medline: 37379067]
12. Johnson SB, King AJ, Warner EL, Aneja S, Kann BH, Bylund CL. Using ChatGPT to evaluate cancer myths and misconceptions: artificial intelligence and cancer information. JNCI Cancer Spectr 2023 Mar 01;7(2):pkad015 [FREE Full text] [doi: 10.1093/jncics/pkad015] [Medline: 36929393]
13. Sedaghat S. Early applications of ChatGPT in medical practice, education and research. Clin Med (Lond) 2023 May 21;23(3):278-279. [doi: 10.7861/clinmed.2023-0078] [Medline: 37085182]

Abbreviations

AI: artificial intelligence

Edited by T Leung, T de Azevedo Cardoso; submitted 01.03.23; peer-reviewed by L Zhuang, R Marshall; comments to author 15.06.23; revised version received 22.07.23; accepted 27.07.23; published 14.08.23.

Please cite as:
Shao CY, Li H, Liu XL, Li C, Yang LQ, Zhang YJ, Luo J, Zhao J
Appropriateness and Comprehensiveness of Using ChatGPT for Perioperative Patient Education in Thoracic Surgery in Different Language Contexts: Survey Study
Interact J Med Res 2023;12:e46900
URL: https://www.i-jmr.org/2023/1/e46900
doi:10.2196/46900
PMID:37578819

Original Paper

Outsourcing the Management of Reusable Medical Devices in a Chain-Wide Care Setting: Mixed Methods Feasibility Study

Bart A C Noort[1], PhD; Paul Buijs[1], IR, PhD; Oskar Roemeling[2], PhD

[1]Department of Operations, Faculty of Economics and Business, University of Groningen, Groningen, Netherlands

[2]Department of Innovation Management & Strategy, Faculty of Economics and Business, University of Groningen, Groningen, Netherlands

Corresponding Author:
Bart A C Noort, PhD
Department of Operations
Faculty of Economics and Business
University of Groningen
Nettelbosje 2
Groningen, 9747 AE
Netherlands
Phone: 31 631986866
Email: a.c.noort@rug.nl

Abstract

Background: Managing reusable medical devices incurs substantial health care costs and complexity, particularly in integrated care settings. This complexity hampers care quality, safety, and costs. Studying logistical innovations within integrated care can provide insights to medical devices use among staff effectively.

Objective: This study aimed to establish the feasibility of a logistical intervention through outsourcing and a web portal. The goal was to provide insights into users' acceptability of the intervention, on whether the intervention was successfully implemented, and on the intervention's preliminary efficacy, thus benefiting practitioners and researchers.

Methods: This paper presents a mixed methods feasibility study at a large chain-wide health care provider in the Netherlands. The intervention entailed outsourcing noncritical reusable medical devices and introducing a web portal for device management. A questionnaire gauged perceived ordering and delivery times, satisfaction with the ordering and delivery process, compliance with safety and hygiene certification, and effects on the care delivery process. Qualitative data in the form of observations, documentation, and interviews were used to identify implementing challenges. Using on-site stocktaking and data from information systems, we analyzed the utilization, costs, and rental time of medical devices before and after the intervention for wheelchairs and anti–pressure ulcer mattresses.

Results: Looking at the acceptability of the intervention, a high user satisfaction with the ordering and delivery process was reported (rated on a 5-point Likert scale). With respect to preliminary efficacy, we noted a reduction in the utilization of wheelchairs (on average, 1106, SD 106 fewer utilization d/mo), and a halted increase in the utilization of anti–pressure ulcer mattresses. In addition, nurses who used the web portal reported shorter ordering times for wheelchairs (−2.7 min) and anti–pressure ulcer mattresses (−3.1 min), as well as shorter delivery times for wheelchairs (−0.5 d). Moreover, an increase in device certification was reported (average score of 1.9, SD 1.0), indicating higher levels of safety and hygiene standards. In theory, these improvements should translate into better outcomes in terms of costs and the quality of care. However, we were unable to establish a reduction in total care costs or a reduced rental time per device. Furthermore, respondents did not identify improvements in safety or the quality of care. Although implementation challenges related to the diverse supply base and complexities with different care financers were observed, the overall implementation of the intervention was considered successful.

Conclusions: This study confirms the feasibility of our intervention, in terms of acceptability, implementation success, and preliminary efficacy. The integrated management of medical devices should enable a reduction in costs, required devices, and material waste, as well as higher quality care. However, several challenges remain related to the implementation of such interventions.

(Interact J Med Res 2023;12:e41409) doi:10.2196/41409

KEYWORDS
health care logistics; outsourcing; web ordering portal; medical devices; feasibility study; device management

Introduction

Background

Managing medical devices is a complex but important element in the delivery of high-quality care and has considerable effect on the financial bottom line of health care providers [1,2]. If not managed well, there can be various downsides, such as health care organizations holding too many devices in relation to demand [3,4], physicians needing to wait for delivery of the devices they need urgently [5], nurses possibly not having enough time available for the direct care process [6], and using devices that do not comply with safety and hygiene standards [7]. Hence, health care organizations spend considerable time and resources on seeking solutions to mitigate these problems, such as opting for joint procurement, applying lean manufacturing principles, using tracking technologies, and outsourcing [2,8]. In this paper, we are especially interested in outsourcing as a potential route to better medical device management.

Achieving better management of medical devices is particularly challenging when one considers that health care providers are still working toward more integrated care delivery [9]. Although such integrated care developments seem promising, operational and organizational challenges are often encountered when operating across care disciplines [10,11]; for example, patients may appreciate being able to retain the same wheelchair when moving between care locations, but this creates additional logistics or administrative tasks for nursing staff. Furthermore, the stringent requirements on the traceability and safety of medical devices under the 2021 European Union (EU) Medical Device Regulation challenge manufacturers and care providers to improve their medical device management [12]. The outsourcing of logistics activities may address some of these challenges by reducing coordination and planning efforts and improving their results [13,14]. Whether such outsourcing solutions, in this case related to managing medical devices, would be perceived as acceptable by nursing staff and indeed improve care delivery need to be carefully assessed.

When pursuing integrated care delivery, supportive IT becomes pivotal, including in relation to medical device management [15]. Care professionals generally have a positive attitude toward the implementation of IT [16]. However, its implementation and adoption are complicated by the multistakeholder context involving patients, medical staff, and other organizational staff [17]. Previous studies have underlined the importance of the usability of technology in gaining the support of core medical staff, thereby contributing to a successful implementation and increasing the likelihood of a positive effect on care delivery and other outcomes [18,19]. In this study, we focused on the feasibility of using a web portal aimed at enhancing medical device management through outsourcing.

Research Aims

We conducted a mixed methods study to assess the feasibility of a logistical intervention, where a health care provider outsourced several noncritical reusable medical devices to a third-party provider and introduced a web portal to support medical device management. Feasibility is based on the extent to which an intervention is considered acceptable by the users; whether the intervention is successfully implemented; and, finally, the intervention's preliminary efficacy. Our aims with regard to outsourcing were to reduce the (1) total utilization days of the medical devices; (2) delivery time; and (3) time and money spent on cleaning, maintaining, and internally transporting the devices. We aimed to achieve our goals by using a web portal that supports medical staff by reducing the time taken to order a device when needed and to deregister it when no longer needed and that provides them with easy insight into device availability from their own location. With this study, we provide practitioners and researchers with knowledge on the potential benefits of, and challenges related to, outsourcing medical devices, assisted by IT, in an integrated care provider setting.

Methods

Research Context

The study took place between January 2019 and June 2022 in the context of a large regional health care provider in the Netherlands. The organization, which was established after a merger of a hospital, several nursing homes, and home care providers, consists of 1 hospital, 17 nursing homes, and 3 residential care centers; it also provides home-based care at several locations. Since the merger, the health care provider has moved toward becoming a more integrated care organization (eg, by implementing an organization-wide electronic medical record system). It was the first such organization to receive chain-wide accreditation worldwide from the Joint Commission International [20].

The Dutch health care system is organized around 3 domains, largely based on the principles of *managed competition* [21]. The curative care domain is regulated by the Health Insurance Act. Private health insurers compete for beneficiaries, whereas private not-for-profit health care providers compete for patients. The long-term care domain is regulated under the Dutch Long-term Care Act, with a national tax-based budget. Budgeting responsibility for long-term care is regionally executed by so-called care offices. The home care domain falls under the responsibility of local government, regulated by the Social Support Act. Given that the studied health care provider is active in all 3 care domains—hospital, residential, and home care—it has contractual agreements with all regional financing bodies: health insurers, care offices, and local government. This leads to a complex setting where the health care provider is required to determine which medical device is associated with which financing body and to understand the conditions that govern the use of these specific devices. Within this health care context, we focused on the management of noncritical reusable medical devices. Reusable medical devices are defined as those that health care providers can reprocess and reuse for >1 patient [22]. Typically, reusable medical devices require calibration, maintenance, repair, user training, cleaning, and decommissioning [23]. In our research, noncritical devices are typified by the fact that they are not invasive, such as wheelchairs and patient lifts. Critical or semicritical medical devices, such as surgical instruments or equipment for diagnostics [22], are beyond the scope of our research.

Before the intervention, on-site stocktaking showed that several medical devices were unused, and there was little insight into the actual stock held. Approximately 40% of the devices physically identified were not registered, whereas several devices registered as available on-site were untraceable. Exploratory meetings with staff and on-site observations indicated that the existing way of managing devices led to considerable time spent searching by employees, patients having to wait, and high inventory and management costs.

Our focus on reusable medical devices provides a setting demarcated by relatively low complexity in which to study how to structurally benefit from logistical solutions such as outsourcing and web ordering portals in a chain-wide provider.

Intervention

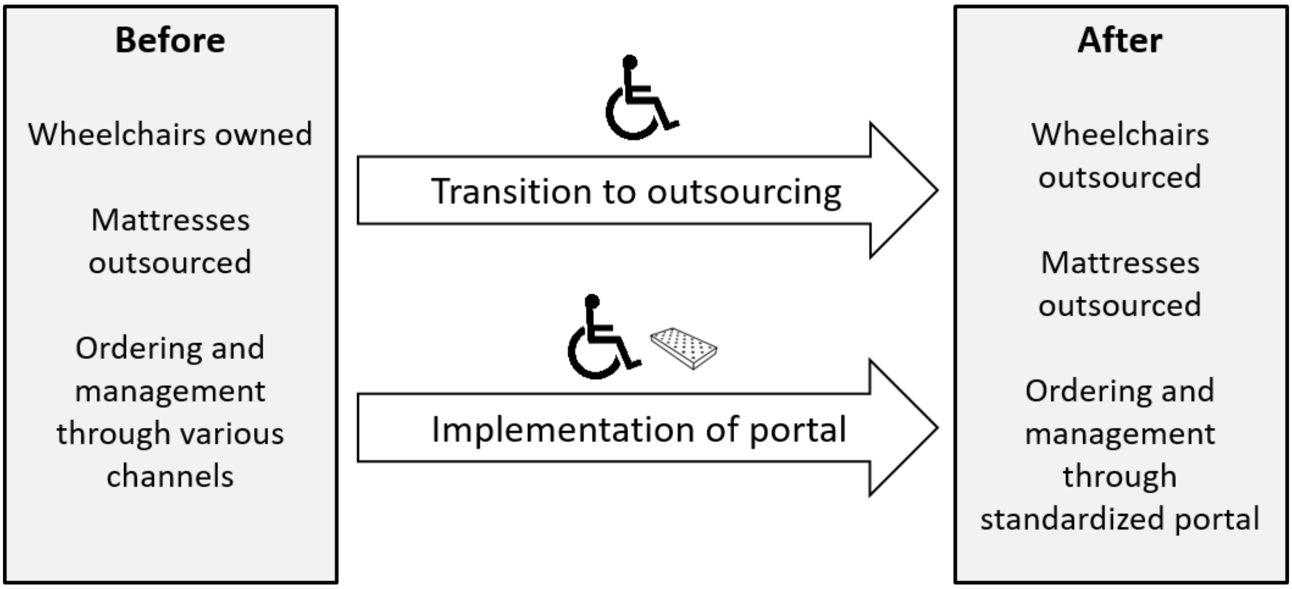
The intervention consisted of 2 parts. The first part involved a transition to the full outsourcing of 8 types of medical devices: anti–pressure ulcer mattresses, bariatric care beds, bed trapezes, lifting slings, low-low beds, patient lifts, standard wheelchairs, and shower stretchers. Of these, bed trapezes, lifting slings, low-low beds, shower stretchers, and standard wheelchairs were largely (>90%) managed and owned by the health care provider before the intervention. Anti–pressure ulcer mattresses, bariatric care beds, and patient lifts were already outsourced for most care locations (>90% of the devices were outsourced). From April 2021 onward, all 8 device types have been fully outsourced. Since then, the health care provider has been paying a rental fee per device per day to a third-party medical device supplier that transports, stores, cleans, and maintains the devices. The second part of the intervention involved implementing a web portal for nurses to support and simplify the process of registering and deregistering an increased number of rented medical devices (implemented on September 1, 2021). Before the intervention, nurses ordered medical devices via telephone or email, either via the internal purchasing department or directly through an external device supplier. The portal provides a standardized system for ordering these devices and gives nursing staff insight into the number of devices ordered and the devices currently available on-site. Overall, the intervention aimed to unify the management of noncritical reusable medical devices across the entire organization.

Study Setup

This study can be typified as a mixed methods feasibility study. It is based on several data sources: medical device rental data, interviews, on-site observations, and a small-sample questionnaire. The study has a before-and-after design and addresses 3 of the typical goals of feasibility studies, as classified by Bowen et al [24]: whether the outsourcing of reusable medical devices and the introduction of a web ordering portal are acceptable, implementable, and effective.

In this paper, we focused on 2 types of devices: wheelchairs and anti–pressure ulcer mattresses. Each device type was exposed to different aspects of the intervention during our study: wheelchairs were undergoing the transition to outsourcing and the implementation of the web portal, whereas, for anti–pressure ulcer mattresses, the web portal was the main change in managing the devices. A comparison between these 2 types of devices thus enables a feasibility assessment of the transition to outsourcing separately from the implementation of the web portal. A schematic overview of the intervention is shown in Figure 1.

Figure 1. Schematic overview of the intervention: transition from owning to outsourcing wheelchairs and the implementation of a web portal for ordering wheelchairs and anti–pressure ulcer mattresses.



Web Portal Data Collection

Stocktaking and Rental Data

To establish the utilization days of the devices managed and owned by the health care provider before the intervention, device stocktaking was undertaken at all care locations in 2019. The aim was to establish (1) whether devices that were shown in the enterprise resource planning (ERP) system were present on-site and whether devices on-site were shown in the ERP system, (2) whether devices had received maintenance within the required period, and (3) whether devices were ready for use in terms of cleanliness and functioning. The utilization days of outsourced devices were determined based on rental data from the third-party supplier that showed the number of newly ordered devices and the rental time per device for each care location. Medical devices used for providing home care are not included in our analysis because, under the Dutch Long-term Care Act, they are not the responsibility of the care provider.

Interviews and On-Site Observations

We collected qualitative data through on-site observations and interviews to understand the supply and management of noncritical reusable medical devices; to explore areas for improving delivery time, patient care, cost reduction, and nursing staff satisfaction; and to reveal potential challenges that may occur when implementing these changes. Available supporting documentation was also analyzed. This data collection also helped explain how the transition to outsourcing and the implementation of the web portal came into effect. During 2 rounds of interviews (2019 and 2021), a total of 30 interviews carried out in Dutch were held with internal and external stakeholders to provide deeper insight into challenges regarding the management of noncritical reusable medical devices. The interviewed participants were nursing staff, team leaders, care location managers, logistics managers, and care purchasers from municipalities and health insurers. The interview guide (translated into English) can be found in Multimedia Appendix 1.

User Experience and Care Quality Questionnaire

User satisfaction related to the intervention was evaluated by means of an anonymous questionnaire (in Dutch). After the intervention, between April 28, 2022, and June 8, 2022, after an informal announcement by the project leader of the care provider, nurses from all care locations were approached by email and invited to complete a questionnaire. Questions related to the ease of use of the web portal, ordering and delivery times for devices, and the perceived consequences of the intervention in terms of the quality of care and patient satisfaction. Each question had responses using a 5-point Likert scale ranging from 1=*fully agree* to 5=*fully disagree*. The 2 types of devices, wheelchairs and anti–pressure ulcer mattresses, were used as examples in the questionnaire. The questionnaire (translated into English) can be found in Multimedia Appendix 2.

An overview of the data collection and timing is provided in Table 1.

Table 1. An overview of the used data sources, linked to the different study phases, as well as study objectives.

Data source	Phase of data collection	Objective
Stocktaking	Before the intervention	• Preliminary efficacy
Rental data	Before and after the intervention	• Preliminary efficacy
Interviews and observations	Before and after the intervention	• Acceptability • Implementation success
Questionnaire	After the intervention	• Acceptability • Preliminary efficacy

Data Analysis

The utilization days of owned (wheelchairs) and rented (anti–pressure ulcer mattresses) devices before the intervention were compared with the situation after the intervention using stacked charts, showing the utilization of medical devices in days per month per care location. The stocktaking data provided the number of owned and rented devices present at each care location before the intervention. All the various device brands and models were clustered into basic types (wheelchairs or anti–pressure ulcer mattresses) to ensure comparability before and after the intervention. The number of owned devices was converted to utilization in days per month. Average rental time per newly rented device was calculated and analyzed for 4 care locations (rental data only). Total device utilization and average utilization time were analyzed by means of 2 linear regression models (Excel 2016 [Microsoft Corporation]). Table 2 provides an overview of the dependent and independent variables, which were defined before the analysis. Autocorrelation between outsourcing and the portal was checked based on variance inflation factors and was <3 in all models.

The questionnaire data were analyzed by distinguishing three user groups: (1) nurses who used the web portal for both wheelchairs and anti–pressure ulcer mattresses, (2) nurses who used the web portal for only 1 type of device, and (3) nurses who had never used the web portal or had stopped using it. The questionnaire responses were analyzed by comparing the average difference between reported ordering time and delivery time before and after the intervention and on the average score per item on questions regarding ordering satisfaction and the perceived effects of the intervention. Differences in average ordering satisfaction and perceived effects among the user groups were assessed by means of a 2-tailed *t* test (Excel 2016). Differences in the average reported ordering time and delivery time before and after the intervention were assessed by means of a paired 2-tailed *t* test.

The interview data were analyzed with the aim to thoroughly understand the challenges that emerged from the transition to outsourcing and the implementation of the web portal—and how the care provider dealt with these challenges—as well as to highlight the challenges that the intervention did not resolve. Led by the first author, we performed the analysis using ATLAS.ti software (ATLAS.ti Scientific Software Development GmbH) and Excel, following the 3-step grounded theory approach of Corbin and Strauss [25] as discussed in the study by Karlsson [26]. This process follows a cycle where categories are inductively formed from the data, followed by a deductive analysis where findings are refined and validated based on existing concepts. In the first step, open coding, the transcripts were summarized and short text codes assigned. In the second step, axial coding, we established more general categories related to typical characteristics, challenges, enabling conditions, and consequences related to medical device use within the health care organization. In the third step, selective coding, 2 typical patterns emerged that seemed to have an effect on the implementation process of the portal and the outsourcing contract. The first pattern related to the scale of uptake of the intervention, leading to ambiguity in the organization. The second pattern concerned the effect of market incentives, hampering suppliers in offering a wide product portfolio. The coding process and intermediate outcomes were frequently discussed between the first and second authors to achieve consensus in categorization and established patterns.

Table 2. An overview of the linear regression models and included variables.

Model	Dependent variable	Independent variables
1a	Total device utilization (wheelchairs)	Time (mo), web portal (dummy), and outsourcing (dummy)
1b	Total device utilization (anti–pressure ulcer mattresses)	Time (mo) and web portal (dummy)
2a	Average device utilization time (wheelchairs)	Time (mo), web portal (dummy), outsourcing (dummy), and location
2b	Average device utilization time (anti–pressure ulcer mattresses)	Time (mo), web portal (dummy), and location

Ethical Considerations

The data regarding rental of medical devices were routinely collected by the health care provider for administrative purposes and not linked to individual patients. The data processed in this study are not considered to constitute medical research involving human participants as defined in the 1964 Helsinki declaration. As such, the statutes of the institutional review board of the faculty of economics and business of the University of Groningen indicate that the study does not require ethics approval by a review committee. Nevertheless, all the methods described in this study were conducted in accordance with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

All interview and questionnaire respondents were informed about the purpose of the data collection and the data processing procedure, and they consented to participate.

Results

Overview

The results are presented in 3 parts. In the first part, we evaluated the acceptability of the intervention by nursing staff. In the second part, we addressed to what extent the intervention was successfully implemented and what enabling or limiting factors have been established. Finally, in the third part, we offered a preliminary assessment of the efficacy of the intervention in terms of device utilization, costs, ordering and delivery time, quality of care, and compliance with safety and hygiene standards.

Acceptability

Questionnaire Responses

User satisfaction, as measured in the questionnaire responses, indicates how the nursing staff rate the new ordering process that is based on outsourcing and the web portal. Nurses (N=45) who were responsible for ordering reusable medical devices completed the questionnaire. Of the 45 nurses, 28 (62%) reported using the web portal for ordering both wheelchairs and anti–pressure ulcer mattresses, 10 (22%) used the web portal only for anti–pressure ulcer mattresses and not for wheelchairs, and 7 (16%) reported not using the web portal and sticking with the earlier procedure of telephoning and emailing.

Satisfaction With the Ordering Process

User satisfaction with the ordering process, with specific attention to the use of the web portal, was generally rated highly (Table 3). Nurses who ordered both types of devices through the web portal reported an average satisfaction score of 1.6 (SD 0.6; using a Likert scale ranging from 1 to 5, with 1=*most satisfied* and 5=*least satisfied*). Nearly half (24/45, 53%) of the respondents fully agreed with the statement that the ordering process is easy. Nurses who only ordered anti–pressure ulcer mattresses via the web portal reported an average satisfaction score of 2.3 (SD 0.6), which was a statistically significant (P=.003) lower rating than that of nurses who ordered both types of equipment via the web portal. Respondents who did not order both types of equipment via the web portal (17/45, 38%) reported either being unaware of the web portal, unsure of how to use it, or dissatisfied with its functionality.

Table 3. Satisfaction with the web portal ordering process (rated using a 5-point Likert scale ranging from 1=fully agree to 5=fully disagree)[a].

	Ordering is easy, mean (SD)	Ordering is clear, mean (SD)	Ordering is fast, mean (SD)	The portal is accessible, mean (SD)	The portal is stable, mean (SD)	The portal is user friendly, mean (SD)	Score, mean (SD)
Ordering via the web portal (n=28)	1.7 (1.0)	1.8 (1.1)	1.6 (0.9)	1.6 (0.7)	1.5 (0.7)	1.7 (0.8)	1.6 (0.6)
Ordering mattress via the web portal and wheelchair via email or telephone (n=10)	1.8 (0.9)	1.9 (0.9)	1.9 (1.0)	1.7 (0.9)	1.8 (0.9)	2.0 (0.7)	2.3 (0.6)
Average	1.7 (0.9)	1.8 (1.1)	1.7 (0.9)	1.6 (0.8)	1.6 (0.8)	1.8 (0.8)	1.8 (0.7)

[a]Overall satisfaction with ordering via web portal versus satisfaction with ordering mattress via portal: *P*=.003 (average scores, based on a 2-tailed *t* test with unequal variances).

Satisfaction With Device Management

With respect to user satisfaction with medical device management in general, the *insight into availability* dimension was rated the highest, with an average rating of 1.8 (SD 1.1; Table 4). Of the 45 respondents, 27 (60%) fully agreed with the statement that there is now a better insight into available devices. Whether there was an improvement in the product range was rated with an average score of 2.1 (SD 1.0). Perceived increase in the ease of transferring devices between locations was rated the lowest (2.4, SD 1.3), which is a score suggesting only a minor improvement. No statistically significant differences between users or (partial) nonusers of the web portal were found.

Table 4. Overall satisfaction with the intervention (rated using a 5-point Likert scale ranging from 1=fully agree to 5=fully disagree)[a].

	Product range is larger, mean (SD)	Better insight into available devices, mean (SD)	Easier to transfer devices between care locations, mean (SD)	Score, mean (SD)
Ordering via the web portal (n=28)	2.1 (1.0)	1.7 (1.1)	2.1 (1.3)	2.0 (0.9)
Ordering mattress via the web portal and wheelchair via email or telephone (n=10)				
Wheelchair	2.1 (1.2)	1.7 (1.0)	2.3 (1.0)	2.0 (0.9)
Mattress	2.0 (0.9)	1.3 (0.7)	3.0 (1.5)	2.1 (0.8)
Ordering via a different method (n=7)	2.1 (1.1)	2.4 (1.1)	2.0 (1.2)	2.2 (0.9)
Average	2.1 (1.0)	1.8 (1.1)	2.4 (1.3)	2.0 (0.9)

[a]Overall satisfaction with ordering via web portal versus satisfaction with ordering mattress via portal (wheelchair): *P*=.92, overall satisfaction with ordering via web portal versus satisfaction with ordering mattress via portal (mattress): *P*=.80, overall satisfaction with ordering via web portal versus satisfaction with ordering via different method: *P*=.63, overall satisfaction with ordering mattress via web portal (wheelchair) versus satisfaction with ordering mattress via web portal (mattress): *P*=.91, overall satisfaction with ordering mattress via web portal (wheelchair) versus satisfaction with ordering via different method: *P*=.74, and overall satisfaction with ordering mattress via web portal (mattress) versus satisfaction with ordering via different method: *P*=.81 (average scores, based on a 2-tailed *t* test with unequal variances).

Implementation

Overview

The acceptability assessment in the previous subsection indicates generally positive findings, and the implementation of the intervention was successful for all 8 types of medical devices. Nevertheless, evaluating the implementation of the intervention also highlighted a number of challenges. The limited uptake of the web portal by nurses (31/45, 69%) shows that there is room for improvement. On the basis of the observational and interview data, we explain in the following subsections how the implementation of the intervention is hampered by the complex context of the care market in which this chain-wide care provider is embedded.

Uptake of the Improved Ordering Process

Although most of the nurses reported that they do follow the new ordering process, several mentioned that they still use other ordering methods, such as ordering via email or telephone or by ordering from other suppliers. To an extent, this is inevitable in the current situation because some of the more specialized devices, such as customized wheelchairs, are only available from other suppliers and not from the one involved in the intervention. A problem is that this then leads to ambiguity about what ordering process should be used for which type of device, thereby structurally hampering a full uptake of the intervention. This is exemplified by several remarks in response to the questionnaire:

> *I do not order wheelchairs via [supplier of study] but via the occupational therapy department, who arrange it with another supplier.* [Head nurse 1]

> *Besides working as a nurse, I also work as an occupational therapy coach, so I order devices via email.* [Head nurse 1]

In addition, nurses reported that there are still many device types where the internal ordering process is experienced as too cumbersome and slow:

> *I particularly have problems with the purchasing process when something is not available from [supplier of study]...I need to go through the purchasing and infection prevention department, the whole process sometimes takes 4 to 6 weeks. The patient may have already passed away by then.* [Occupational health nurse 2]

These quotations illustrate that nursing staff will only fully adapt to, and benefit from, the new ordering process when it is implemented on an organization-wide scale, with a larger range of device types available. However, given the diverse supply base as well as the various care financers involved, this seems challenging.

Effect of a Complex Market

At the outset of our study, we expected the chain-wide health care provider in this study to have a unique opportunity to simplify the management of reusable medical devices; namely, because the provider is active across the 3 care domains in the Netherlands (ie, curative care, long-term care, and social care), we expected that it would be able to pool devices across these domains, thereby reducing its device supply base and the number of ordering methods. However, in practice, device suppliers tend to specialize their service and product range to a specific care domain, thereby forming a barrier to fulfilling our initial expectations:

> *It is partly a very practical issue. Device suppliers have contracts with a selection of health care providers and device manufacturers...When it comes to maintenance, some spare parts can only be supplied by a specific manufacturer, which does not always align with the device type demanded by the health care provider. At the same time, our goal is—and it will be the same for [name of care provider]—to deliver affordable and personalized care for clients, and that sometimes creates tensions.* [Policy advisor of municipality]

Furthermore, the care path of a single patient can span multiple domains. In this situation, the ordering methods and device suppliers ordinarily change during the care process. Consequently, to meet individual patient needs and to prevent long waiting times, nursing staff put much effort and time into the process of ordering devices. Essentially, they take care of the service and coordination duties that should partly be the responsibility of device suppliers and care financers:

> *When a patient goes home after rehabilitation, we sometimes sell them our bed trapeze; otherwise, the patient has to order it, and it may not arrive in time. Also, we may order a wheelchair for short-term use under the Social Care Act from the supplier. If a patient needs longer-term care, we also order wheelchairs under the Long-term Care Act or other devices via the care insurer...So, unfortunately, we put quite some time and effort into coordinating the device process for patients.* [Occupational health nurse 1]

These findings show that the multitude of stakeholders and processes that are part of health care markets limit the extent to which nursing staff benefit from the transition to outsourcing and the implementation of the web portal and hamper a structural implementation of such solutions.

Preliminary Efficacy

On the basis of comparing the use of wheelchairs and anti–pressure ulcer mattresses before and after the intervention, we now present the findings of our preliminary efficacy analysis.

Utilization of Reusable Medical Devices

Figure 2 is based on our analysis of the utilization data and shows that wheelchair utilization decreased significantly (P=.03) by 1106 (SD 106) days per month on average (January 2019 to March 2021: n=5079, May 2021 to January 2022: n=3972). This reduction in total rental time amounts to the most clearly visible pattern across the 8 devices included in the intervention. For the other types of devices, such as lifting slings and low-low beds, we also observed considerable reductions in utilization days, albeit with a less consistent drop. The total number of utilization days of anti–pressure ulcer mattresses gradually increased over time (Figure 3). On the basis of on-site meetings, this increase can be explained by an increase in the number of patients with a high care burden. After the implementation of the web portal, this rising pattern stabilized, which was confirmed by the regression model (P<.001).

As indicated in Figures 4 and 5, the average rental time per rented device per month did not change over time for either wheelchairs or anti–pressure ulcer mattresses after the intervention. We observed similar patterns for the other types of devices.

Figure 2. Utilization days per month for wheelchairs. As of May 2021, wheelchairs were fully delivered on the basis of the outsourcing contract. Before this date, 90% were locally owned by care departments. Color tones relate to different care locations. Outcomes of linear regression model 1a: outsourcing: *P*=.03, portal: *P*=.82, month: *P*=.50.

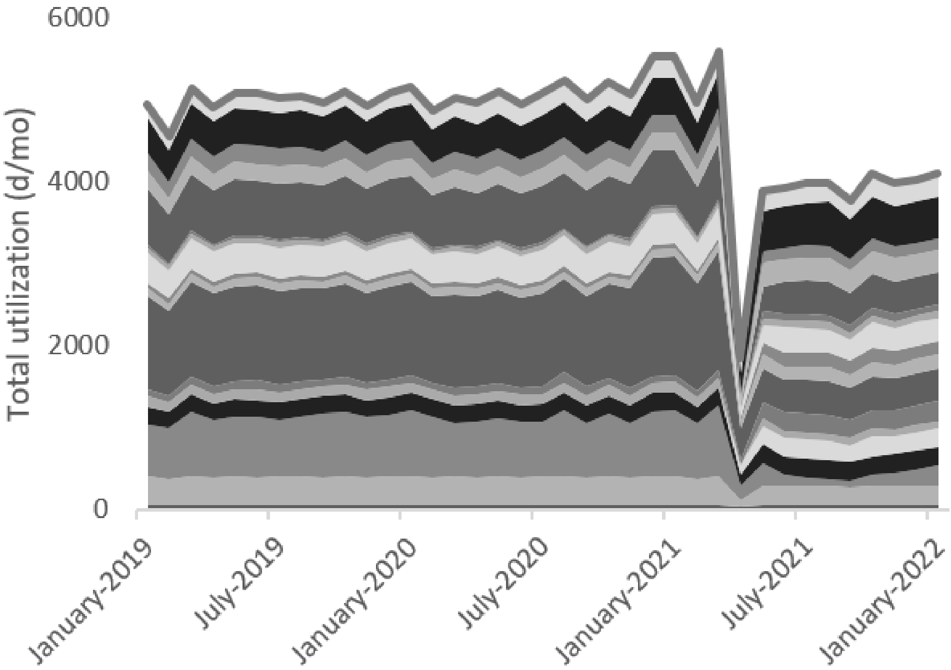


Figure 3. Utilization days per month for anti–pressure ulcer mattresses. Anti–pressure ulcer mattresses were delivered on an outsourcing basis during the entire period shown in the graph. Color tones relate to different care locations. Outcomes of linear regression model 1b: portal: *P*<.001, month: *P*<.001.

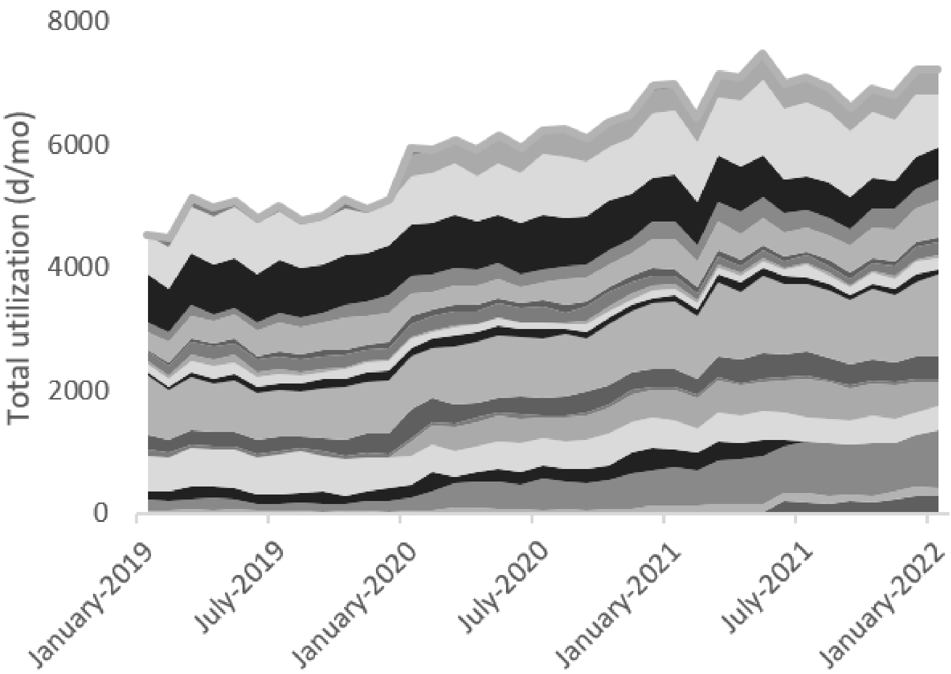


Figure 4. Average rental time per rented device per month of wheelchairs for 4 care locations. Outcomes of linear regression model 2a: outsourcing: *P*=.11, portal: *P*=.91, month: *P*=.95, location: *P*<.001.

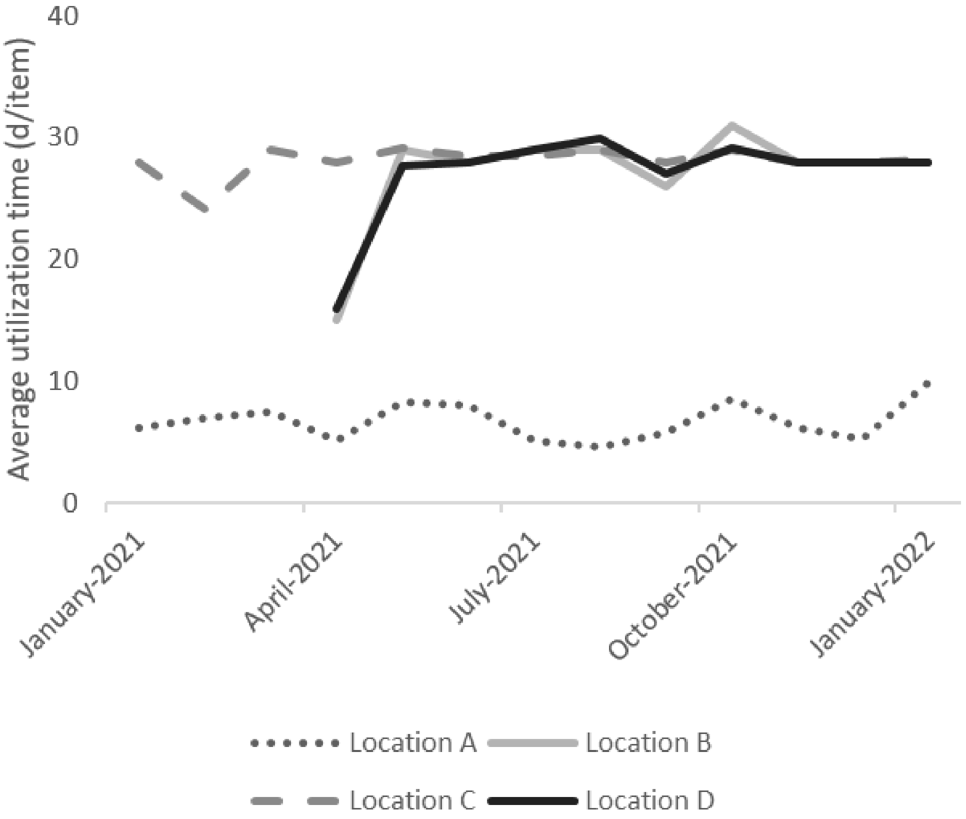
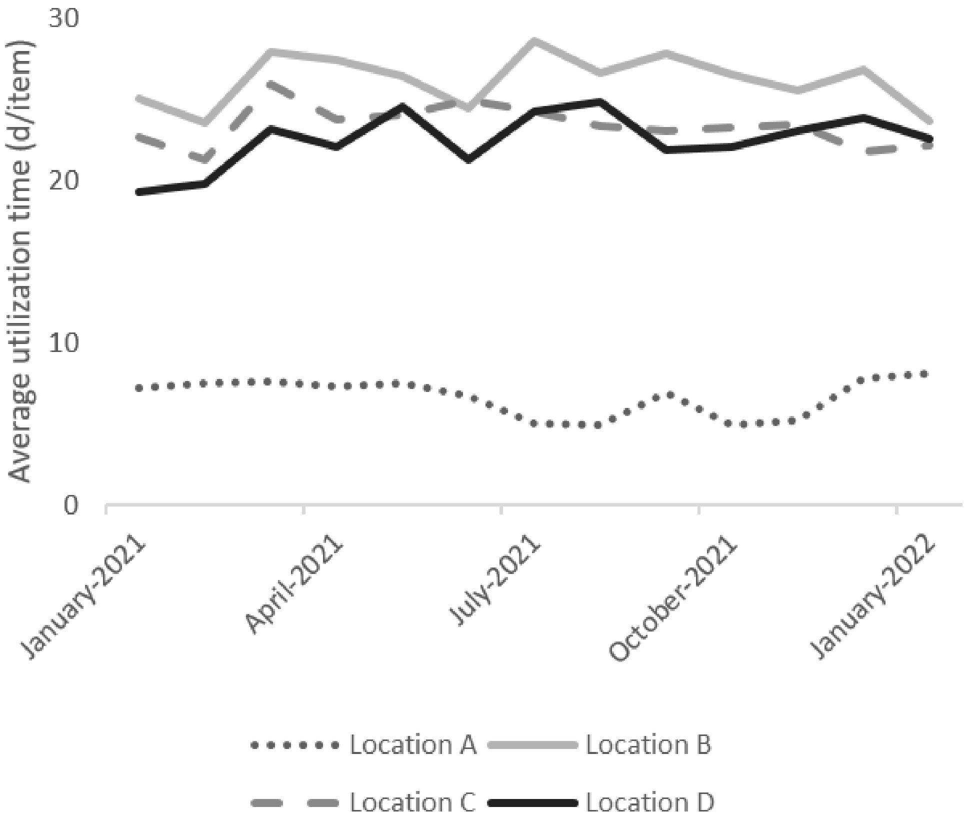


Figure 5. Average rental time per rented device per month of anti–pressure ulcer mattresses for 4 care locations. Outcomes of linear regression model 2b: portal: *P*=.71, month: *P*=.73, location: *P*<.001.



Costs of Renting Medical Equipment

The cost of renting anti–pressure ulcer mattresses was on average US $32,381 per month after the intervention compared with US $25,445 per month before the intervention. The cost of renting wheelchairs was on average US $4501 per month after the intervention. No accurate data were available on ownership costs for wheelchairs, including the actual purchase price and the maintenance, procurement, storage, and material-handling costs. In addition, for both types of devices, there were no data available on the historical patterns of care delivery volumes and the disease burden of patients. Hence, a valid comparison of costs before and after the intervention was deemed infeasible by the researchers involved as well as the managers at the care provider under study.

Ordering and Delivery Time

The questionnaire results show how nurses perceived the ordering and delivery times before and after the intervention. Nurses who ordered both types of devices via the web portal reported statistically significant reduced ordering times for both wheelchairs (*P*=.04) and anti–pressure ulcer mattresses (*P*=.03) and statistically significant reduced delivery times for wheelchairs only (*P*=.01). Nurses who only ordered anti–pressure ulcer mattresses via the web portal and ordered wheelchairs via other channels reported reduced ordering times, albeit not statistically significant, for both devices (*P*=.35 and *P*=.26). These results are summarized in Table 5. Gauging the perceived changes in delivery time, nurses rated the item *devices are available faster* with a score of 2.2 (SD 1.1; using a 5-point Likert scale ranging from 1=*fully agree* to 5=*fully disagree*), indicating some improvement.

Table 5. Reported change in ordering (min) and delivery (d) times for wheelchairs and anti–pressure ulcer mattresses after implementation of the web portal[a].

	Ordering time (min)				Delivery time (d)			
	Before, mean (SD)	After, mean (SD)	δ	*P* value	Before, mean (SD)	After, mean (SD)	δ	*P* value
Ordering via the web portal (both devices; n=28)								
Wheelchair	9.9 (7.4)	7.2 (3.1)	−2.7	.04	3.8 (2.1)	3.3 (1.8)	−0.5	.01
Mattress	10.7 (7.9)	7.6 (3.7)	−3.1	.03	2.4 (0.9)	2.1 (0.5)	−0.3	.07
Ordering mattress via the portal and wheelchair via email or telephone (n=10)								
Wheelchair	6.8 (6.4)	5.5 (3.9)	−1.3	.35	4.4 (1.9)	4.6 (1.8)	0.2	.59
Mattress	8.2 (8.4)	6.0 (3.4)	−2.2	.26	2.6 (0.9)	2.5 (1.2)	−0.1	.34

[a]*P* values are based on a paired 2-tailed *t* test between ordering time and delivery time before and after the intervention.

Quality of Care and Compliance With Safety and Hygiene Standards

Compliance with safety and hygiene standards was measured by responses to the statement *there are fewer uncertified devices at the location*. This item was rated positively by nurses who ordered both device types via the web portal with a score of 1.9 (SD 1.0; using a 5-point Likert scale ranging from 1=*fully agree* to 5=*fully disagree*). Of the 45 respondents, 22 (49%) fully agreed that there were fewer uncertified devices.

On the basis of the averaged questionnaire responses (Table 6), nurses did not report any difference in perceived patient satisfaction, care quality, safety, and care outcomes when comparing the old and new ordering processes for both device types.

Table 6. Assessment of the intervention with respect to the quality of care (rated using a 5-point Likert scale ranging from 1=fully agree to 5=fully disagree)[a].

	Patients are more satisfied, mean (SD)	Care quality is higher, mean (SD)	Care is safer, mean (SD)	Better care outcomes, mean (SD)	Score, mean (SD)
Ordering via the web portal (n=28)	2.3 (1.0)	2.3 (1.2)	2.3 (1.2)	2.3 (1.2)	2.3 (1.1)
Ordering mattress via the portal and wheelchair via email or telephone (n=10)					
Wheelchair	2.6 (1.0)	2.3 (0.9)	2.6 (0.9)	2.3 (0.9)	2.4 (0.9)
Mattress	2.4 (0.9)	2.1 (0.8)	1.9 (0.8)	2.0 (0.8)	2.1 (0.8)
Ordering via a different method (n=7)	2.4 (1.4)	2.1 (1.1)	2.3 (1.3)	2.4 (1.5)	2.3 (1.3)
Average	2.4 (1.0)	2.2 (1.0)	2.3 (1.1)	2.3 (1.1)	2.3 (1.0)

[a]Effects of ordering via web portal versus effects of ordering mattress via web portal (wheelchair): *P*=.72, effects of ordering via web portal versus effects of ordering mattress via web portal (mattress): *P*=.53, effects of ordering via web portal versus effects of ordering via different method: *P*=.99, effects of ordering mattress via web portal (wheelchair) versus effects of ordering mattress via web portal (mattress): *P*=.39, effects of ordering mattress via web portal (wheelchair) versus effects of ordering via different method: *P*=.83, and effects of ordering mattress via web portal (mattress) versus effects of ordering via different method: *P*=.69 (average scores, based on a 2-tailed *t* test with unequal variances).

Discussion

Principal Findings

This study addresses the feasibility of a 2-fold logistical intervention aimed at outsourcing the management of reusable medical devices and the introduction of a web portal facilitating the ordering and use of these devices. The findings show that, with respect to acceptability, user satisfaction with the ordering and delivery process was high. Concerning preliminary efficacy, a reduction in the required number of wheelchairs, which was significantly related to outsourcing (*P*=.03), was observed. The increase in the required number of anti–pressure ulcer mattresses stabilized after the introduction of the web portal. For both wheelchairs and anti–pressure ulcer mattresses, we found shorter reported ordering and delivery times, but rental times per device did not reduce. Hence, the web portal may support the ordering and delivery processes, but there is no indication that it triggers a more efficient use of devices. After the intervention, a higher degree of device certification was reported, thereby increasing compliance with safety and hygiene standards. In theory, these improvements should translate into better outcomes in terms of costs and the quality of care. However, based on the findings, we could not establish that a reduction in overall costs had been achieved, and nurses did not report improvements in safety and the quality of care. Although, for all 8 types of reported medical devices, the implementation of the intervention was successful, based on observations and interviews, several challenges were highlighted related to improving the chain-wide management of medical devices. These challenges relate to the diverse nature of the supply base and complexities with having multiple care financers. In essence, these findings show the difficulties of managing, and catering to, the various stakeholder interests involved in the care chain, and this will be more elaborately reflected upon in the following subsection.

Comparison With Previous Research

Improving Health Care Logistics Performance

The literature on health care logistics suggests that better management of medical devices—for example, through better forecasting, standardized replenishment, and delivery procedures—supported by materials management information systems can lead to a reduction in device stock and increase the occupancy rate of devices [2,7]. In terms of the outcomes of the care process, quality can be improved by reducing device stock-outs and delivery disruption. Hence, it is not surprising that, for health care managers, anticipated cost savings and quality improvements are important drivers when considering new ways of managing medical devices [27,28]. To a certain extent, our findings concur with the outcomes of these previous studies. The ordering process, device utilization, and compliance with safety and hygiene standards certainly improved after the outsourcing of reusable medical devices and the implementation of a web portal for ordering these devices. Indeed, the web portal is an important precondition in that it supports the ordering process and increases the measurability of device management. Nevertheless, we also observed that improvements related to reduced costs and the increased quality of care are not necessarily as straightforward as often assumed. We discuss potential reasons for this discrepancy in the next subsection.

Measurability of Costs and Quality of Care

We recognize that there is a lot of freedom in deciding which expenses and savings to include in comparing the outcomes of an intervention, and, as a result, establishing changes in overall costs is especially challenging [1]. Our case and the potential cost savings of the intervention provide a good example. The health care provider we studied did not account for the human resource costs related to the devices it owned (eg, costs related to maintenance and storage). Moreover, the provider did not take service delivery costs into account when evaluating the expenditure on devices. Similarly, an increase in expenditure on devices, as we observed for anti–pressure ulcer mattresses, may be misinterpreted as a rise in the costs of device management, rather than, at least in part in our case, an increase in patient volume and care burden. More generally, health care providers often have to deal with low-visibility and low-quality data [2,7]. Although providers increasingly work with resource planning systems (eg, ERP) and medical record information systems, and they are increasingly standardizing care service products, such as diagnosis-related treatment combinations, the transparency of costs in the health care sector remains an issue [29,30].

With respect to the quality of care, it would be highly beneficial if health care providers were able to show positive effects from better device management. Our questionnaire indicated that nursing staff did perceive an increase in the speed of device delivery and improvements in terms of compliance with safety and hygiene standards. We would expect this to contribute to the satisfaction of patients by reducing waiting time and increasing the available capacity, that is, the amount of time nurses have available for the direct care process. However, indicators related to patient satisfaction and safety remain hard to monitor because quality inspections are usually based on samples and periodic inspections. This exemplifies the usefulness of using patient-reported outcomes and experience measures [31,32], as well as linking such patient reports with clinical health records [33]. Such improved transparency should be considered not only for meeting external accountability demands but also for being able to demonstrate the effective rollout of improvements internally.

Structurally Embedding Organizational Changes

Our findings, supported by insights from previous studies [7,29-33], illustrate how transparent care processes and outcomes are key preconditions for structurally improving chain-wide care delivery and organizational changes. Because of various hidden costs and limited quality performance metrics, it was challenging for managers to show the true costs related to owning, as against renting, medical devices. Hence, it was difficult to justify the rental costs involved with outsourcing where all the costs (hidden or open) are part of the rental fee. At the same time, nursing staff noted that the intervention was rolled out on a relatively small scale in terms of the number of included types of equipment, leading to a confusing mixture of both new and old ordering procedures and complex coordination. This resulted in a situation where nursing staff felt that they were receiving insufficient top-down support, whereas, at the same time, top management perceived limited success from the intervention because not all nursing staff adhered to the new procedures. However, ultimately, it is likely that the studied intervention does lead to quality improvement and cost savings, yet it remains difficult to measure such improvements objectively.

Reflecting on the studied case enables us to provide recommendations for practice. Given the recently increased EU standards regarding the traceability and safety of medical devices [12], the presented intervention is of importance for medical device manufacturers, suppliers, and care providers. When looking at the dynamic circumstances of daily care delivery, one cannot expect manufacturers to achieve full compliance with standards by themselves. Hence, we recommend that medical device suppliers and care providers work jointly to improve medical device use in terms of efficiency, quality, and safety through outsourcing agreements supported with IT for care personnel. Moreover, by ensuring traceability and certification, these practices may also support the accessibility of innovative medical devices, another area that seems to be becoming more complex owing to the new EU standards [34,35].

Nevertheless, it seems that realizing a broad rollout of such an intervention and obtaining benefits in terms of cost savings and quality improvements remain a challenge, particularly for chain-wide care providers. This is especially the case because they are dealing with an extensive and complex device supply base. Top management needs to be aware of the related issues and challenges, particularly in a setting of care delivery that crosses the traditional boundaries of curative, long-term, and home care domains. For practice to structurally benefit from solutions such as outsourcing and web ordering portals, it is recommended to aim at reducing the complexity of the supply base, improving communication and support toward nursing staff, and establishing clear performance measures for evaluation purposes.

Limitations

This study assessed the feasibility of an intervention based on qualitative and quantitative data collected before and after the intervention. Several limitations warrant mention. First, although our data indicate clear improvements over time, there are several potentially confounding factors that have not been considered explicitly; for example, although care registration data did not indicate considerable changes in patient numbers or disease burden, this was not part of the statistical analyses. Other examples include developments in treatment methods that may affect device use as well as disruptions in the supply and care process, such as those seen during the COVID-19 pandemic.

Second, our study took place against the backdrop of the introduction of the 2021 EU Medical Device Regulation. This context places our results within a framework of heightened standards for the traceability and safety of medical devices and shows a glimpse of the complex landscape that health care providers must navigate within the newly imposed regulatory constraints.

Third and last, the questionnaire on user experience that provided relevant insights that add to the basic device data was sent out only after the intervention, and it was targeted at nursing staff and not at patients. As such, we lack firsthand experiences from the latter group. This is especially relevant because our respondents did not report any clear effects on patient satisfaction or any effect on care quality, which provides an important direction for further study. This again stresses the importance of measuring patient experiences and outcomes over time for both care and research purposes.

Conclusions

The integrated management of medical devices should lead to reduced costs and fewer required devices, higher quality of care, and reduced material waste. This feasibility study confirms this potential when it comes to acceptability, implementation success, and the preliminary efficacy of the presented intervention. Nevertheless, at a time when the integration of care chains is becoming more important, our research also highlights some of the difficulties when putting integration into practice; for example, in a health care system context that is based on market principles, it remains an immense challenge to achieve a more coherent way of managing medical devices, even for a single care provider when operating across traditional health care boundaries. In addition, the ongoing challenges in achieving transparency on prices and on the quality of care again prove key to measuring the efficacy of integrated medical device management which, in turn, is pivotal to achieving long-term implementation.

Acknowledgments

This work was supported by the Dutch Institute for Advanced Logistics (439.18.306). The authors thank Romée Janson, Iris van der Meulen, Erik Bakker, and Maaike Tjeerdsma for their contributions to the data collection and analysis during the early stages of this study. The authors also thank Pieter Vlot and Björn van Nuland for their support in setting up and executing this study and for their advice regarding the daily practices of reusable medical device management.

Data Availability

The data sets generated and analyzed during this study are not publicly available owing to confidentiality issues and agreements made with the case sites but are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guide.
[DOCX File , 17 KB - ijmr_v12i1e41409_app1.docx]

Multimedia Appendix 2
User experience and care quality questionnaire.
[DOCX File , 26 KB - ijmr_v12i1e41409_app2.docx]

References

1. Abdulsalam Y, Schneller E. Hospital supply expenses: an important ingredient in health services research. Med Care Res Rev 2019 Apr;76(2):240-252. [doi: 10.1177/1077558717719928] [Medline: 29148349]
2. Volland J, Fügener A, Schoenfelder J, Brunner JO. Material logistics in hospitals: a literature review. Omega 2017 Jun;69:82-101. [doi: 10.1016/j.omega.2016.08.004]
3. Landry S, Beaulieu M. The challenges of hospital supply chain management, from central stores to nursing units. In: Denton B, editor. Handbook of Healthcare Operations Management. New York, NY: Springer; Jan 10, 2013.
4. Rosales CR, Magazine M, Rao U. The 2Bin system for controlling medical supplies at point-of-use. Eur J Oper Res 2015 May 16;243(1):271-280. [doi: 10.1016/j.ejor.2014.10.041]
5. Ahmadi E, Masel DT, Metcalf AY, Schuller K. Inventory management of surgical supplies and sterile instruments in hospitals: a literature review. Health Syst (Basingstoke) 2019;8(2):134-151 [FREE Full text] [doi: 10.1080/20476965.2018.1496875] [Medline: 31275574]
6. Bélanger V, Beaulieu M, Landry S, Morales P. Where to locate medical supplies in nursing units: an exploratory study. Supply Chain Forum Int J 2018 Feb 02;19(1):81-89. [doi: 10.1080/16258312.2018.1433438]
7. Moons K, Waeyenbergh G, Pintelon L. Measuring the logistics performance of internal hospital supply chains – a literature study. Omega 2019 Jan;82:205-217. [doi: 10.1016/j.omega.2018.01.007]
8. Zhu X, Tao Y, Zhu R, Wu D, Ming WK. Impact of hospital characteristics and governance structure on the adoption of tracking technologies for clinical and supply chain use: longitudinal study of US hospitals. J Med Internet Res 2022 May 26;24(5):e33742 [FREE Full text] [doi: 10.2196/33742] [Medline: 35617002]
9. Hughes G, Shaw SE, Greenhalgh T. Rethinking integrated care: a systematic hermeneutic review of the literature on integrated care strategies and concepts. Milbank Q 2020 Jun;98(2):446-492 [FREE Full text] [doi: 10.1111/1468-0009.12459] [Medline: 32436330]
10. Ling T, Brereton L, Conklin A, Newbould J, Roland M. Barriers and facilitators to integrating care: experiences from the English Integrated Care Pilots. Int J Integr Care 2012 Jul 27;12(5):e129 [FREE Full text] [doi: 10.5334/ijic.982] [Medline: 23593044]
11. Auschra C. Barriers to the integration of care in inter-organisational settings: a literature review. Int J Integr Care 2018 Jan 16;18(1):5 [FREE Full text] [doi: 10.5334/ijic.3068] [Medline: 29632455]

12. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC. Official Journal of the European Union. 2017 May 5. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2017:117:TOC [accessed 2023-01-17]
13. Azzi A, Persona A, Sgarbossa F, Bonin M. Drug inventory management and distribution: outsourcing logistics to third‐party providers. Strateg Outsourcing Int J 2013;6(1):48-64. [doi: 10.1108/17538291311316063]
14. Iannone R, Lambiase A, Miranda S, Riemma S, Sarno D. Pulling drugs along the supply chain: centralization of hospitals' inventory. Int J Eng Bus Manag 2014 Jan 01;6(21):1-11. [doi: 10.5772/58939]
15. Øvretveit J. Digital technologies supporting person-centered integrated care - a perspective. Int J Integr Care 2017 Sep 25;17(4):6 [FREE Full text] [doi: 10.5334/ijic.3051] [Medline: 29588629]
16. Maassen O, Fritsch S, Gantner J, Deffge S, Kunze J, Marx G, et al. Future mobile device usage, requirements, and expectations of physicians in German University hospitals: web-based survey. J Med Internet Res 2020 Dec 21;22(12):e23955 [FREE Full text] [doi: 10.2196/23955] [Medline: 33346735]
17. Gammon D, Berntsen GK, Koricho AT, Sygna K, Ruland C. The chronic care model and technological research and innovation: a scoping review at the crossroads. J Med Internet Res 2015 Feb 06;17(2):e25 [FREE Full text] [doi: 10.2196/jmir.3547] [Medline: 25677200]
18. Kirkley D, Stein M. Nurses and clinical technology: sources of resistance and strategies for acceptance. Nurs Econ 2004;22(4):216-22, 195. [Medline: 15382401]
19. Tawfik DS, Sinha A, Bayati M, Adair KC, Shanafelt TD, Sexton JB, et al. Frustration with technology and its relation to emotional exhaustion among health care workers: cross-sectional observational study. J Med Internet Res 2021 Jul 06;23(7):e26817 [FREE Full text] [doi: 10.2196/26817] [Medline: 34255674]
20. Enterprise accreditation. Joint Commission International. URL: https://www.jointcommissioninternational.org/accreditation/accreditation-programs/enterprise-accreditation/ [accessed 2023-01-05]
21. Enthoven AC, van de Ven WP. Going Dutch--managed-competition health insurance in the Netherlands. N Engl J Med 2007 Dec 13;357(24):2421-2423. [doi: 10.1056/NEJMp078199] [Medline: 18077805]
22. What are reusable medical devices? U.S. Food & Drug Administration. 2018 Dec 1. URL: https://www.fda.gov/medical-devices/reprocessing-reusable-medical-devices/what-are-reusable-medical-devices [accessed 2022-07-05]
23. Decontamination and reprocessing of medical devices for health-care facilities. World Health Organization. 2016 Sep 5. URL: https://www.who.int/publications/i/item/9789241549851 [accessed 2022-07-05]
24. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]
25. Corbin JM, Strauss A. Grounded theory research: procedures, canons, and evaluative criteria. Qual Sociol 1990;13(1):3-21. [doi: 10.1007/bf00988593]
26. Karlsson C. Research Methods for Operations Management. London, UK: Routledge; 2016.
27. Skipworth H, Delbufalo E, Mena C. Logistics and procurement outsourcing in the healthcare sector: a comparative analysis. Eur Manag J 2020 Jun;38(3):518-532. [doi: 10.1016/j.emj.2020.04.002]
28. Young S, Macinati MS. Health outsourcing/backsourcing. Public Manag Rev 2012 Sep;14(6):771-794. [doi: 10.1080/14719037.2011.642627]
29. Jacobs JC, Barnett PG. Emergent challenges in determining costs for economic evaluations. Pharmacoeconomics 2017 Feb 12;35(2):129-139. [doi: 10.1007/s40273-016-0465-1] [Medline: 27838912]
30. Herzlinger RE. How to solve the cost crisis in health care. Harv Bus Rev 2011 Nov;89(11):22, discussion 23. [Medline: 22111427]
31. Basch E, Barbera L, Kerrigan CL, Velikova G. Implementation of patient-reported outcomes in routine medical care. Am Soc Clin Oncol Educ Book 2018 May;38:122-134. [doi: 10.1200/edbk_200383]
32. Hjollund NH, Valderas JM, Kyte D, Calvert MJ. Health data processes: a framework for analyzing and discussing efficient use and reuse of health data with a focus on patient-reported outcome measures. J Med Internet Res 2019 May 21;21(5):e12412 [FREE Full text] [doi: 10.2196/12412] [Medline: 31115347]
33. Hawley S, Yu J, Bogetic N, Potapova N, Wakefield C, Thompson M, et al. Digitization of measurement-based care pathways in mental health through REDCap and electronic health record integration: development and usability study. J Med Internet Res 2021 May 20;23(5):e25656 [FREE Full text] [doi: 10.2196/25656] [Medline: 34014169]
34. Behan R, Pandit A, Watson M. New EU medical device regulations: impact on the MedTech sector. Med Writ 2017;26(2):20-24 [FREE Full text]
35. Maresova P, Hajek L, Krejcar O, Storek M, Kuca K. New regulations on medical devices in Europe: are they an opportunity for growth? Adm Sci 2020 Mar 12;10(1):16. [doi: 10.3390/admsci10010016]

Abbreviations

ERP: enterprise resource planning
EU: European Union

Edited by T de Azevedo Cardoso; submitted 28.07.22; peer-reviewed by E van der Velde, M Popovic; comments to author 28.12.22; revised version received 18.01.23; accepted 11.08.23; published 19.09.23.

Please cite as:
Noort BAC, Buijs P, Roemeling O
Outsourcing the Management of Reusable Medical Devices in a Chain-Wide Care Setting: Mixed Methods Feasibility Study
Interact J Med Res 2023;12:e41409
URL: https://www.i-jmr.org/2023/1/e41409
doi:10.2196/41409
PMID:37725420

Original Paper

Development of a Core Set of Quality Criteria for Virtual Reality Applications Designed for Older Adults: Multistep Qualitative Study

Alina Napetschnig[1], PhD; Klara Brixius[2], Prof Dr; Wolfgang Deiters[1], Prof Dr

[1]Department of Community Health, Hochschule für Gesundheit Bochum, Bochum, Germany

[2]Institut für Kreislaufforschung und Sportmedizin, Deutsche Sporthochschule Köln, Cologne, Germany

Corresponding Author:
Alina Napetschnig, PhD
Department of Community Health
Hochschule für Gesundheit Bochum
Gesundheitscampus 6-8
Bochum, 44801
Germany
Phone: 49 234 77727 764
Email: alina.napetschnig@gmail.com

Abstract

Background: Virtual reality (VR) applications are gaining growing significance, particularly among older adults. These applications can provide valuable support to older adults by offering immersive VR content that positively influences various aspects of their daily lives, including activities of daily living. Furthermore, VR applications can contribute to the enhancement of cognitive and motor skills, ultimately leading to an improved quality of life for older individuals. Nevertheless, to ensure a positive impact, it is crucial to develop VR experiences that are tailored to the needs and preferences of the users.

Objective: This study aims to develop a core set of quality criteria and guidelines for the development of user-centered VR applications specifically designed for older adults (target group).

Methods: The multistep qualitative study design comprised several key stages, beginning with a systematic literature search. This was followed by a framework analysis aimed at identifying a core set of criteria. Subsequently, these criteria underwent validation through expert workshops. The outcomes achieved through this iterative process were organized and categorized into criteria, accompanied by explanations detailing the underlying categories or codes.

Results: The quality criteria core set for older adults–friendly VR applications has been developed through an iterative process. It is divided into 2 distinct parts, each containing criteria categorized into specific areas. The first part includes the following categories: (1) quality assurance of medical/health content, (2) data protection provisions, (3) quality requirements, (4) consumer protection, and (5) interoperability. The second part includes the following categories: (1) graphic/quality, (2) 3D character/avatar, (3) providing in-game instructions and prompts, (4) interaction, (5) navigation, and (6) promotion of user motivation and loyalty to use. The results imply a differentiated scope as well as a differentiated granularity of the criteria.

Conclusions: Considering the ongoing advancement of VR technology and the diverse needs within the older adult demographic, it is essential to assess the quality criteria core set results on an individual basis.

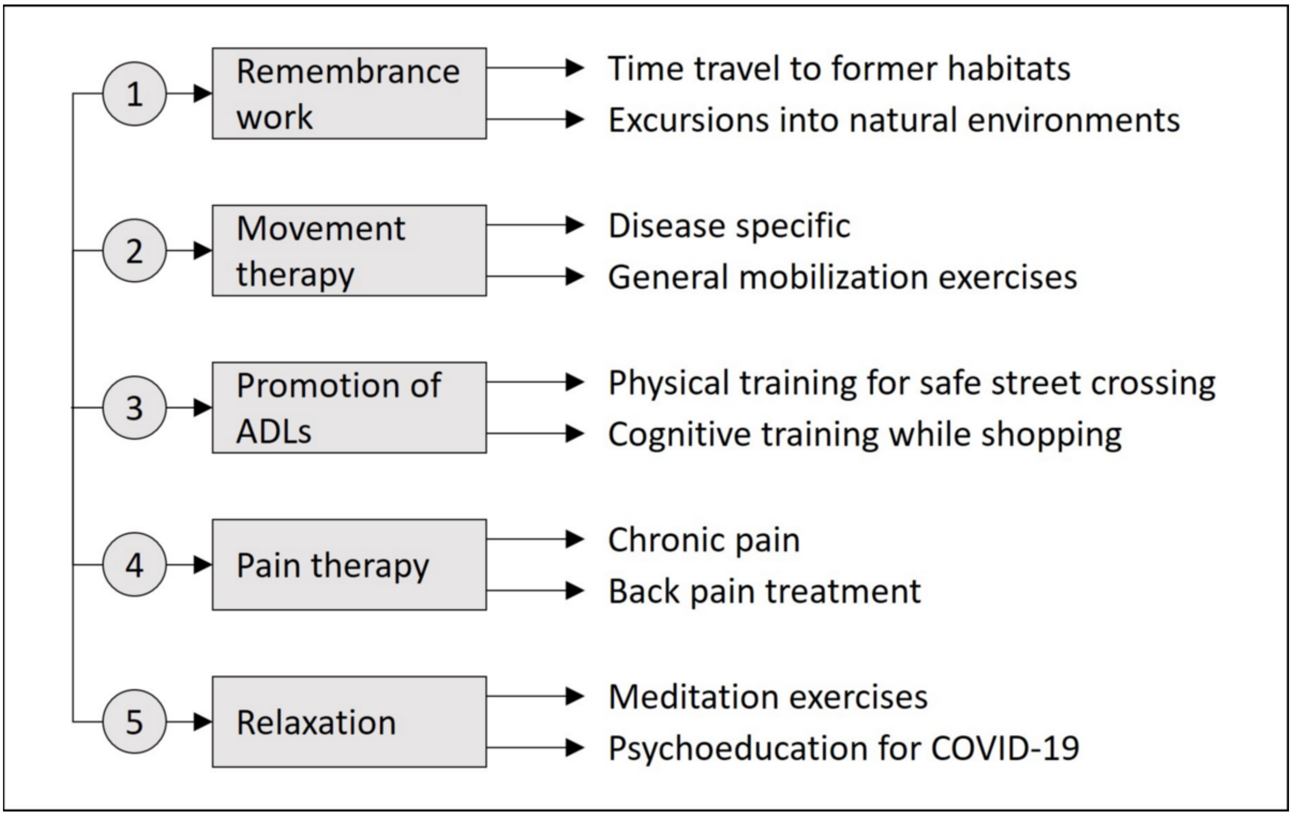
(Interact J Med Res 2023;12:e45433) doi:10.2196/45433

KEYWORDS
virtual reality; older adults; quality criteria; user-centered

Introduction

Virtual reality (VR) goggles are a technology that has been gaining increasing popularity in recent years, and their functionality continues to expand [1]. VR goggles are a technical medium that allows immersion in VR [2]. VR stands for computer-generated environments that enable people to experience and interact in digital worlds [3]. VR applications have found their place in various fields. While they are already a popular choice for leisure, especially among the younger generation, they are increasingly employed in health care, often in the form of "serious games," to address specific human conditions [1,4]. VR applications offer numerous benefits, including use among older adults [2,5,6]. VR applications are intended to specifically promote the competencies of the older adults and to help them achieve more independence in old age [7]. Figure 1 illustrates various application fields in VR for older adults, along with 2 existing exemplary scenarios.

Figure 1. Fields of application and exemplary scenarios in the context of virtual reality and seniors. ADLs: activities of daily living.



Although many interventions have been shown to improve the health status of older adults through the use of VR, there remains a need for research to determine usability and acceptance factors [8]. Furthermore, attention should be paid to target group-specific development [9]. Currently, VR applications in the health care context are typically developed through collaboration between VR developers or computer scientists on the technical side and medical staff or health scientists on the content side [10]. The challenge stems from the unfamiliarity between VR developers and medical personnel, hindering the establishment of comprehensive VR development requirements. Additionally, user-centered development is increasingly complex due to the growing heterogeneity in age-related factors among older adults [6,7,11,12].

VR development for older adults should promote healthy aging and enhance quality of life. The World Health Organization (WHO) describes healthy aging as a process of developing and maintaining the functional capacity that enables well-being in old age [13]. Under the aspect of "healthy aging," the WHO has defined various influencing factors that can have a positive effect on the aging process. These include factors that go beyond the elimination of diseases. As a result, active health promotion throughout one's life and targeted support measures to maintain functional capacity in old age are essential. The term health, in the context of healthy aging, is a person's ability to perform or pursue those things that they prioritize [14]. To fulfill these goals as much as possible, the needs of older adults must be taken into account. At the same time, the feasibility of continuous development of VR must be addressed [15]. Guidelines and best practice frameworks can assist in using VR applications effectively for specific goals [16]. Guidelines developed through collaborative research with experts in VR development and older adult–focused VR applications can serve as an initial orientation guide for tailored and beneficial VR applications for older adults.

This study aimed to establish a core set of quality criteria for older adult–friendly VR applications, which could serve as guidelines for similar target-oriented VR development. It involved considering criteria relevant to older adult–oriented VR application development.

Methods

Overview

Three different approaches were combined in this study. First, a systematic literature review search (SLS) was performed on April 7, 2021, to obtain an overview of existing guidelines or development recommendations for VR applications. The results mainly considered studies on VR application development and technology development guides providing recommendations. Second, the results were clustered in a framework analysis to systematize criteria or categories. Finally, in the third step, expert workshops were conducted with VR development experts and VR application experts with reference to the target group of older adults to check or adapt the validity of the elaborated results.

Systematic Literature Search

To align with the study's objectives, we formulated the following research question: *What universal criteria exist for development recommendations for VR applications and what are their contents?*

The following 3 main keyword groups were examined for the development of the search strategies:

1. (universal) criteria and corresponding synonyms (eg, features, characteristics, features, requirements, quality criteria);
2. development recommendations and corresponding synonyms (eg, development tips, hints, advice, suggestions, design, framework, evaluation); and
3. VR (applications) and corresponding synonyms (eg, virtual system, VR goggles, VR technology, VR head-mounted display, VR headset).

For the 3 groups of keywords, the thesaurus was also reviewed. Second, a search strategy for each database was developed. The detailed strategy for PubMed is shown in Textbox 1. Specific search strategies are outlined in Table 1.

Textbox 1. Search string used for the PubMed search.

(("criteria s"[All Fields] OR "criterias"[All Fields] OR "standards"[MeSH Subheading] OR "standards"[All Fields] OR "criteria"[All Fields]) AND (("develop"[All Fields] OR "develope"[All Fields] OR "developed"[All Fields] OR "developer"[All Fields] OR "developer s"[All Fields] OR "developers"[All Fields] OR "developing"[All Fields] OR "developments"[All Fields] OR "develops"[All Fields] OR "growth and development"[MeSH Subheading] OR ("growth"[All Fields] AND "development"[All Fields]) OR "growth and development"[All Fields] OR "development"[All Fields]) AND ("recommend"[All Fields] OR "recommendable"[All Fields] OR "recommendation"[All Fields] OR "recommendation s"[All Fields] OR "recommendations"[All Fields] OR "recommended"[All Fields] OR "recommending"[All Fields] OR "recommends"[All Fields])) AND ("virtual reality"[MeSH Terms] OR ("virtual"[All Fields] AND "reality"[All Fields]) OR "virtual reality"[All Fields])) OR (("virtual"[All Fields] OR "virtuality"[All Fields] OR "virtualization"[All Fields] OR "virtualized"[All Fields] OR "virtualizing"[All Fields] OR "virtuals"[All Fields]) AND ("google"[All Fields] OR "google s"[All Fields] OR "googled"[All Fields] OR "googling"[All Fields]))

Table 1. Database search strategies.

Database	Search strategy	Results, n
PubMed	["criteria" OR "standard"] AND ["develop" OR "recommend"] AND ["virtual reality" AND "google"]	109
Cochrane Library	criteria OR standard (select Record Title) AND develop OR recommend (select Abstract) AND virtual reality (select Record Title)	72
Embase	criteria OR standard AND develop OR recommend AND virtual reality AND google {related terms are included}	48
CINAHL	criteria OR standard (select TI Title) AND develop OR recommend (select AB Abstract) AND virtual reality AND google (select TI Title)	59
MEDLINE	criteria OR standard AND develop OR recommend AND virtual reality AND google	98
Scopus	criteria OR standard AND develop OR recommend AND virtual reality AND google (TITLE-ABS-KEY)	81
ZB MED	Criteria (open search) AND develop (open search AND virtual reality (title)	173
ACM DL	[[Title: criteria] OR [Title: standard]] AND [[Abstract: develop] OR [Abstract: recommend]] AND [[Title: virtual reality] AND [Title: google]]	121
IEEE Computer Society Digital Library	criteria OR standard (select Document Title) AND develop OR recommend (select Abstract) AND virtual reality AND google (select Document Title)	115

Publications were included in the following scenarios (inclusion criteria): (1) involved an examination (generally valid) of criteria for development recommendations for VR applications; (2) were focused on VR; (3) were written in English or German or both; (4) full-text articles were freely available; (5) involved studies were relevant to the subject (ie, focusing on older adults), and (6) had been published after 2012.

According to Hülsbömer [17], VR has seen a surge in popularity since 2012, leading to a greater emphasis on its development and ongoing discussions and optimizations of development processes. Given the rapid technological advancements and increased demands in the digital age, it is crucial to include relevant literature to ensure the relevance and validity of the core set of quality criteria [18].

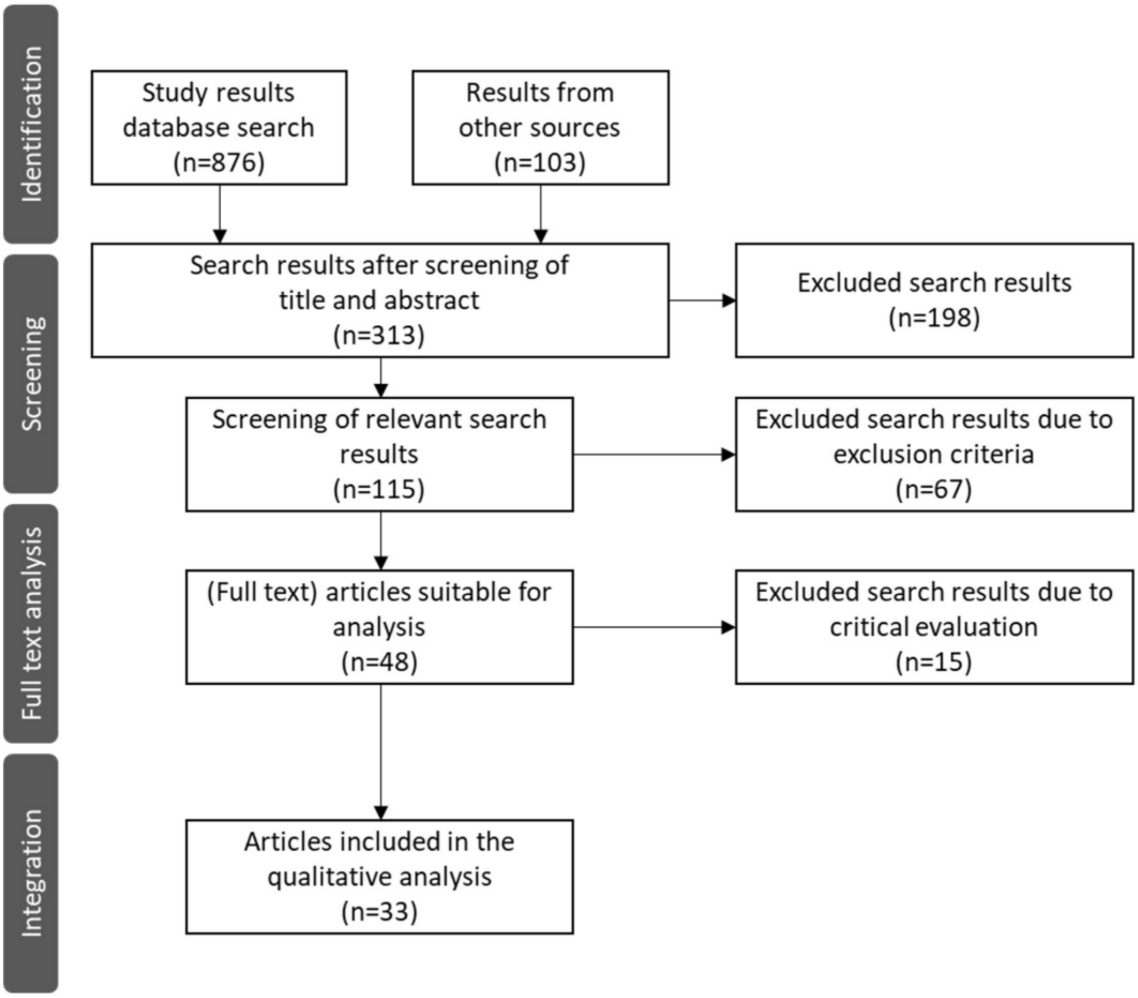
In addition, publication types with the following designs were included: articles in scientific journals, books, book chapters, study reports, guidelines, assessment instruments/assessments, legislative regulations, and reports. As the goal was to provide a comprehensive initial overview of recommendations from the VR development literature, there were no restrictions on the type of study designs reviewed.

Exclusion criteria were as follows: (1) literature results not available in German or English; (2) results written before 2012; (3) study designs in the form of laboratory studies, case reports, and series; and (4) inappropriate publication types, such as book reviews, introductions, forewords, commentaries, position statements, and letters.

For searches in the databases (PubMed, Cochrane Library, Embase, CINAHL, MEDLINE, Scopus, ZB MED, IEEE Computer Society Digital Library, and ACM Digital Library), combinations of all 3 keywords were used, as exemplified in Textbox 1 for the search string used on PubMed. The systematic literature search, including the analysis process, was conducted between February 2021 and April 2021. Some of the publications found in the databases meeting the inclusion and exclusion criteria were partially identified as duplicates. All relevant studies identified by the database searches were downloaded and stored in the literature management software EndNote (Clarivate Plc), which automatically eliminated duplicates. Based on the search strategy, 876 results were identified from the database search and 103 results from other sources. The selection process is illustrated using a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart (Figure 2). The procedure is comparable to that of Moher et al [19]. After screening the titles/abstracts, 313 publications remained, of which 198 studies were excluded, which mainly dealt with informatics-based guideline development and app development (eg, health apps or digital health apps for smartphones). Among the excluded articles, the full text was not freely available in 67. Thereafter, the remaining 115 publications were screened. Prefaces or introductory texts, statements, and general position papers were excluded at this point. Two independent reviewers assessed each title or abstract without influencing each other's decisions. The reviewers first assessed the title/abstracts for eligibility, and in the next stage assessed the full text. To perform the screening process, category formation (inclusion or exclusion) was used with EndNote. A third reviewer determined final eligibility when a discrepancy existed between the initial reviewers. Thus, 48 articles were identified for the full-text analysis. In addition, 15 results were excluded because they were thematically unsuitable, focused on only 1 topic (eg, data privacy), and were therefore not comprehensive enough. In total, 33 studies were included based on the inclusion and exclusion criteria.

The literature search yielded enough freely available literature to establish the initial structure of the core quality criteria. The identified results provided valuable initial insights for the research project. Additionally, the workshop results play a crucial role in the participatory approach, potentially informing content specifics for future older adults. Hence, we deemed the approach suitable and ruled out potential bias. Multimedia Appendix 1 displays all publications included in this study for constructing the initial quality criteria core set [20-51].

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of literature research procedure.



Framework Analysis

The content of the literature findings was analyzed using framework analysis, which involved iterative refinement of data-driven themes [52]. We chose framework analysis for its suitability in analyzing group-level qualitative data in research projects with specific goals, such as co-design. Framework analysis consists of 6 interrelated steps, which were conducted in this study as described in Textbox 2. The analysis was conducted by 2 researchers in Microsoft Excel (Microsoft Corporation). Based on the initial notes, a set of preliminary codes was developed for different aspects based on recurring themes from the literature. When disagreements arose regarding the codes, the 2 researchers discussed their conflicting interpretations and attempted to reach a consensus, possibly leading to the generation of new codes. The saturation of the data was determined by the 2 researchers. They jointly decided when saturation was reached for the formation of the first construct, which formed the basis for the expert workshops. The generation of additional themes and codes or categories and criteria took place in the 2 workshops.

Textbox 2. Framework analysis methodology, modified from Ritchie and Lewis [52].

1. **Familiarization**

   This step involves a comprehensive familiarization with the data. For the development of the quality criteria core set, the individual results of the systematic literature review search (SLS) were reviewed to familiarize with the content. The initial content was developed by forming initial codes and themes. After all literature results had gone through this process, the data set was again fully analyzed.

2. **Identify recurring and important themes**

   The theoretical framework was developed by reviewing the selected results of the SLS and synthesizing the content deemed relevant to answer the research question. Data summaries were prepared to present the data in a concise format. In addition, an initial tabulation was created with preliminary codes that were grouped in the next steps.

3. **Indexing**

   In the indexing phase, the codes were grouped and classified into framework categories. The generated codes and themes were applied to the data summaries.

4. **Charting**

   The multilayered data were systematized and structured according to the theoretical framework. Codes were assigned to categories based on their content match. Initially, some content was extracted from the literature, in terms of both content assignment and its execution. The summaries of the data were reorganized under the generated themes of the theoretical framework and rewritten in a more abstract manner to reflect the themes.

5. **Analyzing**

   The generated data tables were analyzed and discussed by the 2 researchers for content and detail. The analysis included checking for content fit as well as comparability of granularity. Codes with comparable content were merged. In addition, whether contradictions existed between the codes were analyzed. Topic summaries were created to present the findings at a high level in the context of the research question.

6. **Interpretation**

   Interpretation included discussion of content and interpretation of individual categories or codes. It can be assumed that the interpretation leaves room for the development of further criteria that were not generated in the previous steps. Content was also evaluated for complexity and completeness. Descriptions and interpretations of the themes are presented later. Explanations and insights into the themes are considered in the "Discussion" section.

Expert Workshops

The previously elaborated results of the 2 researchers were validated by conducting workshops and involving experts. Two workshops with different experts assessed the results for content fit, completeness, and level of granularity.

The first workshop, held in June 2021 due to COVID-19 pandemic restrictions, took place online and lasted for 2 hours. By involving 12 VR development experts, the previous results were revised using an interactive online collaboration platform. The experts were either self-employed or had over 10 years of experience in VR development research. The external professional reflection helped to refine and specify previous results. For the workshop, we used the MIRO platform [53] to map the previously developed results on digital boards.

The results of the framework analysis presented on the digital board were displayed on sliding digital sticky notes. The experts modified the existing structure by adding elaborations to existing criteria and generating new criteria or content using digital sticky notes. The content was generated by answering the guiding questions. The experts could freely modify the existing content of the template. In a joint discussion round, each content point was taken up, explained, and justified by the experts. In addition, the discussion aspects were recorded and an associated protocol of the respective workshops was used for the revision. The subsequent research steps involved editing the structure of the previous quality criteria core set by incorporating the results of the workshop. This was performed by assigning individual workshop contents, which were added to the categories or explanations of the criteria. New criteria were also generated. Accordingly, the content and structure of the previous version of the core set of quality criteria were adapted.

The second workshop included 12 VR application experts who had experience in the context of VR and older adults. The experts' experience with VR and the target group of older adults was based, for example, on project experience or use in clinical settings (eg, hospitals or nursing homes). The consultation of VR application experts in this step aimed to bridge the gap between the prior findings and empirical observations gathered while working with older adults as part of the research process. The 2-hour online workshop took place in November 2021 and followed a procedure similar to that of the first workshop. The online collaboration platform was used again, and the results developed until this point were presented on the boards of the platform. The given categories or criteria were the same as those of the first workshop. The goal was to further refine and specify the results. The guiding questions were also identical to those of the first workshop to ensure the comparability of methods. The experts engaged in an interactive process to further refine the previous version of the quality criteria core set, followed by subsequent discussions to ensure its specificity and accuracy.

After the 2 workshops, the 2 researchers systematized the results by considering the protocol results of both workshops. The systematization of the results involved assessing the execution of content to ensure comparability in terms of the criteria's level of detail and granularity. Changes were also made to the wording while preserving the original content's meaning, ensuring a consistent structure. The outcome of this process serves as the foundation for VR application development and is referred to as the provisionally valid quality criteria core set.

Ethical Considerations

The study was approved by the Ethics Committee of the German Sport University Cologne (Institute of Pedagogy and Philosophy; protocol number 095/22). As this is an observational study, no additional exemptions and approvals were necessary. Informed consent was given by all workshop participants. No secondary data analysis was performed. Participation in the workshop was voluntary. No personal data or data allowing conclusions to be drawn about the person were collected. All data were collected in an anonymized form. An ethics application was drafted and approved as part of adjacent activities of this research project. The submission of another, separate ethics application for the conduct of the workshops was waived. The Declaration of Helsinki also does not outline such a procedure for conducting a workshop in this particular format. Participants did not receive any compensation for taking part in the workshops.

Results

Systematic Literature Search

Following the completion of the SLS steps, which included identification, screening, full-text analysis, and integration, a total of 33 results were incorporated into the initial version of the quality criteria core set. The SLS uncovered the absence of a standardized set of quality criteria for VR applications, including those designed for older adults. To formulate the core set of quality criteria in this study, guidance from digital health applications (DiGA) in Germany was incorporated. These guidelines, such as AppQ and APPKRI, were selected because they have undergone multiple evaluations and are extensively described and assessed in terms of their content. Furthermore, it is worth noting that certain VR applications also qualify as DiGA, and as such, they were taken into account. The results of various studies provided valuable insights that were deemed essential for shaping the quality criteria core set. Additionally, specific legal provisions, such as those outlined in the Digital Health Care Act [54], were selectively integrated to ensure adherence to regulatory guidelines.

Framework Analysis

The outcome of the framework analysis yielded an initial draft of the quality criteria core set, which encompassed a range of overarching categories and diverse criteria that addressed various content aspects relevant to older adult–friendly VR applications. Because of the differences in content, the development of the quality criteria core set was divided into 2 parts with the following 2 main topics: (1) "General Criteria for VR Application Design," which included the general conditions/criteria of VR applications and (2) "Content Design of VR Applications," which included criteria that are specifically attributed to the thematic content.

The results, that is, the framework analysis codes, framework analysis categories, and derived categories of the topic "General Criteria for VR Application Design," are summarized in Table 2, and those for the topic "Content Design of VR Applications" are listed in Table 3. The framework analysis codes for the general criteria within the first category were *dependability, validity, objectivity, reliability, effectivity, knowledge, utility, treatment, improvement of (health) status, health, quality of life, health apps, psyche,* and *medical device.* As a result of this analysis, the derived framework analysis categories were *quality factors* and *medicine*. These findings led to the creation of the following categories within the first part: (1) medical quality, (2) data protection, (3) information security, (4) technical quality, (5) consumer protection and fairness, (6) interoperability, and (7) usability and motivation. The categories comprising the second part were as follows: (1) graphic/quality, (2) 3D character/avatar, (3) in-game instructions and prompts, (4) interaction, and (5) navigation.

The results of the framework analysis formed the basis for the subsequent workshops, in which the developed contents were discussed with the experts. All derived categories (ie, framework analysis categories) and framework analysis codes were disclosed in the expert workshops.

Table 2. Results of the framework analysis for the first category within the first part of the quality criteria core set focused on the "General Criteria for VR[a] Application Design."

Framework analysis codes (generated from literature)	Framework analysis category	Derived category
• Dependability • Validity • Objectivity • Reliability • Effectivity • Knowledge • Utility	• Quality factors	• Medical quality
• Treatment • Improvement of (health) status • Health • Quality of life • Health apps • Psyche • Medical device	• Medicine	

[a]VR: virtual reality.

Table 3. Results of the framework analysis for the first category within the second part of the quality criteria core set focused on the "Content Design of VR[a] Applications."

Framework analysis codes (generated from literature)	Framework analysis category	Derived category
• Software • Degree of accuracy • Structure • Fineness/detail • Concept • Design	• Graphical representation	• Graphic/quality
• VRISE (virtual reality–induced symptoms and effects) • VR-Technology • Texture • Value • Claim	• Qualitative aspects	

[a]VR: virtual reality.

Expert Workshops

Two expert workshops were conducted, drawing upon the data obtained from the systematic literature search and the framework analysis. In this process, a total of 7 categories were mapped to the first part, which encompassed "general quality criteria." These categories were accompanied by example criteria derived from the literature. The second part involved "specific criteria" and was structured around 5 categories, each supported by example criteria sourced from the literature. The experts actively engaged in the process by collaboratively editing or adding content. The subsequent discussion was guided by the following key questions:

1. Are the categories (1-7, part 1 or 1-5, part 2) complete?

2. Which categories are missing/not appropriate for VR?

3. Are the criteria in each category comprehensive enough? Are criteria missing/can criteria be replaced?

The results of the first workshop with VR development experts are described in the following section.

The categories presented in Tables 2 and 3 were deemed appropriate by the experts, and there were no alterations or criticisms regarding their suitability. However, it is important to highlight that the extent and depth of input from participants varied across the 7 different categories. Notably, the categories "data protection" and "interoperability" received the fewest additional criteria. Conversely, the categories "technical quality" and "user-friendliness and motivation" saw the most substantial contributions. Nevertheless, it was necessary to scrutinize the results to identify and rectify any potential duplication, particularly in instances where the same criteria were described using different terms.

The results of the second workshop with VR application experts revealed the following: The given categories were not criticized or adapted by the application experts, and their suitability was found to be good. The input to the respective categories varied in terms of quantity and quality (ie, level of detail, execution, and description). The fewest criteria were added in the categories "information security" and "interoperability" (n=7 and n=4, respectively), whereas the highest number was added in the categories "technical quality" and "motivation" (n=19 and n=17, respectively). The results were checked for possible duplications.

During the discussion, all categories and criteria were evaluated with the workshop participants. In this process, certain criteria were either retained or excluded for further development of the quality criteria core set. Most of these decisions on the respective contents were made unanimously among the participants. In instances where there were differing opinions, the decision-making process considered the majority agreement as the determining factor.

Following the discussion in the respective expert workshops, the results were revised or adjusted. This included the formation of subcategories as well as further criteria as a result of the reorganization of the content, which was based on the results of the workshops. The criteria were also further developed in terms of their content and categories were restructured. All changes made were based on the comments and opinions of the experts in both workshops. The current version of the quality criteria core set is briefly presented below. The first category of both parts is presented in detail in the respective subcategories (cf. Figures 3-6), whereas the remaining contents are presented in the form of categories and criteria (cf. Tables 4 and 5).

Figure 3. Excerpt of the first part of the quality criteria core set to the subcategory 1.1 Evidence.

Part 1 of the quality criteria core set: Framework conditions for the development of VR applications suitable for senior citizens.

1. Category: quality assurance of medical/health content

This category includes criteria that ensure the quality of content in a medical or health context. This includes criteria that ensure the quality of care for seniors with VR applications.

1.1 Evidence

1.1.1 Content is evidence-based
The content must be backed up by studies that are thematically identical or at least comparable. A thorough evaluation of the studies with regard to criteria such as timeliness, study design, sample, study period, methodology, as well as results and discussion must be carried out. The literature search should conform to established standards. A comparison to current study results on the topic should be considered. The assessment of the comparability of studies (results) and the transfer to the project being undertaken should be reliable. In addition, technical literature should also be included.
1.1.2 Consideration of generally accepted professional and current standards
The implementation of (medical/health) content and procedures should be based on the Lege-Artis rule. Thereby a consideration of the current gold standard takes place. Subject matter experts should be involved in VR development.
1.1.3 Quality assurance of content and procedures
The recommendations for action from guidelines on content or procedures are followed.
1.1.4 Publication of sources used for the content or its implementation
Indication/publication of a list of sources that makes the sources used easily identifiable in terms of their readability and takes into account common citation methods/standards.

The content is derived from the codes or preliminary categories derived during the framework analysis. Based on thematic focus, subcategories were formed to differentiate the categories as well as to detail the criteria. Figure 3 shows a section of subcategory *1.1 Evidence*, which was assigned to the first category *Quality assurance of medical/health content* of the first part. The second subcategory of this part is *1.2 Application safety (before/during use)*, which is shown in Figure 4 with corresponding contents.

Figure 4. Excerpt of the first part of the quality criteria core set to the subcategory 1.2 Application safety (before/ during use).

1.2 Application safety (before/ during use)

1.2.1 Health status assessment
Before using the VR application, the health status of the person using it should be checked. Physical and mental (pre-)illnesses or limitations that could have an influence on VR use should be clarified as part of their medical history. Furthermore, factors such as medication (or its side effects) should be taken into account. In addition, contraindications to VR use should be ruled out. The use of possible aids (glasses, walking stick, hearing aid, rollator, etc.) during VR use must be considered or checked for suitability. Consideration should be given to having medical personnel (with appropriate qualifications, if necessary) present as caregivers during use.
1.2.2 Assessment of physical/mental fitness
Physical factors, such as mobility status or mobility, should be checked by means of appropriate assessments. Assessments should also be carried out to evaluate the mental state.
1.2.3 Cyber/motion sickness assessment
The user should be informed about VR in general and about the VR application to be used. A pre-test should be used to assess the cyber-/motion sickness of the user. Special assessments can be used for a further evaluation of cyber/motion sickness.
1.2.4 Assessment of fall/injury risk during VR application use
The (fall) safety of the user should be ensured as far as possible. The risk of the user falling or being injured during VR use should be checked in advance. In addition to targeted information, the medical history taken, and the performance of physical and psychological assessments, a pre-test would provide initial indications under safety measures to avoid the risk of falling/injury. Moreover, depending on the application and forms of interaction, targeted assessments can be performed to evaluate the risk of falls. Furthermore, an ethics application should be considered. In addition, the insurance relationship should be clarified.
1.2.5 Appropriate use of technology
For the appropriate handling of VR technology and other technologies required for its use, training should be provided by a qualified person on the use of the application for the persons involved. This would involve learning how to operate VR in terms of hardware and software, and how to handle hygiene. If the senior citizens are taught how to use the system, care should be taken to ensure that the instructions are appropriate for the target group.
1.2.6 Securing the hardware
The input device should be certified.
1.2.7 Securing the content
The software should be certified. The best possible guarantee for protection against misuse/malware should be fulfilled, e.g. by using security tools. Attention should be paid to fast data transfer using common procedures.

Figures 5 and 6 refer to the first category graphic/quality of the second part and depict corresponding criteria in the subcategories *1.1 Object-related/environment-related* and *1.2 User-related*, respectively, which were thematically subordinated to the subcategories.

The complete set of quality criteria with all versions can be requested from the authors.

Figure 5. Excerpt of the second part of the quality criteria core set to the subcategory 1.1 Object related/ environment-related.

Part 2 of the Quality Criteria Core Set: Specific Criteria for the Development of VR Applications for Seniors

1. Category: Graphic/ Quality

This category includes criteria that map requirements related to the graphical and qualitative development of VR applications. The criteria are aimed, among other things, to keep the psychological burden on the person using the application as low as possible.

1.1 Object-related/ environment-related

1.1.1 Spatial allocation
The spatial assignment should be clear. A strong contrast between the foreground and background should be chosen. This contributes to the immersion. Depending on the need, blurring can be supportive to the step of distinguishing between reality and virtual reality.
1.1.2 Color space
If possible, warm colors should be selected in the sense of color psychology in order to achieve a more comfortable feeling among the senior citizens. The color depth and the color spectrum (e.g. 256 colors) should be chosen carefully. Color vision disorders should also be taken into account when selecting colors, or alternative representations should be offered.
1.1.3 Objects
The objects should have a small number of polygons. The style should be uniform, so that there is a certain consistency in the object design in terms of level of detail and texture resolution. In addition, the object scaling should be consistent with reality, especially interactive elements should be measured virtually. To reduce unwanted effects in graphics, software solutions should be used to increase quality (e.g. methods such as anti-aliasing, texture baking). The visual level of detail should be high and at the same time clear.
1.1.4 Scenario
Scenes should be designed to be ecologically valid as possible. The virtual environment should not be too dark or too futuristic or similar, so that it has a frightening effect on the person using it.
1.1.5 Content design
The content should be stereoscopic to achieve high immersion. With the use of 3D, greater immersion is achieved compared to 2D. As a result, the sense of presence is stronger. The latency should be as low as possible. The frame rate should be at least 60 fps. The resolution should be as high as possible, at least 1832x1920px is recommended. Good image sharpness should be considered accordingly. The design should be as naturalistic as possible, i.e. light and shadow should be used. In addition, the realism should be increased by having a temporal illusion through a change in the environment. Computationally intensive processes should be avoided. Time periods in which the person using the application has to wait (e.g., when the application is loading) should be kept to a minimum.

Figure 6. Excerpt of the second part of the quality criteria core set to the subcategory 1.2 User-related.

1.2 User-related

1.2.1 Information gift
The information provided in VR should be short and understandable. The use of long texts in the VR application should be avoided. Only essential content should be reproduced in short text form and in a way that is easy to read. The size as well as the contrast of the font should be individually adjustable. The content layout should be centered. Scrollbars should be avoided. An easy-to-read or large font should be chosen. The text should be left-aligned. Visual metaphors should be incorporated whenever possible. The context of use should be clearly defined.
1.2.2 Interaction elements
Due to the partial sensory-motor blurring of senior citizens, the interaction elements should be large. In addition, standardized symbols should be used. The objects in VR should be appropriate for the target group. The interaction mode should be adapted to the respective generation and thus include more mechanical interaction. For a good body perception of the user, visual and proprioceptive information should be combined in an optimized way. A cursor or similar should be provided so that the selected target is clearly indicated (e.g. bright border when an object is selected).
1.2.3 Language
The language should be kept simple and as presupposition-free as possible. Foreign-language terms should be avoided. Terms suitable for senior citizens should be used. Pictograms could be used instead of language or as a support. Language should be natural and accessible to the person using it. Phrases should be positive rather than negative. The form of address should be active and not passive. The language conversion should be done using text or dialog elements.
1.2.4 User interface
The structure of the menu should be clear and include intuitive mapping. The punctuation as well as the execution of commands should be similar. The layout, navigation and terminology used should be simple, clear and consistent. Attention-grabbing elements that are not (no longer) needed should be removed. Complexity should be avoided as much as possible by removing things that are rarely used or not necessary. The elements of the user interface should be easily modifiable by the person using it (e.g. zoom in). The user guidance should be clear, so that choices as well as any selections made are apparent to the person using the system.
1.2.5 Audio
The audio setting should be individual so that it can be selected as desired. Distractions caused by unwanted sounds should be avoided. The audio design should be immersive and take complex soundscapes into account (e.g. 7.1 surround sound). The audio should be senior-friendly (e.g. emphasis, speed, but also phrasing). The quality of the audio should be good and understandable through the use of clear language.

Table 4. Presentation of categories, subcategories, and criteria of the first part of the quality criteria core set.

Category and subcategory	Criteria
2. Data protection provisions	
2.1 Data generation and data storage	• 2.1.1 Earmarking • 2.1.2 Necessity and data economy • 2.1.3 Consent • 2.1.4 Data security • 2.1.5 Ensuring the authenticity of the data • 2.1.6 Information security
2.2 Legitimacies	• 2.2.1 Data protection law • 2.2.2 Data participant rights • 2.2.3 Accesses and access rights
3. Quality requirements	
3.1 Technical safety	• 3.1.1 Robustness against disturbances • 3.1.2 Hygiene
3.2 User-related security	• 3.2.1 Avoiding the risk of collision • 3.2.2 Suitable spectacle ergonomics • 3.2.3 Change of operability • 3.2.4 Traceability application situation
4. Consumer protection	
4.1 Accessibility	• 4.1.1 User support • 4.1.2 Conditions of use
4.2 Transparency	• 4.2.1 Data/software update • 4.2.2 Transparent business models
5. Interoperability	
5.1 Data portability	• 5.1.1 Data extraction
5.2 Technology compatibility	• 5.2 1 Connection of external technologies • 5.2.2 Cross-generational use of technology

Table 5. Presentation of categories, subcategories, and criteria of the second part of the quality criteria core set.

Category and subcategory	Criteria
2. 3D character/avatar	
2.1 Appearance	• 2.1.1 Character • 2.1.2 Appearance/optics/optical representation
2.2 Behaviorism	• 2.2.1 Behavior (facial expressions/gestures) • 2.2.2 Tone and pitch of voice
3. Providing in-game instructions and prompts	
3.1 Didactics	• 3.1.1 Initial learning • 3.1.2 Instructions • 3.1.3 Goal and task design • 3.1.4 Feedback • 3.1.5 Information transmission
3.2 General conditions	• 3.2.1 Useful life • 3.2.2 Local framework conditions
4. Interaction	
4.1 Functionalities	• 4.1.1 Operation • 4.1.2 Design/functionality • 4.1.3 Relevance assessment
4.2 Regulation	• 4.2.1 Evaluation/developments • 4.2.2 Perception/proprioception
5. Navigation	
5.1 Operation	• 5.1.1 Menu control • 5.1.2 Control systems • 5.1.3 Room design
5.2 Taxis	• 5.2.1 Locomotion • 5.2.2 Tracking
6. Promotion of user motivation and loyalty to use	
6.1 Usability	• 6.1.1 Motivation • 6.1.2 Logging • 6.1.3 Configuration
6.2 Target group orientation	• 6.2.1 Adaptability/customizability for different needs and interests • 6.2.2 Practicability • 6.2.3 User-friendliness

Discussion

Overview

The study's objective was to establish a foundational set of quality criteria to provide guidance for the development of VR applications catering to older adults. The findings encompassed a diverse range of quality criteria derived from various qualitative research methods. This study represents the pioneering effort to create a core set of criteria specifically tailored to the development of VR applications for older adults. It introduced a novel qualitative methodological approach that involved a systematic review search, followed by framework analysis and validation through expert workshops, to formulate these criteria.

Principal Findings

The results of the preliminary quality criteria core set provide an initial basis for older adult–friendly VR development. The individual criteria and categories are derived from the results of the SLS, the framework analysis, and the 2 expert workshops. It should be noted that the SLS research results included some of the German literature that was considered in the development process. This is attributed to the fact that the results of the quality criteria core set are primarily applicable to Germany and, as such, consider the specific regulations of the German health care system and legislation. Nevertheless, it is not precluded that these results can be extrapolated to other countries. In such instances, it is advisable to carefully examine the corresponding regulations that may diverge at the national or international level.

The validity of the criteria hinges on the research findings and the latitude for interpretation inherent in the reflective processes of the researchers involved. The expert workshops were thus crucial for the validation of the results. The results represent the criteria developed in the iterative process, which are detailed using explanation examples. Considering the thorough survey conducted across various databases and the consensus among experts during the 2 workshops, it is reasonable to consider the resulting quality criteria core set as a typical representation of a preliminary version. However, the correctness or durability of the results is uncertain. Given the rapid development of technology, particularly in the VR market, it must be assumed that the criteria within the provisional core set of quality criteria are only valid to a limited extent. Nevertheless, this version forms a first building block and serves as a guideline for basic VR development for the target group of older adults.

The categorization of criteria into their respective subcategories (as seen in Figures 3-6) is guided by thematic considerations. However, it is important to acknowledge that thematic overlaps may occur, and the precise classification of criteria into specific subcategories may need to be reevaluated based on the context of the application. For the initial draft of the quality criteria core set, which serves as the foundation, the current structuring within the subcategories and sorting is deemed appropriate.

The distribution of criteria among the respective (sub)categories is somewhat uneven, with varying numbers of criteria in each category. This suggests that certain topic areas receive more extensive coverage than others. An illustrative example of this is the category of data protection, which encompasses a broader range of content within the quality criteria core set compared with other categories. One plausible explanation for this discrepancy is that certain topic areas are subject to stringent standards and requirements, resulting in a greater abundance of detailed explanations compared with other categories.

For a general understanding, these criteria are explained in detail. However, the degree of granularity must be considered: The understanding and interpretation of the explanations vary depending on the individual's level of knowledge, experience, and assessment. Therefore, explanations provided for the individual criteria leave some room for interpretation. The fulfillment of the criteria depends on the respective framework conditions and the hardware. VR head-mounted displays and other VR hardware have different functionalities. Therefore, the requirements developed in the quality criteria form may not be fulfilled or only be insufficiently fulfilled in some cases. The functionalities of the VR hardware also determine the definition and fulfillment of the quality criteria.

The results should reflect older adult–friendly criteria. However, the heterogeneity of the target group must be considered. In old age, numerous physiological and psychological changes occur within individuals. Accordingly, the requirements for VR applications or VR systems may vary within the target group of older adults [55]. Defining older adults by age could assist in refining the criteria, while also considering the stereotypical characteristics associated with each age group. Older adults may use VR in different contexts [56]. To specify the criteria for VR applications suitable for older adults, a classification of application areas could also be helpful. Moreover, the purpose of the VR application is significant, as are considerations about the motivations behind VR application development. The purpose of VR needs to be thoroughly justified. Additionally, older adults' level of technology familiarity and willingness to use VR technology influence these considerations [57,58]. The development of VR applications for older adults requires an analysis of the necessary functionalities relevant to the development project in the initial conceptual phase. The preliminary valid core set of quality criteria serves as a guideline to specify and define further aspects after initial considerations for application development. The validity of the individual criteria must be determined individually and possibly modified depending on the application purpose. In principle, these criteria can serve as valuable guidelines to facilitate more focused design development, beginning from the conceptual stage. By considering these criteria throughout subsequent phases, informed decisions can be made regarding their relevance and integration into the design process.

The significance of quality criteria is continually increasing, and it has become an important topic for discussion. Manser and de Bruin [59] have also called for quality criteria that are thematically related to exergames and described a different approach; however, their study also focused on the target group of older adults [59]. The authors refer to the framework proposed by Li et al [60], which represents a comparable methodological approach to this work but does not clearly map defined quality criteria.

It should be noted that for the development of a targeted VR application for older adults, a participatory approach is important. In addition to people from the creative industry (eg, VR developers, game designers, computer scientists) and people from the relevant application area, the target group should be included. The existing version of the core set of quality criteria for older adult–friendly VR applications will undergo further refinement through the inclusion of older adults in future research projects. The VR development work should thus occur through collaborative and transparent exchanges with both current users and prospective users. Feedback should also be evaluated and taken into account with regard to technical feasibility. In pretest phases and based on prototypes, new insights can be gained that can be used to further specify the core set of quality criteria.

Limitations

Individual categories or criteria, such as "data protection," must be dealt with more comprehensively to ensure the validity of the content. This example category often shows significant changes, as the topic is currently undergoing constant revision and also needs to be adapted to country-specific requirements.

The theoretical construct in the form of the preliminary valid core set of quality criteria will be tested in the next step involving practical implementation. The criteria developed so far will be incorporated into the development work so that practical testing can determine whether the individual criteria can be met. In an upcoming workshop with older adults on the current version of the quality criteria core set, the criteria will be discussed and put into their final form. When discussion and refinement of the content is complete, the version is referred to as the "evaluated and standardized quality criteria core set." Thereafter, the updated results will be presented in a follow-up publication, in which the target group of older adults will be directly included and verification of the theoretically constructed quality criteria will be performed using a VR application in practice.

Conclusions

User-centered development is useful if a product or measure is to reach the target audience. The quality criteria core set should act as a kind of guideline for the implementation of the targeted development work. Attention should be paid to a possible specification of the user groups and the context of use. In addition, the heterogeneity of the target group of older adults should be considered. The quality criteria core set serves as an initial step toward user-centered VR application development, but additional research is needed to build upon the existing results and further enhance the core set.

Data Availability

The data sets are provided in the main manuscript by the authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Included publications according to systematic literature review search (SLS). See also [20-51].
[DOCX File , 24 KB - ijmr_v12i1e45433_app1.docx]

References

1. Cipresso P, Giglioli IAC, Raya MA, Riva G. The past, present, and future of virtual and augmented reality research: a network and cluster analysis of the literature. Front Psychol 2018 Nov 6;9:1-20 [FREE Full text] [doi: 10.3389/fpsyg.2018.02086] [Medline: 30459681]
2. Dermody G, Whitehead L, Wilson G, Glass C. The role of virtual reality in improving health outcomes for community-dwelling older adults: systematic review. J Med Internet Res 2020 Jun 01;22(6):1-17 [FREE Full text] [doi: 10.2196/17331] [Medline: 32478662]
3. Freeman D, Reeve S, Robinson A, Ehlers A, Clark D, Spanlang B, et al. Virtual reality in the assessment, understanding, and treatment of mental health disorders. Psychol Med 2017 Oct;47(14):2393-2400 [FREE Full text] [doi: 10.1017/S003329171700040X] [Medline: 28325167]
4. Liao Y, Tseng H, Lin Y, Wang C, Hsu W. Using virtual reality-based training to improve cognitive function, instrumental activities of daily living and neural efficiency in older adults with mild cognitive impairment. Eur J Phys Rehabil Med 2020 Feb;56(1):47-57 [FREE Full text] [doi: 10.23736/S1973-9087.19.05899-4] [Medline: 31615196]
5. Maggio MG, Latella D, Maresca G, Sciarrone F, Manuli A, Naro A, et al. Virtual reality and cognitive rehabilitation in people with stroke: an overview. J Neurosci Nurs 2019 Apr;51(2):101-105. [doi: 10.1097/JNN.0000000000000423] [Medline: 30649091]
6. Faria AL, Andrade A, Soares L, I Badia SB. Benefits of virtual reality based cognitive rehabilitation through simulated activities of daily living: a randomized controlled trial with stroke patients. J Neuroeng Rehabil 2016 Nov 02;13(1):96-134 [FREE Full text] [doi: 10.1186/s12984-016-0204-z] [Medline: 27806718]
7. Ortet CP, Veloso AI, Vale Costa L. Cycling through 360° virtual reality tourism for senior citizens: empirical analysis of an assistive technology. Sensors (Basel) 2022 Aug 17;22(16):6169 [FREE Full text] [doi: 10.3390/s22166169] [Medline: 36015929]
8. Tobler-Ammann BC, Surer E, Knols RH, Borghese NA, de Bruin ED. User perspectives on exergames designed to explore the hemineglected space for stroke patients with visuospatial neglect: usability study. JMIR Serious Games 2017 Aug 25;5(3):e18 [FREE Full text] [doi: 10.2196/games.8013] [Medline: 28842390]
9. Aydin S, Aktaş B. Developing an integrated VR infrastructure in architectural design education. Front Robot AI 2020;7:495468 [FREE Full text] [doi: 10.3389/frobt.2020.495468] [Medline: 33501294]
10. Kyaw BM, Saxena N, Posadzki P, Vseteckova J, Nikolaou CK, George PP, et al. Virtual reality for health professions education: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019 Jan 22;21(1):e12959 [FREE Full text] [doi: 10.2196/12959] [Medline: 30668519]
11. Chen P, Krch D. Immersive virtual reality treatment for spatial neglect: an agile, user-centered development process. Ann Phys Rehabil Med 2022 May;65(3):101592 [FREE Full text] [doi: 10.1016/j.rehab.2021.101592] [Medline: 34662734]
12. Rowe JW, Kahn RL. Human aging: usual and successful. Science 1987 Jul 10;237(4811):143-149. [doi: 10.1126/science.3299702] [Medline: 3299702]

13. Rudnicka E, Napierała P, Podfigurna A, Męczekalski B, Smolarczyk R, Grymowicz M. The World Health Organization (WHO) approach to healthy ageing. Maturitas 2020 Sep;139:6-11 [FREE Full text] [doi: 10.1016/j.maturitas.2020.05.018] [Medline: 32747042]
14. Dziechciaż M, Filip R. Biological psychological and social determinants of old age: bio-psycho-social aspects of human aging. Ann Agric Environ Med 2014;21(4):835-838 [FREE Full text] [doi: 10.5604/12321966.1129943] [Medline: 25528930]
15. Afifi T, Collins NL, Rand K, Fujiwara K, Mazur A, Otmar C, et al. Testing the feasibility of virtual reality with older adults with cognitive impairments and their family members who live at a distance. Innov Aging 2021;5(2):igab014 [FREE Full text] [doi: 10.1093/geroni/igab014] [Medline: 34632105]
16. Birckhead B, Khalil C, Liu X, Conovitz S, Rizzo A, Danovitch I, et al. Recommendations for methodology of virtual reality clinical trials in health care by an international working group: iterative study. JMIR Ment Health 2019 Jan 31;6(1):e11973 [FREE Full text] [doi: 10.2196/11973] [Medline: 30702436]
17. Hülsbömer S. Hype Cycles der letzten zehn Jahre: Gartner-Trends im Reality Check. COMPUTERWOCHE. 2015 Dec 15. URL: https://www.computerwoche.de/a/gartner-trends-im-reality-check,3070089 [accessed 2023-06-20]
18. Rollwagen I. Zeit und Innovation: Zur Synchronisation von Wirtschaft, Wissenschaft und Politik bei der Genese der Virtual-Reality-Technologien. Bielefeld, Germany: transcript; 2015.
19. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64 [FREE Full text] [doi: 10.7326/0003-4819-151-4-200908180-00135] [Medline: 19622511]
20. Fraunhofer-Institut für Offene Kommunikationssysteme. APPKRI Kriterien für Gesundheits-Apps. Fraunhofer-Institut für Offene Kommunikationssysteme [FOKUS]. 2018. URL: https://ehealth-services.fokus.fraunhofer.de/BMG-APPS [accessed 2021-10-21]
21. Bertelsmann Stiftung. AppQ: Gütekriterien-Kernset für mehr Qualitätstransparenz bei digitalen Gesundheitsanwendungen. Bertelsmann Stiftung. 2019. URL: https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/appq/ [accessed 2021-10-21]
22. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
23. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: a comprehensive quality and therapeutic potential evaluation tool for mobile and web-based eHealth interventions. J Med Internet Res 2017 Mar 21;19(3):e82 [FREE Full text] [doi: 10.2196/jmir.7270] [Medline: 28325712]
24. Haute Autorité de Santé (HAS). Good practice guidelines on health apps and smart devices (mobile health or mhealth). Haute Autorité de Santé (HAS). 2016. URL: https://www.has-sante.fr/jcms/c_2681915/en/good-practice-guidelines-on-health-apps-and-smart-devices-mobile-health-or-mhealth [accessed 2021-10-21]
25. National Health Service Digital. Digital Assessment Questionnaire V2.1. National Health Service Digital. 2018. URL: https://digital.nhs.uk/search?query=digital+assessment+questionnaire [accessed 2021-10-21]
26. American Psychiatric Association (APA). The App Evaluation Model. American Psychiatric Association (APA). 2018. URL: https://www.psychiatry.org/psychiatrists/practice/mental-health-apps/the-app-evaluation-model [accessed 2021-10-21]
27. Bertelsmann Stiftung. AppQ 1.1: Gütekriterien-Kernset für mehr Qualitätstransparenz bei digitalen Gesundheitsanwendungen. Bertelsmann Stiftung. 2020. URL: https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/appq-1-1 [accessed 2021-10-21]
28. Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM). DiGA-Leitfaden. Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM). 2020. URL: https://www.bfarm.de/SharedDocs/Downloads/DE/Service/Beratungsverfahren/DiGA-Leitfaden.html [accessed 2021-10-21]
29. Bundesinstituts für Arzneimittel und Medizinprodukte (BfArM). BfArM-Orientierungshilfe medical apps. Bundesinstituts für Arzneimittel und Medizinprodukte (BfArM). 2016. URL: https://www.bfarm.de/SiteGlobals/Forms/Suche/Servicesuche_Formular.html;jsessionid=EB5D9FFA612077A546DEAD32D188F243.intranet262?nn=468782&resourceId=468548&input_=494988&pageLocale=de&templateQueryString=BfArM-Orientierungshilfe+Medical+Apps&submit.x=0&submit.y=0 [accessed 2021-10-21]
30. Bundesministerium für Gesundheit (BMG). Digitale-Gesundheitsanwendungen-Verordnung (DiGAV). Bundesministerium für Gesundheit (BMG). 2020. URL: https://www.bundesgesundheitsministerium.de/service/gesetze-und-verordnungen/guv-19-lp/digav.html [accessed 2021-10-21]
31. Bertelsmann Stiftung. Digital-Health-Anwendungen für Bürger: Kontext, Typologie und Relevanz aus Public-Health-Perspektive. Bertelsmann Stiftung. 2016. URL: https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/digital-health-anwendungen-fuer-buerger/ [accessed 2021-10-21]
32. Bertelsmann Stiftung. Transfer von Digital-Health-Anwendungen in den Versorgungsalltag (Teil 6), Teil 6: Transparenzmodell Digital-Health-Anwendungen – Grundlagen, Herleitung und Modell. Bertelsmann Stiftung. 2019. URL: https://www.bertelsmann-stiftung.de/de/publikationen/publikation/did/transfer-von-digital-health-anwendungen-in-den-versorgungsalltag-teil-6 [accessed 2021-10-21]

33. Albrecht UV. Chancen und Risiken von Gesundheits-Apps (CHARISMHA). Technische Universität Braunschweig. 2016. URL: https://publikationsserver.tu-braunschweig.de/receive/dbbs_mods_00060000 [accessed 2021-10-21]
34. Bucksch M. Leitfaden für die Entwicklung von Medical Apps: Darauf müssen Hersteller achten. QuickBird Medical. 2021. URL: https://quickbirdmedical.com/medical-app-entwicklung-mdr/ [accessed 2021-10-21]
35. Aktionsforum Gesundheitsinformationssystem (afgis). Gesundheits-App Fact Sheet. Aktionsforum Gesundheitsinformationssystem (afgis). 2012. URL: https://www.afgis.de/standards/gesundheitsapps/ [accessed 2021-10-21]
36. Aktionsforum Gesundheitsinformationssystem (afgis). afgis-Transparenzkriterien. Aktionsforum Gesundheitsinformationssystem (afgis). 2021. URL: https://www.afgis.de/zertifizierung/transparenzkriterien [accessed 2021-10-21]
37. Universitätsklinikum Freiburg. Gesundheits- und Versorgungs-Apps. Universitätsklinikum Freiburg. 2013. URL: https://www.uniklinik-freiburg.de/fileadmin/mediapool/09_zentren/studienzentrum/pdf/Studien/150331_TK-Gesamtbericht_Gesundheits-und_Versorgungs-Apps.pdf [accessed 2021-10-21]
38. Europäische Kommission. Grünbuch über Mobile-Health-Dienste (mHealth). Bundesrat. 2014. URL: https://www.bundesrat.de/SharedDocs/beratungsvorgaenge/2014/0101-0200/0167-14.html [accessed 2021-10-21]
39. Bruder R, Eckert T, Conradt J, Caserman P, Schaub M, Hofmann K, et al. Gütekriterien Serious Games - Langfassung 30.03.2021. TU Darmstadt. 2021. URL: https://tuprints.ulb.tu-darmstadt.de/17872/ [accessed 2021-10-21]
40. Kourtesis P, Korre D, Collina S, Doumas L, MacPherson S. Guidelines for the development of immersive virtual reality software for cognitive neuroscience and neuropsychology: the development of Virtual Reality Everyday Assessment Lab (VR-EAL), a neuropsychological test battery in immersive virtual reality. Front Comput Sci 2020 Jan 14;1:12. [doi: 10.3389/fcomp.2019.00012]
41. Cong X, Li T. Design and development of virtual medical system interface based on VR-AR hybrid technology. Comput Math Methods Med 2020;2020:7108147 [FREE Full text] [doi: 10.1155/2020/7108147] [Medline: 32908580]
42. Vogel J, Schuir J, Thomas O, Teuteberg F. Gestaltung und erprobung einer virtual-reality-anwendung zur unterstützung des prototypings in design-thinking-prozessen. HMD 2020 Mar 26;57(3):432-450. [doi: 10.1365/s40702-020-00608-9]
43. Kourtesis P, Collina S, Doumas LAA, MacPherson SE. Technological competence is a pre-condition for effective implementation of virtual reality head mounted displays in human neuroscience: a technological review and meta-analysis. Front Hum Neurosci 2019;13:342. [doi: 10.3389/fnhum.2019.00342] [Medline: 31632256]
44. Madary M, Metzinger T. Recommendations for good scientific practice and the consumers of VR-technology. Front Robot AI 2016 Feb 19;3:1-23. [doi: 10.3389/frobt.2016.00003]
45. Boletsis C, Cedergren J. VR locomotion in the new era of virtual reality: an empirical comparison of prevalent techniques. Advances in Human-Computer Interaction 2019 Apr 01;2019:1-15. [doi: 10.1155/2019/7420781]
46. LaValle SM. Virtual Reality. Cambridge, UK: Cambridge University Press; 2015.
47. Govea-Valladares EH, Medellin-Castillo HI, Ballesteros J, Rodriguez-Florido MA. On the development of virtual reality scenarios for computer-assisted biomedical applications. J Healthc Eng 2018;2018:1930357 [FREE Full text] [doi: 10.1155/2018/1930357] [Medline: 30245783]
48. Bundesverband Informationswirtschaft, Telekommunikation und neue Medien (Bitkom) e.V. Augmented und Virtual Reality: Potenziale und praktische Anwendung immersiver Technologien. Bitkom. 2021. URL: https://www.bitkom.org/Bitkom/Organisation/Gremien/Augmented-and-Virtual-Reality.html [accessed 2021-10-21]
49. Brennesholtz MS. 3‐1: Invited Paper: VR standards and guidelines. SID Symposium Digest of Technical Papers 49 2018 May;49(1):1-4. [doi: 10.1002/sdtp.12476]
50. Stach M, Kraft R, Probst T, Messner EM, Terhorst Y, Baumeister H, et al. Mobile health app database - a repository for quality ratings of mHealth apps. New York, NY: IEEE; 2020 Presented at: 2020 IEEE 33rd International Symposium on Computer-Based Medical Systems (CBMS); July 28-30, 2020; Rochester, MN p. 427-432. [doi: 10.1109/CBMS49503.2020.00087]
51. Yuan Y. Paving the road for virtual and augmented reality [standards]. IEEE Consumer Electron Mag 2018 Jan;7(1):117-128. [doi: 10.1109/MCE.2017.2755338]
52. Ritchie J, Lewis J. Qualitative Research Practice: A Guide for Social Science Students and Researchers. London, UK: Sage; 2003.
53. MIRO. URL: https://miro.com [accessed 2023-09-14]
54. Gesetz für eine bessere Versorgung durch Digitalisierung und Innovation (Digitale-Versorgung-Gesetz - DVG). Bundesgesetzblatt. 2009 Dec 09. URL: https://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBl&start=%2F%2F%2A%5B%40attr_id=%27bgbl119s2562.pdf%27%5D#__bgbl__%2F%2F*%5B%40attr_id%3D%27bgbl119s2562.pdf%27%5D__1695022237786 [accessed 2023-05-07]
55. Garcia JA. A virtual reality game-like tool for assessing the risk of falling in the elderly. Stud Health Technol Inform 2019 Aug 08;266:63-69. [doi: 10.3233/SHTI190774] [Medline: 31397303]
56. Perez-Marcos D. Virtual reality experiences, embodiment, videogames and their dimensions in neurorehabilitation. J Neuroeng Rehabil 2018 Nov 26;15(1):113 [FREE Full text] [doi: 10.1186/s12984-018-0461-0] [Medline: 30477527]

57. Cook N, Winkler SL. Acceptance, usability and health applications of virtual worlds by older adults: a feasibility study. JMIR Res Protoc 2016 Jun 02;5(2):e81 [FREE Full text] [doi: 10.2196/resprot.5423] [Medline: 27256457]
58. Roccetti M, Prandi C, Mirri S, Salomoni P. Designing human-centric software artifacts with future users: a case study. Hum Cent Comput Inf Sci 2020 Mar 10;10(1):8. [doi: 10.1186/s13673-020-0213-6]
59. Manser P, de Bruin ED. Making the best out of IT: design and development of exergames for older adults with mild neurocognitive disorder - a methodological paper. Front Aging Neurosci 2021 Dec 9;13:734012 [FREE Full text] [doi: 10.3389/fnagi.2021.734012] [Medline: 34955806]
60. Li Y, Muñoz J, Mehrabi S, Middleton L, Cao S, Boger J. Multidisciplinary iterative design of exergames (MIDE): a framework for supporting the design, development, and evaluation of exergames for health. In: Fang X, editor. HCI in Games. Cham, Switzerland: Springer; 2020:128-147.

Abbreviations

DiGA: digital health applications
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SLS: systematic literature review search
VR: virtual reality
WHO: World Health Organization

Edited by T de Azevedo Cardoso; submitted 30.12.22; peer-reviewed by J Lander, PJL Guo, M Roccetti; comments to author 15.03.23; revised version received 07.05.23; accepted 21.08.23; published 27.09.23.

Please cite as:
Napetschnig A, Brixius K, Deiters W
Development of a Core Set of Quality Criteria for Virtual Reality Applications Designed for Older Adults: Multistep Qualitative Study
Interact J Med Res 2023;12:e45433
URL: https://www.i-jmr.org/2023/1/e45433
doi:10.2196/45433
PMID:37756112

Original Paper

A Pragmatic Mapping of Perceptions and Use of Digital Information Systems in Primary Care in Sweden: Survey Study

Anita Sant'Anna[1], PhD; Jens Nygren[2], PhD

[1]Viniam Consulting AB, Halmstad, Sweden

[2]School of Health and Welfare, Halmstad University, Halmstad, Sweden

Corresponding Author:
Jens Nygren, PhD
School of Health and Welfare
Halmstad University
Box 823
Halmstad, 30118
Sweden
Phone: 46 035167100
Email: jens.nygren@hh.se

Abstract

Background: Electronic health records and IT infrastructure in primary care allow for digital documentation and access to information, which can be used to guide evidence-based care and monitor patient safety and quality of care. Quality indicators specified by regulatory authorities can be automatically computed and presented to primary care staff. However, the implementation of digital information systems (DIS) in health care can be challenging, and understanding factors such as relative advantage, compatibility, complexity, trialability, and observability is needed to improve the success and rate of adoption and diffusion.

Objective: This study aims to explore how DIS are used and perceived by health care professionals in primary care.

Methods: This study used quantitative assessment to gather survey data on the use and potential of DIS in health care in Sweden from the perspectives of primary care personnel in various roles. The digital questionnaire was designed to be short and contained 3 sections covering respondent characteristics, current use of platforms, and perceptions of decision support tools. Data were analyzed using descriptive statistics, nonparametric hypothesis testing, ordinal coefficient α, and confirmatory factor analysis.

Results: The study collected responses from participants across 10 regions of Sweden, comprising 31.9% (n=22) from private clinics and 68.1% (n=47) from public clinics. Participants included administrators (18/69, 26.1%), a medical strategist (1/69, 1.4%), and physicians (50/69, 72.5%). Usage frequency varied as follows: 11.6% (n=8) used DIS weekly, 24.6% (n=17) monthly, 27.5% (n=19) a few times a year, 26.1% (n=18) very rarely, and 10.1% (n=7) lacked access. Administrators used DIS more frequently than physicians (P=.005). DIS use centered on quality improvement and identifying high-risk patients, with differences by role. Physicians were more inclined to use DIS out of curiosity (P=.01). Participants desired DIS for patient follow-up, lifestyle guidance, treatment suggestions, reminders, and shared decision-making. Administrators favored predictive analysis (P<.001), while physicians resisted immediate patient identification (P=.03). The 5 innovation attributes showed high internal consistency (α>.7). These factors explained 78.5% of questionnaire variance, relating to complexity, competitive advantage, compatibility, trialability, and observability. Factors 2, 3, and 4 predicted intention to use DIS, with factor 2 alone achieving the best accuracy (root-mean-square=0.513).

Conclusions: Administrators and physicians exhibited role-based DIS use patterns highlighting the need for tailored approaches to promote DIS adoption. The study reveals a link between positive perceptions and intention to use DIS, emphasizing the significance of considering all factors for successful health care integration. The results suggest various directions for future studies. These include refining the trialability and observability questions for increased reliability and validity, investigating a larger sample with more specific target groups to improve generalization, and exploring the relevance of different groups' perspectives and needs in relation to decisions about and use of DIS.

(Interact J Med Res 2023;12:e49973) doi:10.2196/49973

KEYWORDS

digital information systems; implementation; primary care; health care professionals; information system; information systems; usability; adoption; perception; perceptions; technology use; perspective; perspectives

Introduction

The wide adoption of electronic health records and general IT infrastructure in primary care facilitates both digital documentation and access to information. This digital infrastructure can be leveraged to guide evidence-based care and help monitor patient safety and quality of care at the individual or group level [1].

Primary care providers must track a number of quality indicators specified by regulatory authorities. These quality indicators are used to monitor and compare the quality of care across local and regional organizational boundaries [2,3]. Thanks to the wide adoption of electronic health records, these indicators can be automatically computed and presented to primary care staff [4]. In Sweden, these indicators are part of a digital platform called "Quality of Primary Care" (swe. Primärvårdskvalitet), to which the vast majority of primary care providers have access [5]. This platform is an example of a digital information system (DIS).

The purpose of DIS in health care is to facilitate the visualization and analysis of the data contained in electronic health records and administrative data both on an individual and population level [6,7]. In addition to presenting information such as quality indicators, these systems may also provide clinical decision support tools. The use of DIS among administrators, doctors, and nurses includes a spectrum of diverse purposes, intricately intertwined with their distinct roles, specialized training, individual preferences, and intrinsic motivations. This multiplicity of factors influences the unique ways in which each group engages with these systems, shaping their interactions, decision-making processes, and overall contributions within the health care ecosystem [8,9]. They may also have distinct perspectives on the value of DIS, both within their daily duties and the clinic's overall functioning. These viewpoints arise from factors like duties and tasks, desired outcomes, familiarity with technologies, and personal experiences. This variance of perspectives shapes the perception and integration of digital systems, influencing clinical operational efficiency [10-12]. Understanding these different roles and perceptions is essential to developing systems that work well in practice and fulfill users' needs [13,14].

In general, the introduction or implementation of a new DIS is difficult [12,13]. Oftentimes, they do not add enough value, they are too generic and not aligned with local work processes, or they do not consider practical barriers to implementation [15,16]. It is also unknown how DIS are used in practice, by whom, how often, and to what purpose, as well as how they are perceived and what innovative features DIS users wish to see in the future [17]. The introduction of an innovation such as DIS can be understood and studied as part of a process toward adoption and diffusion. In relation to a specific innovation and the intended setting and context for its introduction, there are several factors that affect the adoption process and the diffusion within and across settings [18]. If such factors are understood they can be targeted to facilitate and improve both the success and rate of adoption and diffusion. In this context, 5 attributes of innovation are decisive—relative advantage, compatibility, complexity, trialability, and observability—and have been respectively defined as "the degree to which an innovation is perceived as being better than the idea it supersedes," "the degree to which an innovation is perceived as consistent with the existing values, past experiences, and needs of potential adopters," "the degree to which an innovation is perceived as relatively difficult to understand and use," "the degree to which an innovation may be experimented with on a limited basis," and "the degree to which the results of an innovation are visible to others" [18]. In the context of introducing DIS to improve health care, it is therefore of great importance to understand how DIS are perceived by the stakeholders affected by its introduction, with respect to its different attributes and in relation to different intended uses [12].

This study adopts a pragmatic approach using a familiar DIS within primary care to prompt reflection regarding use and potential development and future applications. Consequently, the study aims to explore how DIS are used and perceived by health care professionals in primary care.

Methods

Study Design

The study has an exploratory design and is based on quantitative assessment, using descriptive statistics, of survey data collected during April-May 2022 in accordance with the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) [19].

Participants

We wanted to reach a large number of primary care personnel in various roles. To make it feasible to reach users of DIS we used two approaches for the distribution of a closed web-based survey: (1) directly emailing publicly listed contact persons at primary care clinics when that information was available and (2) requesting an interest organization to distribute the link to the questionnaire to its members or networks. While using this pragmatic approach enabled the creation of a convenience sample comprising a larger respondent count, it constrained our outreach predominantly to administrators and physicians.

From publicly available information on the web, we curated a total of 44 contact emails from primary care health centers in a local region in southern Sweden and 64 contact emails from a private primary care provider across the country. We contacted the 25 regional chapters of The Swedish Association of General Practice as well as the Innovation Platform at Region Västra Götaland and the Digital Well Arena in Region Värmland and asked them to distribute our questionnaire to people working in the primary care sector.

Data Collection

The digital questionnaire was kept as short as possible so that it would take less than 5 minutes to answer. This was a pragmatic choice intended to maximize the number of respondents in a short period of time. The questionnaire was structured in 3 sections developed for the purpose of the study (Multimedia Appendix 1). It included both self-developed questions of a descriptive nature (sections 1-2) and questions based on a previously validated instrument (section 3). The three sections covered were as follows:

1. Respondent characteristics, including role, whether they work at a private or public provider, and their geographical region (3 questions);
2. Use of current platforms, including access to digital tools, frequency of use, and reason for use (18 questions);
3. Perceptions of decision support tools, including unmet needs, perceptions, and beliefs about future DIS (12 questions).

Most questions were multiple choice on an ordinal scale. Sections 2 and 3 asked about the respondents' level of agreement or disagreement with a number of statements on a 5-level ordinal scale: completely agree, partially agree, indifferent, partially disagree, or completely disagree. Section 3 was developed based on Roger 5 attributes of innovation [16]—relative advantage, compatibility, complexity, trialability, and observability of results—and used a previously developed survey as template for the development [17]. To assess each of these attributes, 2 statements were developed. In addition, 2 final statements were designed to assess the respondents' intention to use DIS.

Statements were developed from previous interviews with primary care physicians, primary care administrators, medical experts in the Life Science industry, and technologists developing decision support tools. The full questionnaire is available in Multimedia Appendix 1. The questionnaire was reviewed for face validity by technologists working with clinical decision support tools, life science professionals, and health innovation and implementation researchers. The answers from the first-week respondents (N=13) were used to check that the list of roles was representative and that they were not indifferent to all items. The questionnaire was updated with additional terms for roles in the first question to better reflect what the first respondents had written under the item "other." No other adjustments were made.

Participants were, during data collection, presented with all questions regardless of their responses to preceding queries. All questions were obligatory, and participants were provided with the alternative to respond "Don't know" or "Neither." Participants were allowed to review their responses before submitting the survey. The data were checked for duplicated entries to ensure that participants had not answered the questionnaire twice by accident. The usability and technical functionality of the electronic questionnaire were tested internally before fielding the questionnaire.

Data Analysis

Questionnaire answers were analyzed with descriptive statistics. Each section was analyzed for differences in answers according to roles (administrators vs physicians). This was done using the Mann-Whitney Wilcoxon Test and a significance level of 0.05, to test the null hypothesis that answers from the 2 different groups follow the same distribution. The test chosen is a nonparametric rank-sum test that has been shown to perform robustly even for small sample sizes of 10 or fewer observations [20,21].

Section 3 of the questionnaire was analyzed for internal consistency, in this case, the level of agreement between items related to the same attribute. This was done using ordinal coefficient α based on a polychoric correlation matrix [22,23].

Confirmatory factor analysis with varimax rotation and 5 factors was also performed on section 3 of the questionnaire to ensure that Roger 5 attributes were indeed captured in the questions. The 2 items related to intention to use presented high internal consistency and were averaged into one factor. The resulting factors were then used to predict the respondent's intention to use DIS. A linear regression model was trained using 50 randomly selected entries. Model performance was tested with the remaining 20 entries. The accuracy of the model was evaluated visually and with root-mean-square (RMS) errors. All statistical analysis was done using RStudio (version 2023.03.0+386, Posit PBC).

Ethical Considerations

The study adheres to the principles outlined in the Declaration of Helsinki and fulfilled the following research requirements: information, consent, confidentiality, and participant safety. Ethical approval for the research was not formally required under Swedish law, as no personal or sensitive information was handled. Each participant received written information encompassing the study's objectives and inception, outlining their role in the study, clarifying the collection of exclusively anonymous data, and delineating the methods for data collection and storage. They were also informed about the voluntary nature of participation, confidentiality, and the option to withdraw their consent at any point, without the need for justification.

Results

We received a total of 70 responses across 10 of the 21 regions of Sweden (Table 1). Of the 69 respondents informing on their workplace and role, 31.9% (n=22) worked at private clinics and 68.1% (n=47) worked at public clinics and, 26.1% (n=18) worked as administrators or head of operations, 1.4% (n=1) had the role as a medical strategist, and the remaining 72.5% (n=50) were physicians. Of the respondents, 11.6% (n=8) used DIS on a weekly basis, 24.6% (n=17) used DIS on a monthly basis, 27.5% (n=19) used DIS a few times a year, 26.1% (n=18) used DIS very rarely, and 10.1% (n=7) said they did not have access to DIS. However, the frequency of use differed significantly by role. Administrators were more likely to use DIS more frequently than physicians (*P*=.005; Table 2).

Table 1. Regional distribution of respondents (N=70).

Region	Respondents, n (%)
Kronoberg	17 (24.3)
Halland	17 (24.3)
Skåne	12 (17.1)
Västra Götaland	8 (11.4)
Uppsala	6 (8.6)
Blekinge	5 (7.1)
Gotland	2 (2.9)
Kalmar	1 (1.4)
Gävleborg	1 (1.4)
Dalarna	1 (1.4)

Table 2. Frequency of use depending on participants' roles (N=68; 2 participants did not inform on both what their role was and frequency of use and were excluded from this table).

	Administrators (N=18), n (%)	Physicians (N=50), n (%)
No access to digital information systems	0	7 (14.0)
Rarely	2 (11.1)	16 (32.0)
Yearly	5 (27.8)	14 (28.0)
Monthly	9 (50.0)	8 (16.0)
Weekly	2 (11.1)	5 (10.0)

In general, a large portion of participants indicated that DIS are not used to follow individual patients' care journeys. On the other hand, a large portion of participants indicated that DIS are used for guiding quality improvement activities and for identifying high-risk or high-cost patients. Differences in the use of DIS were aligned with specific roles. Administrators were significantly more likely than physicians to use DIS for reporting to authorities (*P*=.01), for developing activities to improve quality of care (*P*=.02), and for planning budget and staffing (*P*<.001). On the other hand, physicians were significantly more likely than administrators to use DIS out of curiosity (*P*=.01; Figure 1).

In general, participants indicated that they would like to have DIS to support the follow-up of patients, lifestyle changes for patients, suggestions for treatment options, and reminders for patients to follow treatment, as well as making shared decisions with patients. None of the suggested future developments were seen as negative by either of the roles. However, there were a few areas where administrators and physicians differed in their opinion (Figure 2). For example, administrators were more positive toward predictive analysis of care needs to facilitate planning of budget and staffing (*P*<.001). Physicians on the other hand were more negative toward identifying high-risk or high-cost patients as soon as they visit the clinic (*P*=.03).

Section 3 of the questionnaire included 2 items for each of Roger 5 attributes of innovation: competitive advantage, complexity, compatibility, trialability, and observability. The answers to each attribute were evaluated for internal consistency using the ordinal reliability coefficient α. Starting from the assumption that several items measure the same latent variable, the reliability coefficient indicates how consistent those items are as a group. An α above .9 is considered excellent, whereas an α between .7 and .8 is considered acceptable. Items related to competitive advantage (α=.94), complexity (α=.96), and compatibility (α=.90) presented excellent internal consistency. Trialability (α=.73) and observability (α=.72) presented acceptable internal consistency, indicating a certain dissonance between the 2 items related to these 2 attributes. The additional 2 items referring to intention to use also presented excellent internal consistency (α=.93).

Figure 1. Participants' responses to statements on their use of digital information systems in their everyday practice in response to the question "How well do the following statements agree with why you use the digital tool today?".

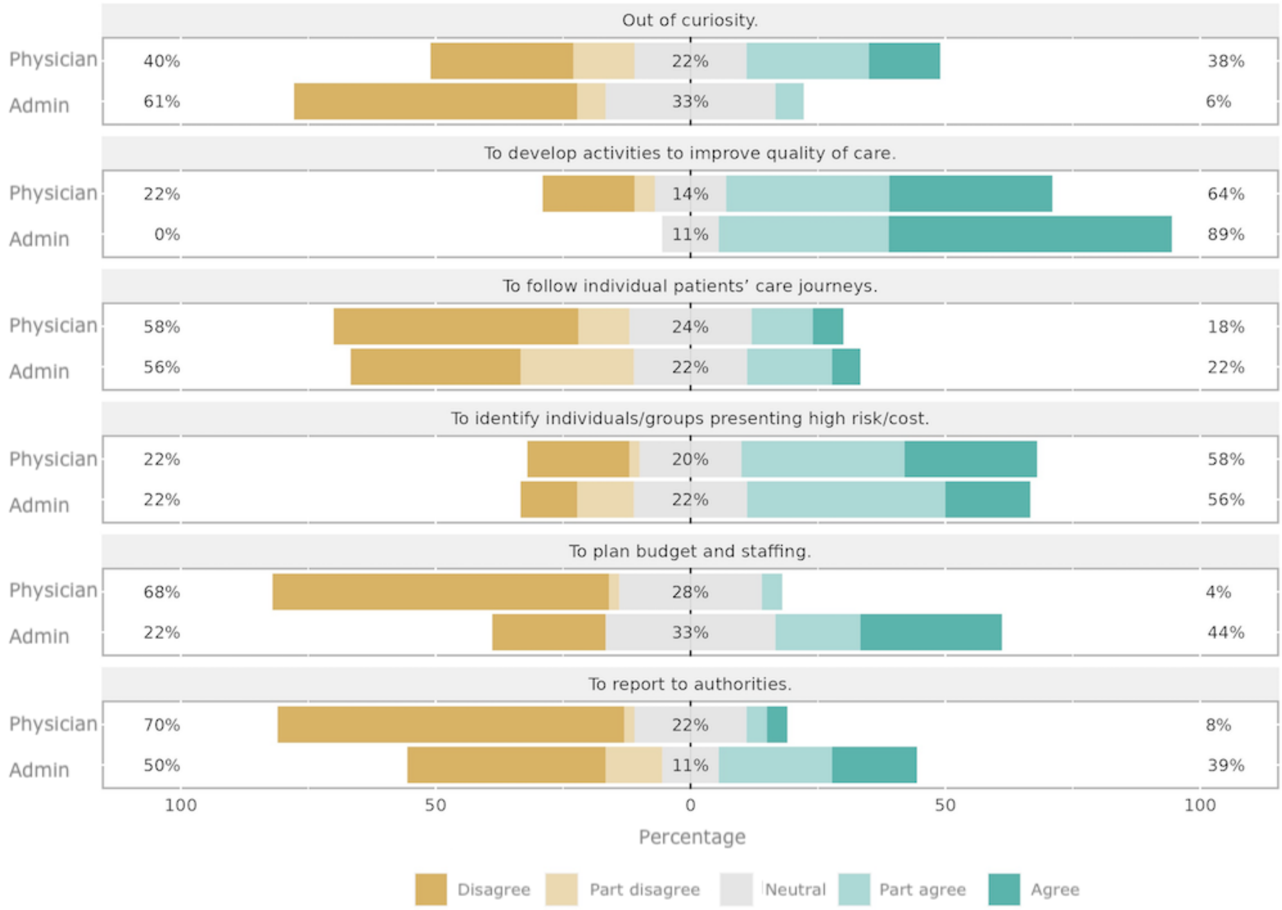
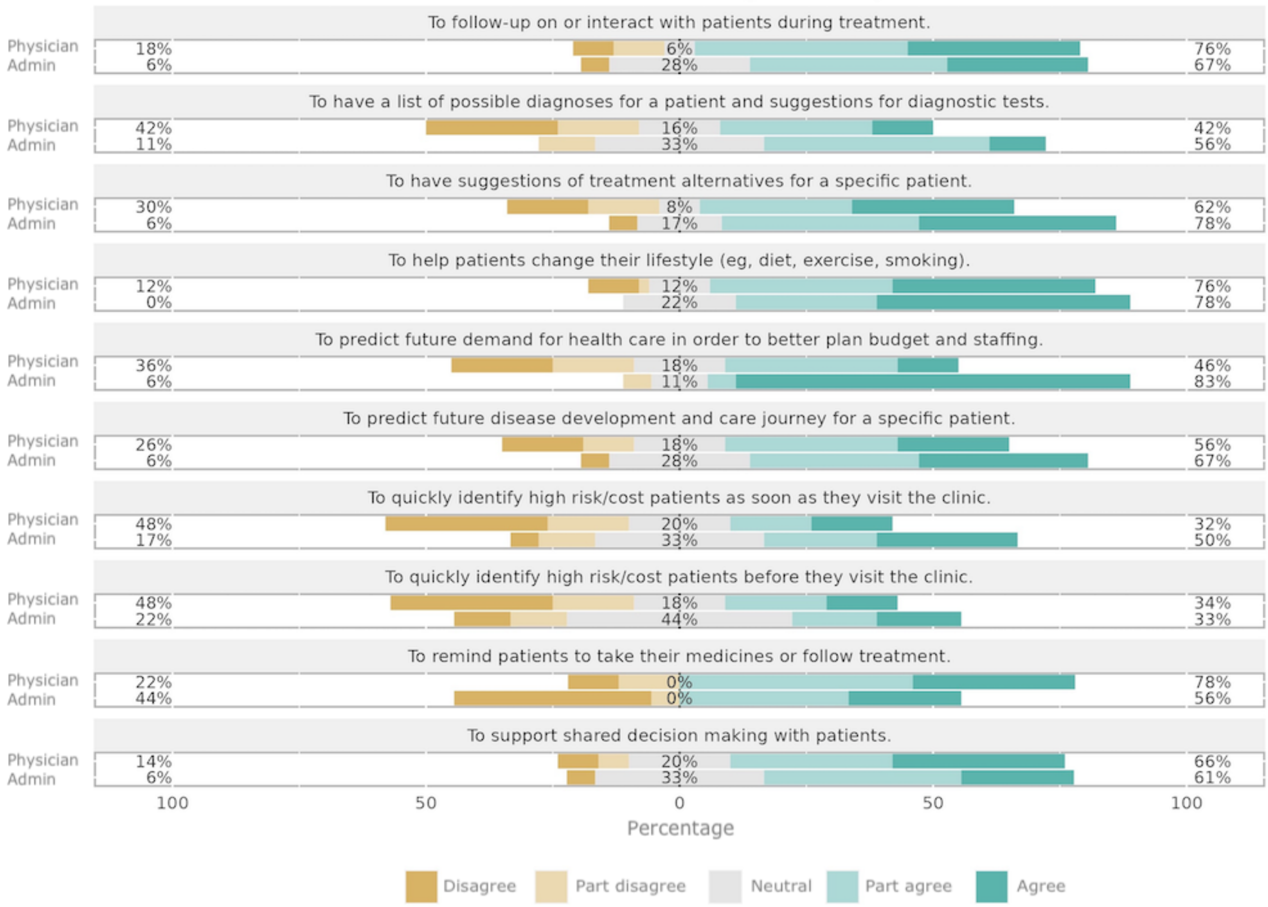


Figure 2. Participants' statements on their use of digital information systems in their everyday practice in response to the question "How well do the following statements agree with your needs?".



The answers to section 3 of the questionnaire were decomposed into 5 factors to test the hypothesis that these questions capture the 5 innovation attributes. Indeed, results show that the resulting 5 factors explain 78.5% of the variance in the data. In addition, each resulting factor was highly linked to one of the five attributes: competitive advantage 1 and 2 (0.090 and 0.247), complexity 1 and 2 (0.005 and 0.218), compatibility 1 and 2 (0.094 and 0.318), trialability 1 and 2 (0.550 and 0.005), and observability 1 and 2 (0.618 and 0.005). We may therefore refer to factor 1 as mostly representing complexity, factor 2 as representing competitive advantage, factor 3 as representing compatibility, factor 4 as representing trialability, and factor 5 as representing observability. For trialability and observability, loadings were high for 1 item but relatively low for the other item. This is consistent with the slightly lower ordinal reliability coefficient for those items, and the high values of uniqueness of variables trialability 1 and observability 1 (Table 3). This indicates that these variables contain unique information that does not conform with the 5 resulting factors.

Administrators were significantly more positive than doctors about DIS being compatible with their work processes (P=.04). In addition, administrators were significantly more positive than physicians in their intention to use DIS (P=.01). Intention to use did not significantly differ between respondents who used DIS frequently (ie, weekly or monthly) and the remaining respondents (P=.30; Figure 3).

All factors are positively correlated with the intention to use. Pearson correlation coefficient was highest for factor 2 (r=0.85), followed in order by factor 3 (r=0.78), factor 5 (r=0.74), factor 4 (r=0.73), and factor 1 (r=0.64). Intention to use was predicted from the 5 factors, starting with factor 2, which had the highest correlation, and consecutively adding more factors in order of correlation. RMS error was used as a measure of accuracy for each model. The best accuracy was obtained using factor 2 only (RMS=0.513), followed by a model using all 5 factors (RMS=0.524). Accuracy results were slightly worse for a model using factors 2 and 3 (RMS=0.523), followed by a model including factors 2, 3, 4, and 5 (RMS=0.532). The worst-performing model included factors 2, 3, and 5 (RMS=0.533).

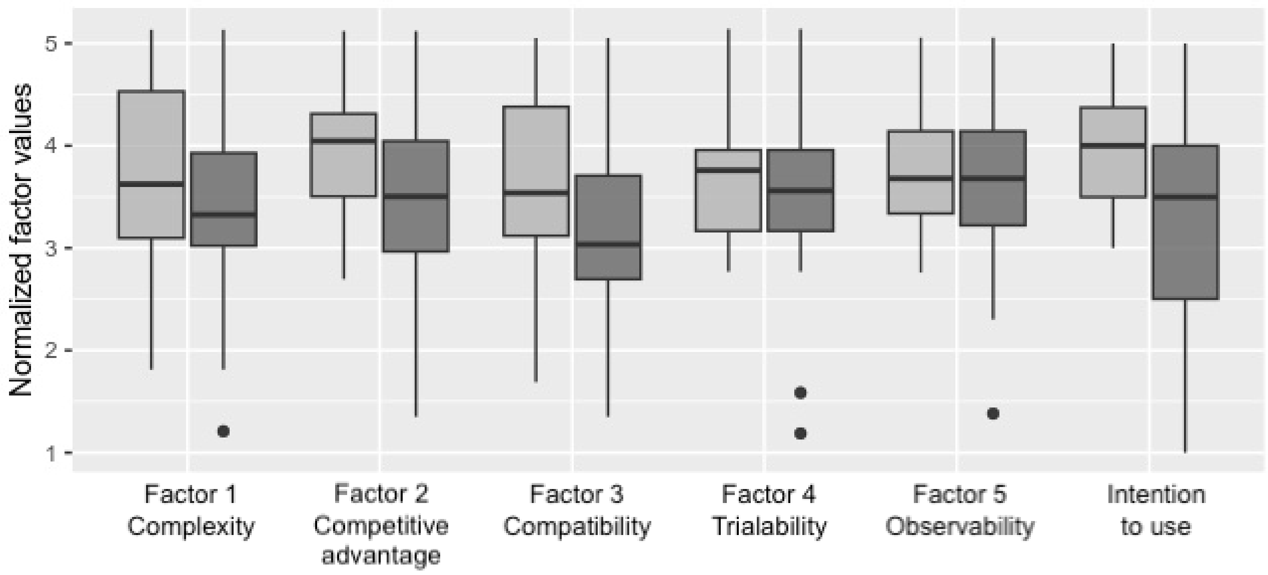
Table 3. Factor loadings of questions in relation to 5 innovation attributes (bold values indicate items with high faction loadings).

Loadings	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Competitive advantage 1	0.158	0.882[a]	0.259	0.170	0.108
Competitive advantage 2	0.224	0.709[a]	0.390	—[b]	0.213
Complexity 1	0.956[a]	—	0.246	0.133	—
Complexity 2	0.820[a]	0.160	0.261	0.125	—
Compatibility 1	0.359	0.342	0.785[a]	0.208	—
Compatibility 2	0.425	0.286	0.632[a]	—	0.124
Trialability 1	0.272	0.404	0.131	0.378[a]	0.230
Trialability 2	0.184	0.182	0.140	0.929[a]	0.213
Observability 1	–0.139	0.450	—	0.246	0.310
Observability 2	—	0.259	—	0.222	0.934[a]

[a]Values indicate items with high faction loadings.

[b]Not available.

Figure 3. Perceptions decision support of digital information systems by role, administrators (light gray) and physicians (dark gray).



Discussion

Principal Findings

In this study, we applied an approach to broadly reach out to users of DIS at a national level. A digital survey with questions about the participants' characteristics and demographics, current use, and perceptions of DIS was distributed through emails directly to potential participants and through email lists and networks. A limited number of responses were received, but still sufficient to address the study's research questions. One-third of the participants held roles as administrators and two-thirds as physicians. Almost two-thirds used DIS often or a few times a year, while approximately one tenth had no access to DIS at all, and thereby responded to the questionnaire based on perceptions rather than own experiences. The administrators used the DIS more often and both the administrators' and the physicians' use followed purposes based on their professional roles, that is, more administrative purposes and more clinical purposes respectively.

Both roles were very positive about using DIS to improve the quality of care and to identify patients and groups at high risk and high costs. In cases where there was a disinterest in the use of DIS, it was associated with the application not being linked to the professional role or work tasks. For example, administrators were less interested in using DIS out of pure curiosity, while physicians were less interested in applying DIS for budget work and staffing. The only application area that was seen as predominantly negative for both roles was to follow individual patients' care pathways. Given the value for both administrators, regarding risk assessment and cost calculation; and doctors, regarding opportunities to improve quality in clinical decisions and care planning, one could expect more positive attitudes toward this purpose of use. Whether this lower interest was due to uncertainty about the meaning of applying DIS to follow individual patients' care pathways, or that, for example, the application was deemed not to be consistent with work tasks or the professional roles of the participants, remains to be investigated. Overall, these findings align with previous research that the attitudes toward using DIS in health care are intricately shaped by a multitude of factors, such as the application area, the professional role, and work tasks. These factors include not only the specific application domains but also the distinct professional roles and inherent tasks of individuals within the health care context [8,9,12]. The findings from this study and existing knowledge underscore the interplay between contextual variables and illustrate the essential role they play in shaping the attitudes and inclinations toward integrating DIS within health care settings [10,12]. To promote the use of DIS in health care, this study and other research in this field [18,24] suggest it may be important to consider these factors and tailor the application of DIS to the needs of different stakeholders.

In all contexts of the introduction of new innovations, stakeholders' perceptions of the innovation have a decisive role in the innovation's successful adoption and diffusion within the organization, and to other similar application areas in other organizations [18]. In order to succeed in introducing DIS in clinical organizations with the aim of improving health care, it is thus important to understand how DIS is perceived by the stakeholders affected by its introduction. To investigate this, we developed a self-assessment instrument for stakeholders' perceptions of DIS based on Roger 5 innovation attributes [18]. The instrument items were constructed based on a previous study of attitudes toward the integration of social media in university education [25]. The construction of 2 items to represent each innovation attribute was validated through an assessment of their internal consistency and factor loadings in relation to the comprehensive spectrum of 5 attributes. The subsequent correlation analysis, linking participant responses with their intentions to use DIS, unveiled a noteworthy positive alignment between higher-rated perceptions of DIS's innovation attributes and an elevated intention to engage with the system. This suggests that a swift preliminary evaluation of stakeholders' favorable attitudes and use intentions can be achieved by evaluating the perceived relative advantage of DIS. However, to attain a deeper and more nuanced understanding, it is prudent to holistically examine all 5 of Roger's proposed innovation adoption factors in conjunction. This approach ensures a comprehensive exploration of the different dynamics influencing stakeholders' inclinations to adopt DIS based on their perceived attributes and attitudes and could facilitate a more informed and contextually sensitive evaluation of their intentions.

Limitations

The study faced a challenge in terms of reaching a sufficient number of participants, and the response rate was relatively low. The findings may thus not fully represent the views and experiences of the entire population of interest, and there could be some sampling bias. More than half of the respondents reported using DIS often. While this may reflect the growing trend of digitalization in health care, it could also mean that the sample was overrepresented with participants with positive experiences and attitudes toward DIS and not representative of the general population, which may include individuals who have limited access or skills in using digital technologies. The study measured the intention to use DIS among both users and nonusers. However, this approach could introduce some variability in the results, as those who have prior experience with such systems may have different motivations and expectations compared to those who have only a perception or awareness of these systems. Therefore, the findings should be interpreted with caution and may not apply to all potential users.

Conclusions

In conclusion, this study used a comprehensive approach to engage a diverse range of DIS users at a national level. The survey responses provided valuable insights despite the limited number received, addressing key research questions. Administrators and physicians demonstrated varying patterns of DIS use aligned with their professional roles, reflecting distinct attitudes toward different application areas. Notably, DIS was positively embraced for quality enhancement and identifying high-risk patients, with concerns arising mainly in relation to following individual care pathways. Stakeholders' perceptions of DIS, influenced by attributes such as professional roles and application contexts, underscore the need for tailored approaches in promoting DIS adoption within health care. By drawing from Roger's innovation attributes, this study developed an assessment instrument to gauge stakeholder perceptions, revealing a correlation between positive perceptions and intention to use DIS. Considering all innovation factors is essential for a comprehensive understanding of stakeholders' attitudes toward DIS integration, offering valuable insights for successful adoption and diffusion in health care settings. Based on the results, there are several potential avenues for future studies. For example, the trialability and observability questions could be refined to improve their reliability and validity. Additionally, a larger sample with more defined target groups could be investigated to enhance the generalization of the findings. Finally, more clearly defined relevance of the groups with respect to decisions about and use of DIS could be explored to gain a deeper understanding of their perspectives and needs.

Acknowledgments

This study was partially funded by CSAM Carmona AB and Halmstad University. Generative artificial intelligence was not used for manuscript writing apart from translation purposes.

Data Availability

The data sets generated or analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

Study conception and design were formulated by AS and JN. AS was responsible for data collection. Analysis and interpretation of results were done by AS and JN. AS and JN collaborated in the preparation of draft paper. Both authors reviewed the results and approved the final version of the paper.

Conflicts of Interest

This study was partially sponsored by a private organization that provides digital decision-support tools and data management solutions to care providers and researchers. The work was, however, conducted independently. All analyses and results presented here were not subject to any influence by the sponsor.

Multimedia Appendix 1
Digital questionnaire developed for the purpose of the study.
[PDF File (Adobe PDF File), 108 KB - ijmr_v12i1e49973_app1.pdf]

References

1. Atasoy H, Greenwood BN, McCullough JS. The digitization of patient care: a review of the effects of electronic health records on health care quality and utilization. Annu Rev Public Health 2019;40(1):487-500 [FREE Full text] [doi: 10.1146/annurev-publhealth-040218-044206] [Medline: 30566385]
2. Cinaroglu S, Baser O. Understanding the relationship between effectiveness and outcome indicators to improve quality in healthcare. Total Qual Manag Bus Excell 2016;29(11-12):1294-1311 [FREE Full text] [doi: 10.1080/14783363.2016.1253467]
3. Horvat A, Filipovic J. Healthcare system quality indicators: the complexity perspective. Total Qual Manag Bus Excell 2018;31(1-2):161-177. [doi: 10.1080/14783363.2017.1421062]
4. Essén A, Scandurra I, Gerrits R, Humphrey G, Johansen MA, Kierkegaard P, et al. Patient access to electronic health records: differences across ten countries. Health Policy Technol 2018;7(1):44-56 [FREE Full text] [doi: 10.1016/j.hlpt.2017.11.003]
5. Sveriges Kommuner och landsting (SKL). Primärvårdskvalitet—ett stöd för kvalitetsarbete i primärvården. URL: https://skl.se/Primarvardskvalitet [accessed 2023-10-05]
6. Saigí-Rubió F, Pereyra-Rodríguez JJ, Torrent-Sellens J, Eguia H, Azzopardi-Muscat N, Novillo-Ortiz D. Routine health information systems in the European context: a systematic review of systematic reviews. Int J Environ Res Public Health 2021;18(9):4622 [FREE Full text] [doi: 10.3390/ijerph18094622] [Medline: 33925384]
7. Epizitone A, Moyane SP, Agbehadji IE. Health information system and health care applications performance in the healthcare arena: a bibliometric analysis. Healthcare (Basel) 2022;10(11):2273 [FREE Full text] [doi: 10.3390/healthcare10112273] [Medline: 36421597]
8. Ahmadian L, Dorosti N, Khajouei R, Gohari SH. Challenges of using hospital information systems by nurses: comparing academic and non-academic hospitals. Electron Physician 2017;9(6):4625-4630 [FREE Full text] [doi: 10.19082/4625] [Medline: 28848639]
9. Abdekhoda M, Ahmadi M, Gohari M, Noruzi A. The effects of organizational contextual factors on physicians' attitude toward adoption of electronic medical records. J Biomed Inform 2015;53:174-179 [FREE Full text] [doi: 10.1016/j.jbi.2014.10.008] [Medline: 25445481]
10. Marwaha JS, Landman AB, Brat GA, Dunn T, Gordon WJ. Deploying digital health tools within large, complex health systems: key considerations for adoption and implementation. NPJ Digit Med 2022;5(1):13 [FREE Full text] [doi: 10.1038/s41746-022-00557-1] [Medline: 35087160]
11. Moore EC, Tolley CL, Bates DW, Slight SP. A systematic review of the impact of health information technology on nurses' time. J Am Med Inform Assoc 2020;27(5):798-807 [FREE Full text] [doi: 10.1093/jamia/ocz231] [Medline: 32159770]
12. Petersson L, Larsson I, Nygren JM, Nilsen P, Neher M, Reed JE, et al. Challenges to implementing artificial intelligence in healthcare: a qualitative interview study with healthcare leaders in Sweden. BMC Health Serv Res 2022;22(1):850 [FREE Full text] [doi: 10.1186/s12913-022-08215-8] [Medline: 35778736]
13. Nilsen P, Seing I, Ericsson C, Birken SA, Schildmeijer K. Characteristics of successful changes in health care organizations: an interview study with physicians, registered nurses and assistant nurses. BMC Health Serv Res 2020;20(1):147 [FREE Full text] [doi: 10.1186/s12913-020-4999-8] [Medline: 32106847]
14. Stanley DJ, Meyer JP, Topolnytsky L. Employee cynicism and resistance to organizational change. J Bus Psychol 2005;19(4):429-459. [doi: 10.1007/s10869-005-4518-2]
15. Augustsson H, Morici BC, Hasson H, von Thiele Schwarz U, Schalling SK, Ingvarsson S, et al. National governance of de-implementation of low-value care: a qualitative study in Sweden. Health Res Policy Syst 2022;20(1):92 [FREE Full text] [doi: 10.1186/s12961-022-00895-2] [Medline: 36050688]
16. Kraus S, Schiavone F, Pluzhnikova A, Invernizzi AC. Digital transformation in healthcare: analyzing the current state-of-research. J Bus Res 2021;123:557-567 [FREE Full text] [doi: 10.1016/j.jbusres.2020.10.030]

17. Van Velthoven MH, Cordon C. Sustainable adoption of digital health innovations: perspectives from a stakeholder workshop. J Med Internet Res 2019;21(3):e11922 [FREE Full text] [doi: 10.2196/11922] [Medline: 30907734]
18. Rogers E. Diffusion of Innovations, 5th Edition. Manhattan, New York City, New York, United States: Simon and Schuster; 2003.
19. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
20. Bridge PD, Sawilowsky SS. Increasing physicians' awareness of the impact of statistics on research outcomes: comparative power of the t-test and and Wilcoxon Rank-Sum test in small samples applied research. J Clin Epidemiol 1999;52(3):229-235. [doi: 10.1016/s0895-4356(98)00168-1] [Medline: 10210240]
21. Divine G, Norton HJ, Hunt R, Dienemann J. Statistical grand rounds: a review of analysis and sample size calculation considerations for Wilcoxon tests. Anesth Analg 2013;117(3):699-710. [doi: 10.1213/ANE.0b013e31827f53d7] [Medline: 23456667]
22. Zumbo BD, Gadermann AM, Zeisser C. Ordinal versions of coefficients alpha and theta for likert rating scales. J Mod Appl Stat Methods 2007;6:21-29 [FREE Full text] [doi: 10.22237/jmasm/1177992180]
23. Gadermann AM, Guhn M, Zumbo BD. Estimating ordinal reliability for likert-type and ordinal item response data: a conceptual, empirical, and practical guide. Pract Assess Res Evaluation 2019;17:3 [FREE Full text]
24. Gagnon MP, Desmartis M, Labrecque M, Car J, Pagliari C, Pluye P, et al. Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. J Med Syst 2012;36(1):241-277 [FREE Full text] [doi: 10.1007/s10916-010-9473-4] [Medline: 20703721]
25. Almohtadi RM, Aldarabah IT. University students' attitudes toward the formal integration of facebook in their education: investigation guided by rogers' attributes of innovation. World J Edu 2021;11(1):20-28 [FREE Full text] [doi: 10.5430/wje.v11n1p20]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
DIS: digital information systems
RMS: root-mean-square

Edited by T de Azevedo Cardoso; submitted 15.06.23; peer-reviewed by A Epizitone, S Nilsson; comments to author 29.07.23; revised version received 01.09.23; accepted 27.09.23; published 25.10.23.

Please cite as:
Sant'Anna A, Nygren J
A Pragmatic Mapping of Perceptions and Use of Digital Information Systems in Primary Care in Sweden: Survey Study
Interact J Med Res 2023;12:e49973
URL: https://www.i-jmr.org/2023/1/e49973
doi:10.2196/49973
PMID:37878357

Original Paper

COVID-19 in Vietnam and Its Impact on Road Trauma: Retrospective Study Based on National Data

Ba Tuan Nguyen[1], MD; Christopher Leigh Blizzard[1], DPhil, Prof Dr; Andrew Palmer[1], DPhil; Huu Tu Nguyen[2], MD, DPhil, Prof Dr; Thang Cong Quyet[2], MD, DPhil, Prof Dr; Viet Tran[1,3,4], MBBS; Mark Nelson[1], MBBS, DPhil

[1]Menzies Institute for Medical Research, University of Tasmania, Hobart, Australia

[2]Hanoi Medical University, Hanoi, Vietnam

[3]Tasmanian School of Medicine, University of Tasmania, Hobart, Australia

[4]Tasmanian Health Service, Hobart, Australia

Corresponding Author:
Ba Tuan Nguyen, MD
Menzies Institute for Medical Research
University of Tasmania
17 Liverpool Street
Hobart, 7000
Australia
Phone: 61 460918919
Email: batuan.nguyen@utas.edu.au

Abstract

Background: Despite significant improvement in the last decade, road trauma remains a substantial contributor to deaths in Vietnam. The COVID-19 pandemic necessitated public health measures that had an unforeseen benefit on road trauma in high-income countries. We investigate if this reduction was also seen in a low- to middle-income country like Vietnam.

Objective: Our aim was to investigate how the COVID-19 pandemic and the government policies implemented in response to it impacted road trauma fatalities in Vietnam. We also compared this impact to other government policies related to road trauma implemented in the preceding 14 years (2007-2020).

Methods: COVID-19 data were extracted from the Vietnamese Ministry of Health database. Road traffic deaths from 2007 to 2021 were derived from the Vietnamese General Statistical Office. We used Stata software (version 17; StataCorp) for statistical analysis. Poisson regression modeling was used to estimate trends in road fatality rates based on annual national mortality data for the 2007-2021 period. The actual change in road traffic mortality in 2021 was compared with calculated figures to demonstrate the effect of COVID-19 on road trauma fatalities. We also compared this impact to other government policies that aimed to reduce traffic-related fatalities from 2007 to 2020.

Results: Between 2007 and 2020, the number of annual road traffic deaths decreased by more than 50%, from 15.3 to 7 per 100,000 population, resulting in an average reduction of 5.4% per annum. We estimated that the road traffic mortality rate declined by 12.1% (95% CI 8.9-15.3%) in 2021 relative to this trend. The actual number of road trauma deaths fell by 16.4%. This reduction was largely seen from August to October 2021 when lockdown and social distancing measures were in force.

Conclusions: In 2021, the road traffic–related death reduction in Vietnam was 3 times greater than the trend seen in the preceding 14 years. The public health response to the COVID-19 pandemic in Vietnam was associated with a third of this reduction. It can thus be concluded that government policies implemented to address the COVID-19 pandemic resulted in a 4.3% decrease in road traffic deaths in 2021. This has been observed in high-income countries, but we have demonstrated this for the first time in a low- and middle-income country.

(Interact J Med Res 2023;12:e40883) doi:10.2196/40883

KEYWORDS

COVID-19; impact; road trauma; low- and middle-income country; LMIC; mortality; pandemic; trauma; social distancing; lockdown; Vietnam; disease; policy; deaths

Introduction

Background

COVID-19 is an infectious disease caused by the SARS-CoV-2 virus, which was discovered in Wuhan, China, on December 31, 2019 [1]. In Vietnam, the first confirmed COVID-19 case was recorded in late January 2020 [2]. During that year, while most countries around the world struggled against the pandemic, Vietnam was considered one of the safest places in terms of COVID-19 community transmission [3]. This was due to a strict "zero-COVID" government policy that included closure of the Vietnam-China border, restrictions to air travel from any affected countries, quarantining of infected patients and all close contacts, mandatory use of masks in public places, closure of all nonessential services, and use of rapid antigen or real-time–polymerase chain reaction testing for all suspected cases [4]. However, this changed quickly in the following year. While most affected countries were reopening, a wave of COVID-19 spread across Vietnam. There were 4 waves of COVID-19 outbreak in Vietnam [5], but the fourth outbreak, which started in April 2021 and was caused by the Delta variant, was the most catastrophic [6]. To control this situation, as was done in other countries [7], social distancing and lockdown strategies were applied in Vietnam by the central government to local government areas according to their disease burden [8]. The Vietnamese government placed approximately a third of the national population (including the 2 largest cities, Hanoi and Ho Chi Minh City) under lockdown starting on July 23, 2021. This lasted for about 3 months and was gradually relaxed in November 2021 as the pandemic was brought under control.

Another contemporary "pandemic" is road trauma, which has been a significant contributor to morbidity and mortality in Vietnam and other low- and middle-income countries (LMICs) for decades [9]. The risk of road traffic death in LMICs over the last decade was more than 3 times higher than in high-income countries (HICs) such as the United States, United Kingdom, and Australia [10]. In the last 15 years in Vietnam, approximately 9000 people per annum have been killed in road accidents, with an equal number hospitalized [11,12]. This problem has been driven by poor infrastructure and the dominance of the motorcycle as the mode of transport, often overladen with passengers and cargo [9]. To address this problem, the Vietnamese government legislated multiple public health measures. In 2008, helmets became mandatory for all motorcycle riders and their passengers [13]. Mandatory seatbelt use in other vehicles was introduced in the same year [13]. Media campaigns and public traffic law education have also been ubiquitous since this time. Subsequently, road deaths have steadily declined [14].

The lockdown and social distancing measures were implemented to reduce transmission and death due to COVID-19 but may have had an unforeseen benefit on road trauma. For example, social distancing may limit the number of passengers in an individual vehicle and, hence, reduce the number of people exposed to trauma should that vehicle crash. Lockdowns also reduce exposure since population movements are limited to essential travel and cross-border movements are restricted. Reductions have been seen in most HICs [15]. For example, in Australia where such interventions were widely enforced, there was a 7% reduction in road trauma deaths in 2020 compared with 2019 [15-17]. Our study aimed to investigate if those reductions were also seen in an LMIC like Vietnam.

Objective

We aimed to investigate the effect of the COVID-19 pandemic and the Vietnamese government's policies in response to it on road trauma fatalities. We also compared this impact with other improvements related to government policies over the past 14 years (2007-2020) regarding road trauma management.

Methods

Data Sources

COVID-19 data were extracted from the Vietnamese Ministry of Health database [18]. Road traffic mortality (2007-2021) was derived from the Vietnamese General Statistical Office [19]. All data sources were anonymous or deidentified for privacy and confidentiality protection.

Statistical Methodology

We used Stata software (version 17; StataCorp) for statistical analysis. Poisson regression modeling, with the logarithm of the national population of Vietnam for each year included as an offset, was used to estimate trends in the rates of road fatalities from the annual national mortality data for the 2007-2021 period. The Poisson regression model related the expected number $E(Y_t)$ of road fatalities Y_t in Vietnam in each year t (t=2007, 2008,…, 2021) to a vector of covariates $\mathbf{x}_t$ for that year according to the log-linear form:

where pop_t is the estimated national population of Vietnam for year t and In(pop_t) is its logarithm entered as an offset (a covariate with a coefficient of unity), and is the vector of coefficients to be estimated. The final form of the linear predictor $\mathbf{x}_t$ is as follows:

where t_{2008}, t_{2012}, t_{2020}, and t_{2021} are step functions that allow the vertical position of the quadratic line of best fit to shift up or down (a level change) in the years nominated (2008, 2012, 2020, and 2021) and subsequent years. For example, the step function t_{2008} was defined as:

The exponential value of the estimated coefficient is an estimate of the percentage change (step up or step down) that occurred in the year 2008. The other step functions were defined in analogous ways. The 4 years involved were chosen a priori as the years that helmet and seatbelt use were made mandatory (2008), traffic safety regulations were strengthened (2012), impaired driving laws were fortified (2020), and the COVID-19 response was implemented (2021). Having annual data only, there was neither capacity nor need to estimate the lag in months between implementation and impact, nor was it necessary to use segmented regression methods. The final form of the linear predictor was selected after fitting interaction terms between each of the step functions and the covariates (t and t^2) for year to allow the slope of the trend line to change. In every case, the null hypothesis that no change in slope had occurred was accepted (P>.05). There was no evidence of extra-Poisson variation (deviance goodness-of-fit test: P=.13) or of autocorrelation in the residuals (Q-statistic: P>.05 at each lag). The actual change in road traffic mortality in 2021 was compared with calculated figures to demonstrate the effect of COVID-19 measures on road trauma fatality.

Ethical Considerations

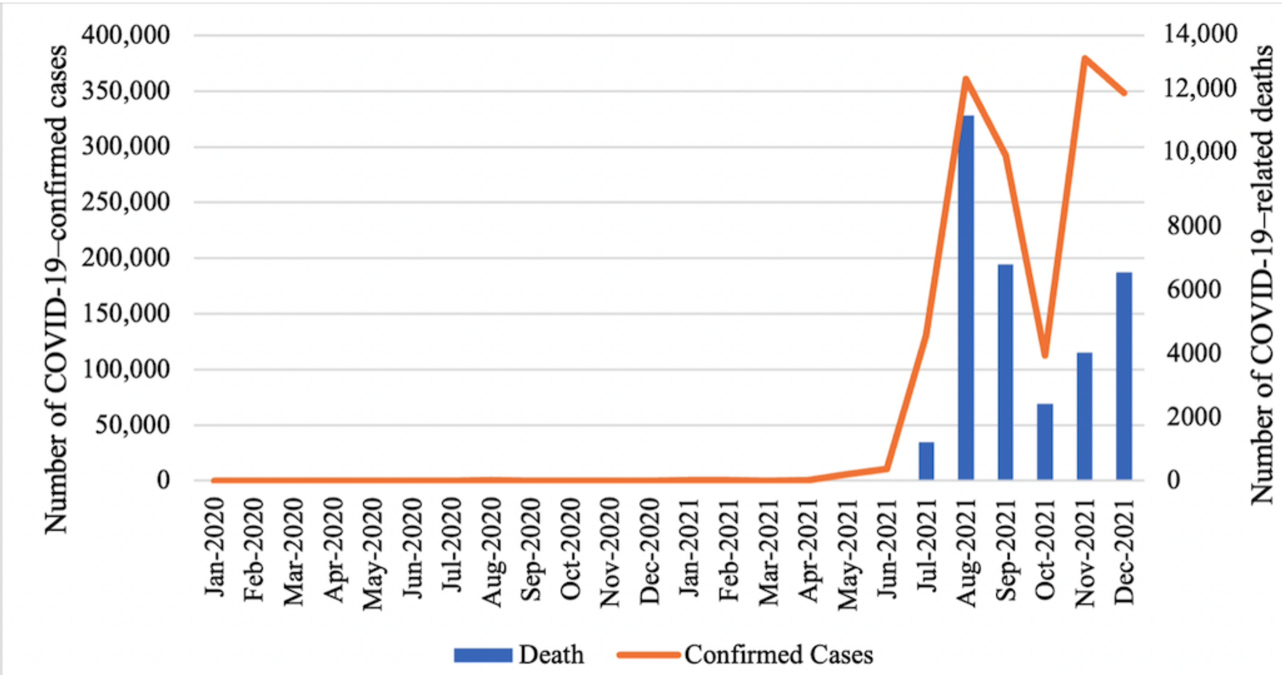
The study was approved for human ethics exemption (H0027318 [H-84839]), issued on April 6, 2022, by the Human Research Ethics Committee of the University of Tasmania.

Results

The COVID-19 Pandemic in Vietnam

There were fewer than 3000 confirmed COVID-19 cases and 35 COVID-19–related deaths during the initial 15 months since the first recorded case in January 2020. However, with the appearance of the Delta variant, the number of confirmed cases escalated exponentially from 373 in April 2021 to 10,730 in June 2021, reaching a peak of over 360,000 in September and a second peak in November (Figure 1). Similarly, the number of deaths due to COVID-19 rapidly escalated from 18 in May 2021 to 11,487 in August 2021. A high death rate persisted until the end of 2021 [6]. In response, the Vietnamese government imposed a lockdown (from July 23 to October 31) on approximately one-third of the country's population (about 32 million people), including in its most populous cities, Hanoi and Ho Chi Minh City.

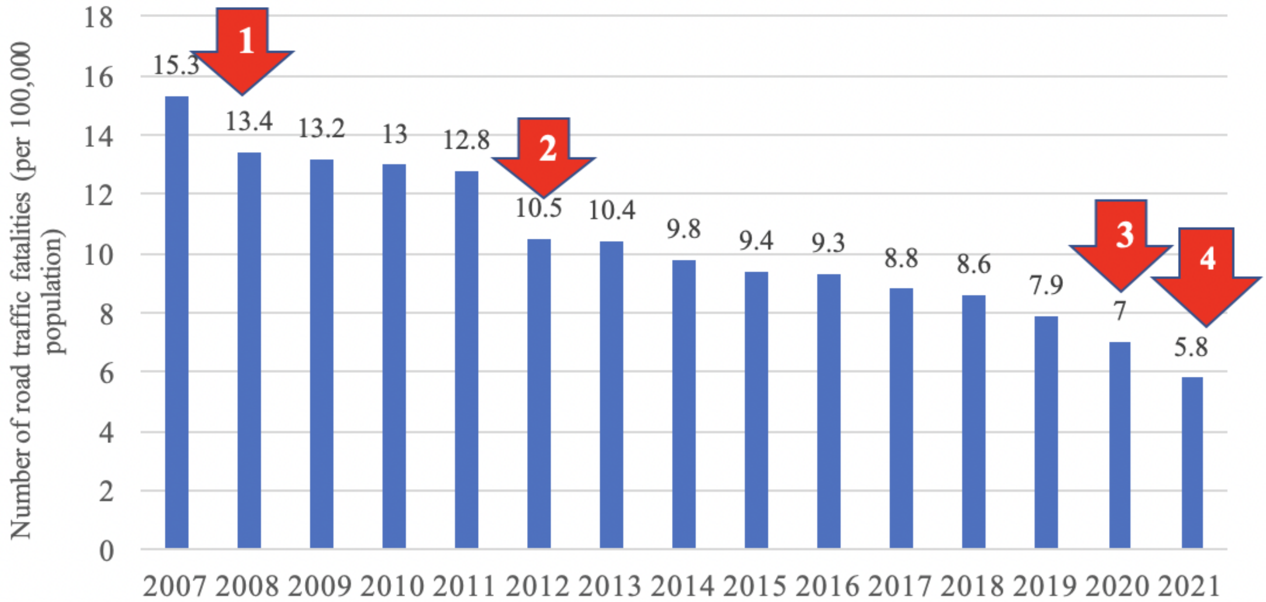
Figure 1. Deaths and confirmed cases of COVID-19 in Vietnam from 2020 to 2021 [18].



Road Trauma Fatalities in Vietnam

Between 2007 and 2020, the rate of road traffic fatalities decreased by more than 50% from 15.3 to 7 per 100,000 population, an average reduction of 5.4% per annum. The greatest improvements were seen in 2008, 2012, and 2020 (13.0%, 18.0%, and 11.4%, respectively; Figure 2). We estimated that road traffic mortality rates declined by 12.1% (95% CI 8.9-15.3%) in 2021 relative to this trend. In 2021, there were 5739 deaths, which was a 16.4% decrease from the previous year [20]. Thus, there was a further reduction of 4.3% in road trauma deaths in 2021 compared to the previous trend.

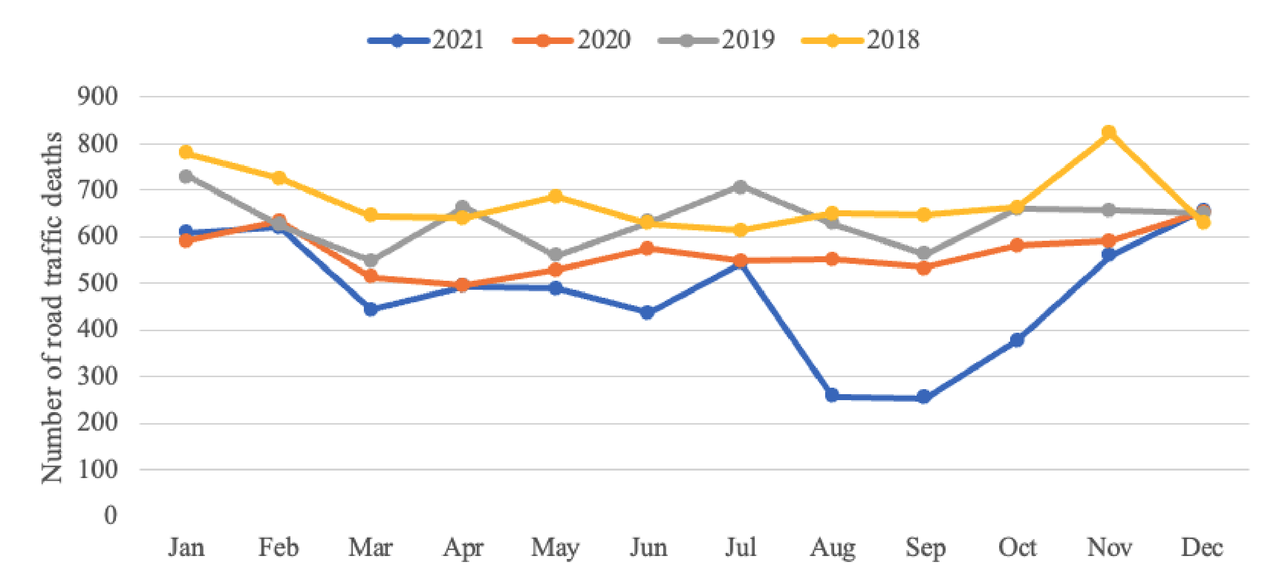
Figure 2. Number of road traffic fatalities per 100,000 population in Vietnam from 2007 to 2021 [12,13,20]. Arrows indicate government policies that were intentionally or otherwise implemented to address this issue: (1) mandatory helmet use, (2) strengthening of the national traffic safety committee (ie, increased presence of traffic police and increased financial penalties for traffic violations), (3) implementation of the zero-alcohol policy, and (4) COVID-19–related measures.



COVID-19's Impact on Road Trauma

During the first 2 calendar months of the lockdown, the number of traffic accident deaths decreased by more than 50% from 543 in July to 257 and 254 in August and September 2021, respectively (Figure 3). It then increased to 377 in October before increasing more substantially to 580 and 680 in November and December, respectively [21].

Figure 3. Monthly statistics on road accident fatalities from 2018 to 2021 in Vietnam [20].



Discussion

Principal Findings

Since the declaration of COVID-19 as a global pandemic by the World Health Organization, the disease has swept through all countries. COVID-19 had minimal impact in Vietnam during the 15 months following its first confirmed case due to the early and decisive policies implemented by the Vietnamese government. However, these early successes were undone by the Delta variant after April 2021. The number of cases and deaths rapidly increased from then onward. Due to vaccine scarcity, the Vietnamese government had to initially rely on lockdown and social distancing policies to control the pandemic. With the availability of community vaccination and its high uptake, lockdown and social distancing measures were relaxed by November 2021. The number of COVID-19 cases reached a new peak, but the fatality rate was only a third of that in August 2021 [22]. According to data from the World Health Organization, countries worldwide have experienced these trends regardless of social economic standing [23].

Road Trauma: A Steady Improvement in the Past 14 Years

Road trauma is a leading cause of death in Vietnam [13]. Like other LMICs, common causes include poor road safety and law enforcement, impaired driving, and speeding. Vietnam has been ranked 50 out of 183 recorded countries for road traffic deaths per 100,000 population [24]. Consequently, reducing road traffic deaths has been a priority of the Vietnamese government. A range of government decrees, regulations, laws, and policies have been issued and updated to improve the population's health and financial circumstances. As a result of these initiatives, road trauma deaths have been decreasing by an average of 5.4% per annum over the 2007-2020 period, with the greatest reductions seen in 2008, 2012, and 2020 (Figure 1). The first reduction coincided with the implementation of mandatory motorcycle helmet use legislation [25]. The second was associated with the national traffic safety committee increasing traffic police presence on the road and enforcement of increased financial penalties for traffic violations [11]. The third was associated with the introduction of zero-alcohol laws for all drivers and passengers [26].

A "Positive" Side of the Pandemic

We found that road trauma deaths in Vietnam in 2021 declined at a greater rate compared to the long-term trend. This was likely in response to the COVID-19 pandemic lockdown. In 2021, we found a 16.4 % improvement in road trauma deaths compared with an average 5.4% reduction over the preceding 14 years. This reduction fluctuated month to month. A reason for this distortion was that during the 3 months of lockdown, the number of vehicles on the road decreased by 84%, largely contributing to the total year reduction of 37% [27]. Likewise, Yasin et al [15] compared April 2020 to April 2019 and found that 32 out of 36 countries had a sharp drop in traffic volume, leading to a substantial reduction in road crashes and road trauma deaths.

It could be argued that the mortality rate was already declining due to other factors, and the additional decline beyond the preceding trend in 2021 (4.3%) was a delayed effect of the "zero-alcohol" policy from 2020 and not attributable to COVID-19. The temporal relationship of road trauma deaths in relation to the lockdown (Figure 3) supports our attribution beyond the zero-alcohol policy.

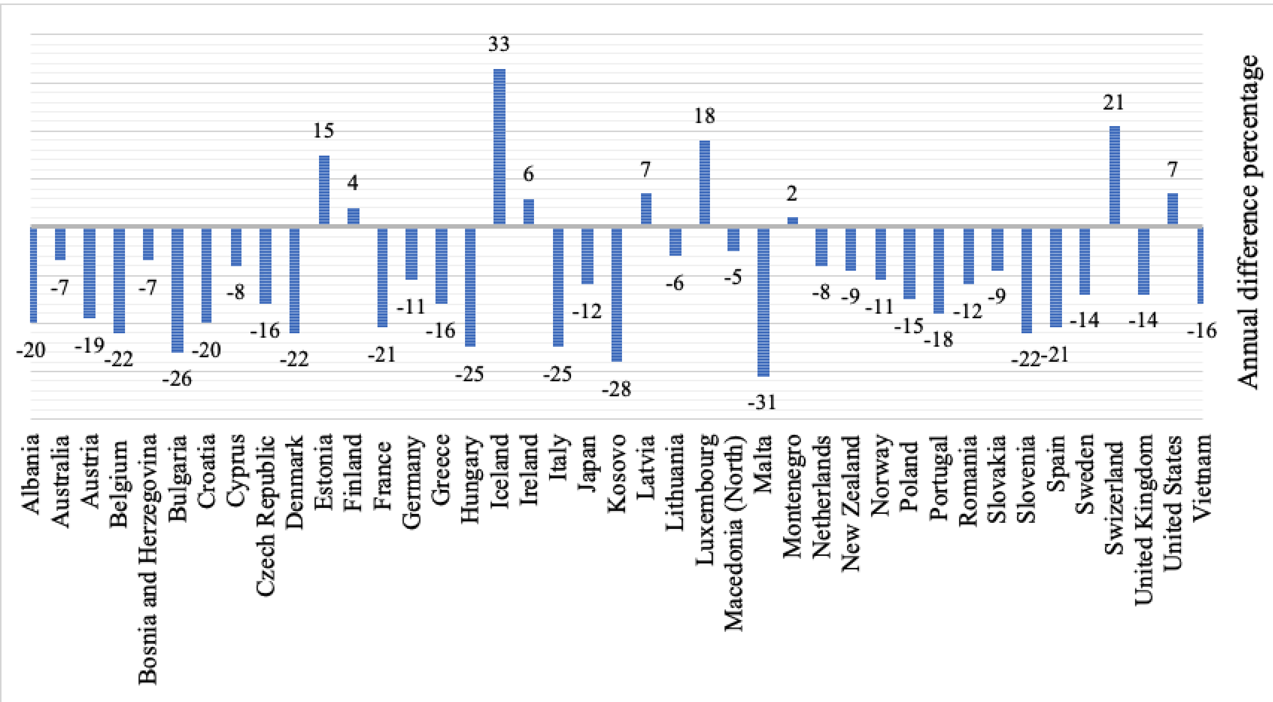
Effects of the COVID-19 Pandemic on Road Trauma Deaths Globally

A similar global trend was seen in most HICs where data have been published, with most countries showing a reduction in the annual road accident fatality rate (annual difference percentages) ranging from 5% in Macedonia to 31% in Malta. Some exceptions were seen in several countries with increases observed in Estonia (15%), Finland (4%), Iceland (33%), and the United States (7%) (Figure 4).

The explanation for this may be that these countries had not applied public health measures such as lockdowns with as much vigor or as extensively. There may also have been an increased use of private vehicles due to a fear of using public transport [36], leading to increased road accidents [26,27].

A lesson learned from the COVID-19 pandemic is that one approach to limiting road trauma deaths and injuries is to reduce vehicular traffic. Though this seems a trivial observation and interpretation, it does suggest that the Vietnamese government should prioritize the improvement of public transport such as bus and train systems to reduce private vehicle use and subsequently road trauma and deaths.

Figure 4. Comparison of the annual difference percentage in road trauma deaths prior to and during the COVID-19 pandemic in Vietnam and other countries [16,17,27-35].



Limitations

This study would have benefited from additional national and international comparisons. We were limited in performing this analysis due to dissimilarity between national and international data not only for LMICs but also for HICs [37]. In this paper, we did not use information from international sources such as the World Health Organization, the World Bank, or the United Nations since the data have been insufficient in the last 15 years and highly variable. For these reasons, we chose original data sourced from the Vietnamese General Statistical Office and Ministry of Health.

Conclusion

In 2021, the reduction in road traffic deaths in Vietnam was 3 times greater than the trend over the preceding 14 years. The public health response to the COVID-19 pandemic in Vietnam was associated with this reduction. This has also been observed in HICs but was demonstrated by us for the first time in an LMIC.

Acknowledgments

This work was funded by the Elphinstone Group and Broadreach Holdings Pty Ltd. BTN holds a University of Tasmania scholarship.

Data Availability

COVID-19 data can be found on the Ministry of Health website [18]. Data on road traffic mortality can be found on the General Statistical Office website [19]. All sources are available to the public and free to access.

Authors' Contributions

BTN and MN conceived the idea for the study. BTN, TCQ, and HTN collected the data. BTN and VT drafted the manuscript. BTN and CLB are responsible for the statistical analyses. AP, CLB, TCQ, and HTN revised the manuscript. MN contributed to the critical revision of the manuscript for important intellectual content and approved the final version. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Baloch S, Baloch MA, Zheng T, Pei X. The coronavirus disease 2019 (COVID-19) pandemic. Tohoku J Exp Med 2020 Apr;250(4):271-278 [FREE Full text] [doi: 10.1620/tjem.250.271] [Medline: 32321874]
2. Van Cuong L, Giang HTN, Linh LK, Shah J, Van Sy L, Hung TH, et al. The first Vietnamese case of COVID-19 acquired from China. Lancet Infect Dis 2020 Apr;20(4):408-409 [FREE Full text] [doi: 10.1016/S1473-3099(20)30111-0] [Medline: 32085849]
3. Van Tan L. COVID-19 control in Vietnam. Nat Immunol 2021 Mar;22(3):261. [doi: 10.1038/s41590-021-00882-9] [Medline: 33627879]
4. McKee KL, Crandell IC, Hanlon AL. County-level social distancing and policy impact in the United States: a dynamical systems model. JMIR Public Health Surveill 2020 Dec 23;6(4):e23902. [doi: 10.2196/23902]
5. Vu DC, Nguyen TH, Ho TL. May the Vietnam response have reduced daily new cases of COVID-19 in the country? Disaster Med Public Health Prep 2022 Feb 14:1-11 [FREE Full text] [doi: 10.1017/dmp.2022.33] [Medline: 35152938]
6. Minh LHN, Khoi Quan N, Le TN, Khanh PNQ, Huy NT. COVID-19 timeline of Vietnam: important milestones through four waves of the pandemic and lesson learned. Front Public Health 2021 Nov 24;9:709067 [FREE Full text] [doi: 10.3389/fpubh.2021.709067] [Medline: 34900885]
7. Guest JL, Del Rio C, Sanchez T. The three steps needed to end the COVID-19 pandemic: bold public health leadership, rapid innovations, and courageous political will. JMIR Public Health Surveill 2020 Apr 06;6(2):e19043 [FREE Full text] [doi: 10.2196/19043] [Medline: 32240972]
8. Nguyen T, Wong ZS, Wang L, Thanh TT, Van Nguyen H, Gilmour S. Rapid impact assessments of COVID-19 control measures against the Delta variant and short-term projections of new confirmed cases in Vietnam. J Glob Health 2021;11:03118 [FREE Full text] [doi: 10.7189/jogh.11.03118] [Medline: 34987776]
9. Rossiter ND. Trauma-the forgotten pandemic? Int Orthop 2022 Jan 14;46(1):3-11 [FREE Full text] [doi: 10.1007/s00264-021-05213-z] [Medline: 34519840]
10. Global status report on road safety. World Health Organization. 2018. URL: https://www.who.int/publications/i/item/9789241565684 [accessed 2023-01-31]
11. Viet Hung K. For an effective national road safety observatory. National Traffic Safety Committee. 2019 Dec 03. URL: https://tinyurl.com/3usn7cr7 [accessed 2023-01-31]

12. Road safety performance review, Viet Nam: project on strengthening the national road safety management capacities of selected developing countries, and countries with economies in transition. United Nations. 2018. URL: https://unece.org/DAM/trans/roadsafe/unda/RSPR_Viet_Nam_FULL_e.pdf [accessed 2023-01-31]
13. Le VD. Vietnam national road safety goals and action plan: opportunities and challenges. United Nations ESCAP. 2019. URL: https://tinyurl.com/5ufmty36 [accessed 2023-01-31]
14. Anh TTV, Burlacu AF, Small M, Paala M, Duc NH, Tri LH. Road safety data assessment in Vietnam for the establishment of a national road safety observatory. Vietnam Transport Knowledge Series. Hanoi, Vietnam: World Bank; 2021. URL: https://openknowledge.worldbank.org/handle/10986/35979 [accessed 2023-01-31]
15. Yasin YJ, Grivna M, Abu-Zidan FM. Global impact of COVID-19 pandemic on road traffic collisions. World J Emerg Surg 2021 Sep 28;16(1):51 [FREE Full text] [doi: 10.1186/s13017-021-00395-8] [Medline: 34583713]
16. Road safety report annual 2020. International Transport Forum. 2020 Oct 26. URL: https://www.itf?oecd.org/sites/default/files/docs/irtad?road?safety?annual?report?2020_0.pdf [accessed 2023-01-31]
17. Road deaths Australia. Australian Government; Department of Infrastructure, Transport, Regional Development, and Communications. 2020 Apr. URL: https://www.bitre.gov.au/sites/default/files/documents/rda_apr_2020.pdf [accessed 2023-01-31]
18. Portal of the Ministry of Health about the COVID-19 pandemic. Ministry of Health, Vietnam. URL: https://covid19.gov.vn/ban-tin-covid-19.htm [accessed 2023-01-31]
19. Search results for "social economic situation report by years". General Statistics Office, Vietnam. URL: https://www.gso.gov.vn/en/?s=SOCIO-ECONOMIC+SITUATION+REPORT&lang=en [accessed 2023-01-31]
20. Nguyen TTH. Press release socio-economic situation in the fourth quarter and 2021. General Statistics Office, Vietnam. 2021 Dec 29. URL: https://www.gso.gov.vn/en/data-and-statistics/2022/01/press-release-socio-economic-situation-in-the-fourth-quarter-and-2021/ [accessed 2023-01-31]
21. Vietnam locks down capital Hanoi as COVID-19 infections soar. Al Jazeera. 2020 Jul 24. URL: https://www.aljazeera.com/news/2021/7/24/vietnam-locks-down-capital-hanoi-as-covid-19-infections-soar [accessed 2023-01-31]
22. Mathieu E, Ritchie H, Rodés-Guirao L, Appel C, Giattino C, Hasell J, et al. Coronavirus (COVID-19) vaccinations. Our World in Data. 2020. URL: https://ourworldindata.org/covid-vaccinations?country=OWID_WRL [accessed 2023-01-31]
23. WHO coronavirus (COVID-19) dashboard. World Health Organization. 2023. URL: https://covid19.who.int/ [accessed 2023-01-31]
24. Road traffic accidents. World Life Expectancy. URL: https://www.worldlifeexpectancy.com/ [accessed 2023-01-31]
25. Olson Z, Staples JA, Mock C, Nguyen NP, Bachani AM, Nugent R, et al. Helmet regulation in Vietnam: impact on health, equity and medical impoverishment. Inj Prev 2016 Aug;22(4):233-238 [FREE Full text] [doi: 10.1136/injuryprev-2015-041650] [Medline: 26728008]
26. Park K. New law on prevention and control of alcohol related harms in Vietnam. J Glob Health Sci 2019;1(2). [doi: 10.35500/jghs.2019.1.e49]
27. Pham HC. Impacts of COVID-19 on urban mobility: experience of major cities in Vietnam. Transport Development and Strategy Institute, United Nations ESCAP. 2020 Nov. URL: https://www.unescap.org/sites/default/files/5_COVID-19%20and%20Mobility%20Experience%20of%20Major%20Cities%20in%20Viet%20Nam.pdf [accessed 2023-01-31]
28. Mobility and transport. European Commission. URL: https://ec.europa.eu/transport/modes/road/news/2021?0420?road?safety_en [accessed 2023-01-31]
29. Early estimate of motor vehicle traffic fatalities in 2020. US Department of Transportation, National Highway Traffic Safety Administration. 2021 Jun. URL: https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/813118 [accessed 2023-01-31]
30. Fatalities for 2020: annual statistics for Western Balkans. Transport Community Permanent Secretariat. 2021 Apr 15. URL: https://www.mit-ks.net/repository/docs/2021_04_22_081823_Annual-Statistics-2020.pdf [accessed 2023-01-31]
31. Motor vehicle. National Safety Council. 2021. URL: https://injuryfacts.nsc.org/motor?vehicle/overview/preliminary?estimates/ [accessed 2023-01-31]
32. Katrakazas C, Michelaraki E, Sekadakis M, Yannis G. A descriptive analysis of the effect of the COVID-19 pandemic on driving behavior and road safety. Transp Res Interdiscip Perspect 2020 Sep;7:100186 [FREE Full text] [doi: 10.1016/j.trip.2020.100186] [Medline: 34173462]
33. 2020 road deaths down on 2019. New Zealand Police. 2021 Jan 01. URL: https://www.police.govt.nz/news/release/2020-road-deaths-down-2019 [accessed 2021-05-25]
34. Jacob S, Mwagiru D, Thakur I, Moghadam A, Oh T, Hsu J. Impact of societal restrictions and lockdown on trauma admissions during the COVID-19 pandemic: a single-centre cross-sectional observational study. ANZ J Surg 2020 Nov;90(11):2227-2231. [doi: 10.1111/ans.16307] [Medline: 32894624]
35. Road traffic estimates in Great Britain: 2020. Gov.uk, Department for Transport. 2021 Apr 28. URL: https://www.gov.uk/government/statistics/road-traffic-estimates-in-great-britain-2020 [accessed 2021-06-05]

36. Pawar DS, Yadav AK, Akolekar N, Velaga NR. Impact of physical distancing due to novel coronavirus (SARS-CoV-2) on daily travel for work during transition to lockdown. Transp Res Interdiscip Perspect 2020 Sep;7:100203 [FREE Full text] [doi: 10.1016/j.trip.2020.100203] [Medline: 34173467]
37. Razzaghi A, Soori H, Abadi A, Khosravi A. World Health Organization's estimates of death related to road traffic crashes and their discrepancy with other countries' national report. J Inj Violence Res 2020 Oct 06;12(3) [FREE Full text] [doi: 10.5249/jivr.v12i3.1425] [Medline: 33021247]

Abbreviations

HIC: high-income country
LMIC: low- and middle-income country

Edited by A Mavragani; submitted 05.09.22; peer-reviewed by T Thang, N Phuong; comments to author 07.10.22; revised version received 19.12.22; accepted 31.01.23; published 06.02.23.

Please cite as:
Nguyen BT, Blizzard CL, Palmer A, Nguyen HT, Cong Quyet T, Tran V, Nelson M
COVID-19 in Vietnam and Its Impact on Road Trauma: Retrospective Study Based on National Data
Interact J Med Res 2023;12:e40883
URL: https://www.i-jmr.org/2023/1/e40883
doi:10.2196/40883
PMID:36718815

Original Paper

Creation of a Holistic Platform for Health Boosting Using a Blockchain-Based Approach: Development Study

Juan Lopez-Barreiro[1*], MSc; Luis Alvarez-Sabucedo[2*], PhD; Jose-Luis Garcia-Soidan[1*], PhD; Juan M Santos-Gago[2*], PhD

[1]Faculty of Education and Sport Sciences, University of Vigo, Pontevedra, Spain

[2]AtlanTTic, University of Vigo, Vigo, Spain

[*]all authors contributed equally

Corresponding Author:
Juan Lopez-Barreiro, MSc
Faculty of Education and Sport Sciences
University of Vigo
Campus A Xunqueira, s/n
Pontevedra, 36005
Spain
Phone: 34 610669712
Email: juan.lopez.barreiro@uvigo.es

Abstract

Background: Low adherence to healthy habits, which is associated with a higher risk of disease and death, among citizens of Organization for Economic Co-operation and Development countries is a serious concern. The World Health Organization (WHO) and the physical activity (PA) guidelines for Americans provide recommendations on PA and healthy diets. To promote these habits, we suggest using a blockchain-based platform, using the PA Messaging Framework to deliver messages and rewards to users. Blockchain is a decentralized secure platform for data management, which can be used for value-added controls and services such as smart contracts (SCs), oracles, and decentralized applications (dApps). Of note, there is a substantial penetration of blockchain technologies in the field of PA, but there is a need for more implementations of dApps to take advantage of features such as nonfungible tokens.

Objective: This study aimed to create a comprehensive platform for promoting healthy habits, using scientific evidence and blockchain technology. The platform will use gamification to encourage healthy PA and eating habits; in addition, it will monitor the activities through noninvasive means, evaluate them using open-source software, and follow up through blockchain messages.

Methods: A literature search was conducted on the use of blockchain technology in the field of PA and healthy eating. On the basis of the results of this search, it is possible to define an innovative platform for promoting and monitoring healthy habits through health-related challenges on a dApp. Contact with the user will be maintained through messages following a proposed model in the literature to improve adherence to the challenges.

Results: The proposed strategy is based on a dApp that relies on blockchain technology. The challenges include PA and healthy eating habits based on the recommendations of the WHO and the Food and Agriculture Organization. The system is constituted of a blockchain network where challenge-related achievements are stored and verified using SCs. The user interacts with the system through a dApp that runs on their local device, monitors the challenge, and self-authenticates by providing their public and private keys. The SC verifies challenge fulfillment and generates messages, and the information stored in the network can be used to encourage competition among participants. The ultimate goal is to create a habit of healthy activities through rewards and peer competition.

Conclusions: The use of blockchain technology has the potential to improve people's quality of life through the development of relevant services. In this work, strategies using gamification and blockchain are proposed for monitoring healthy activities, with a focus on transparency and reward allocation. The results are promising, but compliance with the General Data Protection Regulation is still a concern. Personal data are stored on personal devices, whereas challenge data are recorded on the blockchain.

(Interact J Med Res 2023;12:e44135) doi:10.2196/44135

KEYWORDS
blockchain; exercise; gamification; habits; healthy lifestyle; physical fitness

Introduction

Background

In modern societies, many of the deaths and diseases that occur could easily be avoided if people adopt healthy lifestyle habits [1-4]. Therefore, the governments of the Organization for Economic Co-operation and Development (OECD) countries are especially interested in promoting healthy lifestyle habits among their citizens and have been making relevant policies. The problem observed in these policies, however, is the low adherence to these habits among the general population. It seems, therefore, that the difficulty lies not in defining these habits but in generating a culture based on them.

On the one hand, recommendations for the practice of PA and its benefits for people's health, based on scientific evidence, can be found in the physical activity (PA) guidelines of the World Health Organization (WHO) [1] and the PA guidelines for Americans (PAG) [2]. These guidelines state that, in general, for all populations, some exercise is better than none. If people who do not practice any PA just start doing so, they will obtain health benefits. It is recommended that people with sedentary habits should perform PA following the principles of load progression [5]. People who perform moderate-intensity PA can gradually begin performing vigorous PA. In addition to the practice of PA, a healthy diet is recommended, which involves reducing sugar, fat, and salt consumption and limiting the consumption of processed foods and foods containing saturated fats.

On the other hand, the guidelines recommend the consumption of fruits, vegetables, legumes, nuts, and whole grains, as well as the consumption of at least 5 servings of fruits and vegetables daily [6-12]. The guidelines also emphasize that poor dietary habits together with a lack of PA greatly increase the risk of contracting noncommunicable diseases [6,11].

This is why it is of great social value to provide tools to promote these healthy habits in the population and monitor adherence. To this end, this paper explores the use of a platform to promote these habits and monitor adherence taking advantage of blockchain (BC) for gamification techniques. To carry out this gamification, the Physical Activity Messaging Framework [13] is used to organize the delivery of the most appropriate messages and rewards to the user to encourage their participation. These messages are categorized as generic, targeted, tailored, tailored personalized, and generic personalized, and as we progress through them, they become more relevant and more personal to the user.

Generic messages are those that apply to any person in general, regardless of their particularities, and that inform about the benefits of PA practice (eg, "Performing PA is good for your health") [14]. By contrast, targeted messages are relevant to a specific group [15], in this case, the general population of adults, and specifically highlight the benefits of PA practice in this group (eg, "Adults should perform 30 minutes of moderate PA per day to improve their cardiovascular health"). To engage the user in a more personal way, tailored messages are used. These messages use specific data about each individual user (eg, their goals) to make the message more relevant [15] (eg, "You are only 10 minutes away from reaching your weekly PA goal. Achieve it and improve your cardiovascular health!").

A personalization layer can be added to these messages, which consists of adding data that are not related to PA, such as the name of the user, to increase the salience and proximity of the message [15]. Thus, this feature can be added to generic messages (eg, "Hi Manuel! Doing PA is good for your health") and to tailored messages (eg, "Hi Rosa! You are only 10 minutes away from reaching your weekly PA practice goal. Achieve it and improve your cardiovascular health!").

In addition, messages can be classified according to whether they are framed to highlight the benefit of meeting the proposed objectives (gain framed; eg, "PA practice reduces the risk of heart disease, hypertension, and type 2 diabetes") or to point out the harms of not meeting them (loss framed; eg, "Not performing PA increases the risk of heart disease, hypertension, and type 2 diabetes") [16]. Messages aimed at highlighting the benefits of performing PA are generally recommended to promote PA practice [14,16]. By contrast, messages emphasizing the harms of not performing PA may also be recommended in certain cases, such as back injuries, where it may be beneficial to increase the perceived risk of not performing PA to engage users [17,18].

BC is a technology that provides features such as decentralization, transparency, open source, autonomy, immutability, and anonymity [19]; it can be conceptualized as a new model for the externalization of trust in information management in a distributed environment [20]. It consists of generating a general ledger, in which, using accounting terminology, the information is stored. This information, by the nature of the system itself, becomes immutable. To this end, it relies on a *peer-to-peer* structure in which the nodes or members participating in the system collaborate with each other to guarantee the inviolability of the data and their high availability, subject neither to the failure of a server nor to the management of a third party. This latter aspect is what allows it to become the appropriate tool when it is not desirable to rely on third parties. The nodes within the network themselves validate the records and add them to a chain of blocks (hence the name of the technology), which constitutes the aforementioned ledger of records [21].

When an agent wishes to enter a new record in this ledger, the agreement of all members of the ledger's host network is needed before the record can be validated. This is done by using a specific protocol called a *consensus algorithm*, which establishes the criteria for the acceptance of an element in the chain of records. The 2 most common consensus algorithms are *proof of work* (PoW) [22], used in the bitcoin network, whereby miners must solve a complex mathematical problem to justify the inclusion of the new block, and *proof of authority* (PoA)

[22], which allows the inclusion of new records based on the relevance of the miner making the proposal. These algorithms are only a small sample of the plethora of proposals in the literature.

This technology is achieving high market penetration as a solution for information storage and verification in a wide variety of domains, ranging from cryptocurrencies to the traceability of food and pharmaceutical products [23,24]. As shown in a recent systematic review [25], a significant penetration is observed in the field of PA but with a poor leveraging of the special features that BC offers to implement value-added controls and services such as smart contracts (SCs), oracles, and decentralized applications (dApps), each of which is described in Textbox 1.

Textbox 1. Descriptions of smart contracts (SCs), oracles, and decentralized applications (dApps).

- SCs: these are executable codes that run on top of the blockchain to facilitate, execute, and enforce an agreement among untrustworthy parties without the involvement of a trusted third party. SCs provide network automation and the ability to convert paper contracts into digital contracts [26].
- Oracles: SCs cannot take into consideration information that is not registered in the network itself. To alleviate this shortcoming and to improve the functionality of the network in general, oracles are introduced. Oracles are responsible for registering data captured from the real world in the network without the intervention of a human user [27].
- dApps: these are applications that run on one or more clients using information hosted in a distributed manner in a blockchain network, taking advantage of the characteristics of these networks. By making use of SCs, operations and verifications of conditions (such as those imposed in a challenge in a gamification context) are carried out without any human intervention [28].

As reported by Cai et al [28], the implementation of dApps is required to exploit another important feature existing within this environment, namely nonfungible tokens (NFTs). An NFT is an encrypted digital asset, a special type of cryptographic token that represents something unique. NFTs serve to prove that a certain user is in possession of a token that is unique, traceable, and exchangeable; they are very useful in certain gamification contexts to reward users for achieving their goals [29].

Objectives

On the basis of the points outlined so far, this work proposes the creation of a platform to facilitate the inculcation of healthy lifestyle habits and practices through a gamification strategy. The objective was to engage users—society as a whole—in activities that can become healthy habits. These healthy habits will be both sporting and nutritional.

One of the highlights of this platform is its great potential in terms of gamification. This tool provides a very functional support for the monitoring of the data hosted on it without human intervention (eg, the validation of the challenges presented to the users).

In addition, within this platform, the possibility of defining challenges in a highly parameterized way is contemplated so that different state or private agencies can, in due course, propose their own challenges and make them available to users.

We can say, therefore, that the objective of this work is to present a holistic platform for the support of health-related challenges among the population, using scientific evidence with the support of BC technology. It is intended to provide a mechanism to encourage and set healthy PA and eating habits among the general population by using gamification techniques through challenges.

This platform will integrate the noninvasive monitoring of the proposed activities, evaluation through open-source software, and follow-up using BC through messages addressed to the end user that will allow them to adhere to this activity.

Methods

As a first step, the state of the art in this regard was checked, through a literature search, to get an idea about the use of BC technology in the field of PA and healthy eating [25]. On the basis of the results of this search, it is possible to define an innovative platform for the promotion and monitoring of healthy lifestyle habits based on scientific evidence. The goal is to introduce healthy habits concretized in activities defined within different health-related challenges through the use of a dApp for the general population.

One of the key aspects with regard to improving users' adherence to the training program embedded in the challenges is to maintain contact with the user. To organize this information delivery to the user, messages will be used following the model proposed in the study by Williamson et al [13].

Results

The aforementioned review of the current literature shows a significant increase in scientific production related to BC technology in recent years, as is also indicated in 2 bibliometric reviews [24,30]. Among the large number of existing works that take advantage of this technology, the following are worth mentioning.

BC and PA and Health Care

In the literature, it is possible to find several works that combine BC technology with PA practice and health care. Among them is the study by Alsalamah et al [31], who proposed a platform to incentivize PA practice and encourage a healthy lifestyle through gamification and rewarding of users for meeting their goals using cryptocurrencies. Another noteworthy study is the one by Frikha et al [32], who stored users' health data in electronic health records to diagnose and treat patients more easily and cost-effectively. Other notable works are those by Jamil et al [33] and Jamil et al [34]; the authors assigned training and diet programs to each user based on their anthropometric and body composition data. Furthermore, in the study by Jamil et al [34], the authors allowed the transfer of the user profile among different sports centers.

BC and Sport

Other trending works have made contributions that are restricted to the sporting field, ranging from data capture and management to predictions of sporting performance. This is the case with the study by Cao et al [35], who developed a model to predict performance and improve training; the study by Hong and Park [36], who captured players' performance data to make tactical decisions in situ and in real time; and the study by Yu [37], who developed a model to improve and guide training using athletes' physiological data. Moreover, we have the study by Ma [38] in which the author filtered data from users' gait patterns; as well as the study by Shan and Mai [39], who proposed a system to capture and manage athletes' fitness data in real time. Finally, there is the study by Mulyati et al [40] in which a model was developed to store data regarding belt promotions and grades in taekwondo, bringing transparency and immutability to the scores.

BC and Active Aging

There are also very diverse contributions related to the incorporation of BC into active aging. Khezr et al [41] developed a system that provides alerts when normal behavioral patterns change. Rahman et al [42] assigned therapies based on users' treatment needs. Rahman et al [43] developed a system to control smart home devices using gestural recognition tools. Rupasinghe et al [44] determined the risk factors for falls and developed a model to predict them. Silva et al [45] captured physiological data of patients and made them secure and interoperable through BC. Spinsante et al [46] proposed an app to promote active aging and assess the level of PA practice and quality of life. Finally, Velmovitsky et al [47] proposed a system for users to control informed consent for their participation in studies at all times.

Table 1 shows a synthesis of the state of the art in different technological aspects, such as the use of SCs and oracles, support for cryptocurrencies and NFTs, training and dietary programs based on scientific evidence, and end-user delivery support.

Table 1. Analysis of studies related to blockchain and physical activity and health care, sport, and active aging.

Domain and reference	SC[a]	Oracle	Cryptocurrencies	NFT[b]	Evidence based	End-user delivery support
Physical activity and health care						
Alsalamah et al [31]	Yes	No	Yes	No	No	Web dApp[c] and mobile app
Frikha et al [32]	Yes	No	No	No	No	Web application and mobile app
Jamil et al [33]	Yes	No	No	No	No	Web application
Jamil et al [34]	Yes	No	No	No	No	Web application
Sport						
Cao et al [35]	No	No	No	No	No	Not described
Hong and Park [36]	No	No	No	No	No	Not described
Ma [38]	No	No	No	No	No	Not described
Mulyati et al [40]	No	No	No	No	No	Web application and mobile dApp
Shan and Mai [39]	No	No	No	No	No	Not described
Yu [37]	No	No	No	No	No	Not described
Active aging						
Khezr et al [41]	Yes	No	No	No	No	Not described
Rahman et al [42]	Yes	No	No	No	No	Not described
Rahman et al [43]	Yes	No	No	No	No	Web application and mobile dApp
Rupasinghe et al [44]	Yes	No	No	No	No	Not described
Silva et al [45]	No	No	No	No	No	Web application and mobile app
Spinsante et al [46]	No	No	No	No	No	Web application and mobile app
Velmovitsky et al [47]	Yes	No	No	No	No	Not described

[a]SC: smart contract.

[b]NFT: nonfungible token.

[c]dApp: decentralized application.

The studies cited (Table 1) dealt with the introduction of BC in the field of PA and health care, sport, and active aging. However, most of them (14/17, 82%) show very limited development, which shows us the initial stage of development of this technology in the field concerned. Only 9 (53%) of the 17 studies make use of SCs [31-34,41-44,47]. Among those describing the access policy, most (10/17, 59%) use private and authorized networks; of the 17 studies, only 1 (6%) uses a public network, and 1 (6%) uses an authorized consortium. In addition, only the study by Alsalamah et al [31] incentivizes using cryptocurrencies as a reward. None of the cited works make use of NFTs, and none base their training or dietary plans on scientific evidence. Regarding the delivery medium, most used web applications or mobile apps, and only 3 (18%) of the 17 studies leveraged dApps [31,40,43].

On the basis of this review of the state of the art and relying on the Physical Activity Messaging Framework and BC technology, challenges will be proposed to the general population and monitored through the use of a dApp that relies on the information and SCs stored in the BC.

These challenges are composed of (1) a series of PAs and specific healthy eating habits that generate benefits for the user when performed with the proposed sequencing and periodicity and (2) the messages corresponding to each challenge.

The PAs included in these challenges are based on scientific evidence following the recommendations for the general adult population found in the PA guidelines of the WHO [1] and the PAG [2], whereas the proposed healthy eating habits are based on the recommendations of the WHO and the Food and Agriculture Organization (FAO) [7,11]. We list here in a concrete and clear way the PA practice recommendations for the general adult population that will be the basis for the subsequent creation of the different challenges that users will have to complete to obtain their rewards (it is recommended to exceed the upper limits of moderate and vigorous PA or perform a combination of both):

- Moderate PA per week: 150 to 300 minutes
- Vigorous PA per week: 75 to 50 minutes
- Strength PA per week: ≥2 sessions

Among the aforementioned recommendations, we find different PA modalities such as aerobic exercise (muscle movement in a rhythmic way and maintained over time), muscle strengthening (strength training and weight lifting), bone strengthening (produces a force in the bones that promotes their growth and strength), balance training (improves the ability to resist internal or external forces of the body that cause falls), and multicomponent training (a combination of aerobic PA, balance training, and muscle strengthening), which will bring some benefit to the user when performed [2]. Of note, Momma et al [48], in their recent systematic review and meta-analysis of cohort studies on muscle-strengthening activities, highlighted the reduction in the risk of all-cause mortality, cardiovascular disease, cancer, and diabetes in participants by 10% to 17% [48]. Regarding healthy eating habits, the WHO recommends restricting sugar consumption to <10% of total daily calories, fat consumption to <30% of total daily calories, and salt consumption to <5 g daily, as well as limiting the consumption of processed foods and foods containing saturated fats to <10% of total calorie intake and foods containing trans fats to <1% of total calorie intake. By contrast, the guidelines recommend the consumption of fruits, vegetables, legumes, nuts, and whole grains, as well as the consumption of at least 5 servings of fruits and vegetables daily [6-12]. Aune et al [6] report a 31% decrease in the risk of contracting diseases with a daily intake of 800 g of fruits and vegetables, a 19% decrease with a daily intake of 600 g of fruits, and a 25% decrease with a daily intake of 600 g of vegetables; Leenders et al [49] suggest an increase in longevity with fruit and vegetable consumption; and Chowdhury et al [50] report that individuals consuming a well-balanced diet are healthier with a strong immune system and have a reduced risk of contracting infectious diseases such as COVID-19.

On the basis of the aforementioned recommendations for healthy habits and the scientific evidence that supports each activity, 4 challenges are generated (summarized schematically in Textbox 2).

Textbox 2. Explanatory summary of the 4 proposed challenges.

Challenge 1

- Name: High-intensity interval training (HIIT) *7-minute workout*
- Description: the challenge consists of performing 4 three-round sessions of the *HIIT 7-minute workout* in a week
- Included activities: *HIIT 7-minute workout*

Challenge 2

- Name: *Walk more than 10,000 steps every day*
- Description: the challenge consists of walking >10,000 steps every day of the week
- Included activities: walking >10,000 steps

Challenge 3

- Name: *Balance training*
- Description: the challenge consists of performing at least 2 days of eccentric training using gliding disks in a week
- Included activities: eccentric training protocol using gliding disks

Challenge 4

- Name: *Eat at least 5 servings of fruits and vegetables per day*
- Description: the challenge consists of eating at least 5 servings of fruits and vegetables every day of the week
- Included activities: eating at least 5 servings of fruits and vegetables

Common information for all 4 challenges

- Duration: 7 days
- Types of messages
  - Generic *gain framed*
  - Targeted *loss framed*
  - Tailored *gain framed*
  - Tailored and personalized *gain framed*

Challenge 1: *High-Intensity Interval Training 7-Minute Workout*

This challenge consists of the user performing the *high-intensity interval training (HIIT) 7-minute workout* 4 days per week. The basis for this challenge comes from the WHO and PAG recommendation to combine aerobic PA and muscle-strengthening PA and from the training proposed in the study by Klika and Jordan [51], in which PA training is performed only with body weight aerobic PA and muscle strengthening [2,52]. The training consists of repeating 2 or 3 sets of the *HIIT 7-minute workout* [51]. On the basis of the WHO and PAG vigorous PA practice recommendations, in this challenge, the user will be asked to perform 3 sets daily of the *HIIT 7-minute workout* at least 4 times per week.

If no workout has been performed after 2 days from the start of the challenge, the user will receive “PA practice improves your physical and mental health” as a generic message to highlight the benefit of meeting their goals.

After 3 days from the start of the challenge without performing any training, the user will receive “Not performing your strength training sessions will worsen your health” as a targeted message framed to highlight the harms of not meeting the PA and strength training goals.

When the user has completed 2 training sessions, they will receive “Cheer up! You have been strength training this week, keep it up to improve your quality of life” as a tailored message framed around the benefit.

Finally, when the user reaches their goal of 4 strength workouts per week, they will receive “Great job, [name of user]! You’ve completed this challenge, keep it up—you’re decreasing your chance of getting heart disease by more than 10%!” as a personalized tailored message based on the virtues of performing PA.

Challenge 2: *Walk More Than 10,000 Steps Every Day*

This challenge consists of the user walking >10,000 steps daily all 7 days of the week, based on the results of the recent systematic review and meta-analysis conducted by Jayedi et al [52], in which a clear decrease in the risk of all-cause mortality is observed when walking >10,000 steps daily, in addition to a 12% decrease in the risk of all-cause mortality with each increment of 1000 steps per day. The user will be asked to replace sedentary time with PA practice and walk >10,000 steps every day for 7 consecutive days.

After 1 day from the start of the challenge, if the user has not walked 10,000 steps, they will receive "The practice of PA reduces the risk of heart disease, hypertension, and type 2 diabetes" as a generic message to highlight the benefit of meeting their goals.

After 3 days from the start of the challenge without walking 10,000 steps, the user will receive "Not reaching your daily step goals will worsen your quality of life" as a targeted message framed to highlight the harms of not meeting daily step goals.

When the user has walked >10,000 steps for 4 consecutive days, they will receive "Cheer up! You have reached your daily goal again, keep it up to improve your cardiovascular health" as a tailored message framed around the benefit.

Finally, when the user manages to reach their daily step goal for 7 consecutive days, they will receive "Congratulations, [name of user]! You have completed this challenge, keep it up, you have just improved your physical and mental health!" as a personalized tailored message based on the virtues of performing PA.

Challenge 3: *Balance Training*

This challenge consists of the user performing strength training ≥2 days per week. To meet this goal, the user will be asked to perform the eccentric training exercises using sliding disks [53] at least twice a week.

After 2 days from the start of the challenge without performing any training, the user will receive "PA practice increases your longevity" as a generic message to highlight the benefit of meeting their goals.

After 4 days from the start of the challenge without performing any training, the user will receive "By not performing your balance training you are increasing the probability of falling" as a targeted message framed to highlight the harms of not meeting their weekly training goal.

When the user has completed 1 workout, they will receive "Cheer up! You've had a workout this week, keep it up to improve your balance" as a tailored message framed around the benefit.

Finally, when the user reaches their goal of 4 strength workouts per week, they will receive "Great job, [name of user]! You've completed this challenge, keep it up, you've just improved your balance and bone health!" as a personalized tailored message based on the virtues of performing balance training.

Challenge 4: *Eat at Least 5 Servings of Fruits and Vegetables per Day*

This challenge consists of the user eating at least 5 servings of fruits and vegetables per day, based on the results of the recent systematic review and meta-analysis conducted by Aune et al [6] as well as the recommendations of the WHO [11] and the FAO [7]. Both organizations recommend the consumption of at least 5 servings of fruits and vegetables per day, and Aune et al [6] report a 31% decrease in the risk of contracting diseases with a daily intake of 800 g of fruits and vegetables, a 19% decrease with a daily intake of 600 g of fruits, and a 25% decrease with a daily intake of 600 g of vegetables. To perform this challenge, the user will be asked to consume at least 5 servings of fruits and vegetables per day (1 serving is approximately 150 g) on all 7 days of the week.

After 1 day from the start of the challenge, if the user has not consumed at least 5 servings of fruits and vegetables, they will receive "WHO recommends the consumption of fruits and vegetables to reduce the risk of heart disease, hypertension, and type 2 diabetes" as a generic message to highlight the benefit of meeting their goals.

After 3 days from the start of the challenge without consuming the 5 daily portions, the user will receive "If you don't eat at least five servings of fruits and vegetables a day, you increase your risk of disease" as a targeted message framed to highlight the harms of not meeting the daily goals.

When the user has reached the goal of eating at least 5 servings of fruits and vegetables a day for 4 consecutive days, they will receive "Cheers! You have reached your daily goal again, keep it up to increase your life expectancy" as a tailored gain-framed message.

Finally, when the user manages to reach their daily goal of eating at least 5 servings of fruits and vegetables for 7 consecutive days, they will receive "Congratulations, [name of user]! You have completed this challenge, keep it up, you are reducing the probability of being diagnosed with cancer by more than 10%!" as a personalized tailored message based on the virtues of consuming fruits and vegetables.

Architectural Perspective

From an architectural perspective, the system is fundamentally constituted through a BC network. In this network, challenge-related achievements are stored, and their verification is performed using SCs. As mentioned in the previous sections, the objective of the system is to provide a motivating user experience so that participants feel engaged in the proposed activity and thus adhere to the challenges introduced in the system. By using this registration and verification tool, users can be assured of the veracity of their achievements.

To operate within the system, the user must make use of the dApp provided for this purpose. This application will run on the user's local device and be responsible for managing the user's identity and sending for publication on the BC network the data registered for the event in which the user is taking part. This monitoring of activities related to the challenge itself should be carried out in the least invasive way possible.

The BC network used for this purpose was hosted on an external service provider that runs the Hyperledger nodes with support for Web3 applications. In particular, tests were performed using support from Kaleido [54].

According to the proposed model, the SC defined for each challenge automates challenge-specific decision-making, performing tasks such as the following:

- Verifying the fulfillment of the conditions for each challenge: once the conditions for the challenge in question have been met, the established rewards are assigned.
- Generating the established messages: these messages correspond to certain challenge conditions that are analyzed by the SC. Thus, when a user does not perform the walking PA on a particular day, a corresponding alert is generated and sent to the user.

For the interaction with the BC network, the deployed nodes offer a representational state transfer application programming interface that allows the invocation of remote services in a simple way. A description of the most relevant procedures can be found in Table 2. The user self-authenticates when sending data by providing their public key and creating an encrypted field with their private key to validate the information sent. In other words, the access credentials will be managed only by the client device. It is also worth noting that any user or node can obtain a complete list of records in the chain and obtain the messages that have not yet been delivered.

Table 2. Description of the most relevant procedures.

Action	Resource	Purpose	Input parameters
POST	/v1/records	Add a challenge record	• Challenge ID • User log-in • Public key • Challenge facts • Encrypted hash with private key
GET	/v1/records	Obtain the complete data string or the data referring to a log-in or challenge	• User log-in (optional) • Challenge (optional) • Initial date (optional)
GET	/v1/awards	Obtain the rewards associated with a log-in	• User log-in
GET	/v1/messages	Obtain a user's pending messages	• User log-in (optional) • Initial date (optional)

As an example, in the case of challenge 1, *HIIT 7-minute workout*, the user must perform 3 sets of the 7-minute HIIT workout per day for at least 4 days per week. The user must manually record the sets performed each day using the dApp provided for this purpose. In addition, the user must attach a JPEG file demonstrating the completion of the training (eg, a screenshot of the heart rate variation intervals during the HIIT execution). Subsequently, the SC corresponding to this challenge, illustrated in Algorithm 1 within Textbox 3, performs the verification of the established conditions for this particular challenge. This mechanism supports the handling of messages sent to users as well as the allocation of a reward in the form of a transfer of the network's own cryptocurrency as a reward.

The information, which is stored in the network, can be used as a support to encourage competition among participants. To this end, using these data, dashboards can be generated showing the most involved user in the activity—the one who has walked the most steps, performed the most sets of the *HIIT 7-minute workout*, or consumed the most number of servings of fruits and vegetables—or any other parameter that may be interesting and can be used to encourage participation. The idea is to achieve a critical mass of users among whom a habit of healthy activities is inculcated through this system of rewards and competition among peers.

Of note, the tests carried out in the laboratory after the implementation of the BC network showed satisfactory results in its functioning.

Textbox 3. Algorithm 1: smart contract snippet for checking the high-intensity interval training challenge.

```
function checkNewAward(recordChallengeHIIT memory actualRecord) public returns (bool success) {      bool result = false;
     uint currentChain=0;
     uint userSearched = actualRecord.user;
     uint lastHiit = block.timestamp; // today      for (uint i = 0 ; i < recordCollection.length - 1; i++) {             if (recordCollection[i].user == userSearched)
{                 if ( ((recordCollection[i].date) >= (lastHiit + (4 days ))) && ( (recordCollection[i].date) < (lastHiit + 3 days )) )
               {
               currentChain = currentChain + 8;
               }
               if ( ((recordCollection[i].date) >= (lastHiit + (3 days ))) && ( (recordCollection[i].date) < (lastHiit + 2 days )) )
               {
                 currentChain = currentChain + 4;
               }
               if ((recordCollection[i].date >= (lastHiit + 2 days )) && (recordCollection[i].date < (lastHiit + 1 days )))
               {
                 currentChain = currentChain + 2;
               }
               if ((recordCollection[i].date >= (lastHiit + 1 days )) && (recordCollection[i].date < lastHiit ))
               {
                 currentChain = currentChain + 1;
               }
             }
         if (currentChain==15)
         {
           success = true;
           awardChallengeHIIT memory data;
           data.date = block.timestamp;
           data.user = actualRecord.user;
           data.award = "Award";
           addAward(data);
         }
     }
  return result;
  }
```

Discussion

With regard to the objectives and hypotheses set out in this work, we have been able to create a tool that encourages healthy lifestyle habits in the population through challenges. BC technology will be key to the implementation of these habits and the monitoring of compliance in the least intrusive way and without the need to rely on trusted third parties.

Limitations and Future Work

The limitations of this work include the limited consideration of General Data Protection Regulation (GDPR) implications and the manual need for information upload. In future work and to overcome the latter limitation, we propose the introduction of artificial intelligence techniques and the use of wearables connected to the dApp, a method similar to that used in the study by Santos-Gago et al [55].

Comparison With Prior Work

Regarding the characteristics considered relevant in the 17 articles cited in the *Results* section, the following aspects are worth highlighting in comparison with our proposal.

Regarding the access policy, our platform is formed by a permissioned network. Therefore, only authorized nodes will be able to participate in the platform, as is the case with 3 (18%)

of the 17 studies [33,34,36]. Other approaches [31,32,42,43,47] involved the use of a permissioned private BC, Rupasinghe et al [44] used a permissioned consortium BC, and Mulyati et al [40] used a public BC network. The other studies (7/17, 41%) do not indicate the type of network used.

Concerning SC, 8 (47%) of the 17 studies [35-40,45,46] did not report on their implementation, in contrast to our approach, which is similar to that of 9 (53%) of the 17 studies [31-34,41-44,47], which also made use of this special BC feature.

Regarding the use of cryptocurrencies to incentivize users, only Alsalamah et al [31] take advantage of this feature; the other works cited (16/17, 94%) do not indicate the use of cryptocurrencies. In our work, too, this feature is not exploited.

Concerning end-user delivery support, we implemented a dApp to be able to offer all the features that only BC can provide, similar to the approach used in 3 (18%) of the 17 studies [31,40,43].

Finally, regarding the use of oracles, NFTs, and the proposed PA and scientific evidence–based feeding, none of the aforementioned works indicate the use of these features. On the one hand, our proposal does not involve the use of oracles either. Nevertheless, according to the suggested model, it is possible to include oracles as actors with minor updates on the low level of the designed system. On the other hand, we use rewards to achieve a gamification experience and engage users in healthy lifestyle habits through challenges. We have based these challenges, composed of PA and healthy eating habits, on scientific evidence, supported by relevant organizations such as the WHO and the FAO.

Conclusions

The emergence of disruptive technologies such as BC has opened the door to new possibilities in the provision of services to society. This work explores the potential of this technology in the development of services that improve people's quality of life. To this end, strategies have been developed that allow, using gamification, the monitoring of adherence to new healthy habits in a simple way and, consequently, help to increase adherence.

The use of BC technology has been fundamental for meeting these objectives. In the review of previous works, it can be observed that the potential of BC has not been fully exploited. In our model, the aim is to show how to fully use this technology. Consequently, it is worth highlighting the following aspects of the platform:

- In an autonomous manner, without the need for supervision by a human agent and without the possibility of blockage, the verification of challenge completion is carried out. In the same process, the reward allocation is carried out, which cannot be interfered with by any system agent.
- All participants in the system can transparently verify the status of challenges at all times, thus increasing system transparency.
- As trust in the data resides in the network itself, there is no need to rely on a third party. This eliminates distrust because the promoter of the challenge is not known at first hand.

By contrast, when using this technology, there are certain deployment aspects that must be taken into account, including, primarily, the fact that future practitioners must be aware that once an SC is deployed, it cannot be modified, as could be the case with other technologies where the software is easily updatable. This is why it is very important to adequately test the system in development before deploying it in production.

This technology offers other elements that can improve users' adherence to the system but have not yet been properly implemented in this prototype. We are talking about both the use of NFTs to reward the fulfillment of certain challenges or meta-challenges and the use of oracles for the unsupervised acquisition of information to eliminate the need for user input and improve SC decision-making.

Although the system is still pending functional validation in a realistic environment, the experimental result has been satisfactory. A simple tool has been generated for the user, with a scalable and inexpensive deployment for service providers and with great potential for improving people's health. In this aspect, the adequate generation of training plans has played a fundamental role. These have been obtained from validated medical sources (eg, the WHO and the PAG) and therefore offer a high level of confidence.

A negative aspect of the system, pending more rigorous treatment, is compliance with the GDPR. This legal framework establishes a series of conditions, such as the elimination of user information when the user demands it. However, it should be noted that, in our proposal, no personal or medical data are recorded directly on the BC. The personal data are stored in the personal device, and the data regarding the completion of the different challenges are recorded on the BC. In fact, there are already critical voices regarding these aspects, and they are calling for a revision of the legal framework to facilitate the adoption of these new technologies [56].

In conclusion, a tool has been created through which healthy lifestyle habits can be inculcated in terms of both PA and healthy eating. Furthermore, it has been automated in the most transparent, safe, and least intrusive way possible using BC technology. Thereby, a tool to reduce the risk of all-cause mortality and to increase the well-being of society has been developed.

Acknowledgments

This research was cofunded by grant PID2020-115137RB-I00 funded by the Ministry of Science and Innovation (MCIN/AEI/10.13039/501100011033) and by the *Predoctoral grant program* of the Xunta de Galicia (Regional Ministry of Culture, Education, and University Organization).

Authors' Contributions

JL-B and LA-S were responsible for the conceptualization of the study as well as the software. All authors were responsible for the methodology, formal analysis, and data curation. The original draft was prepared by JL-B and LA-S. All authors reviewed and edited the draft and have read and approved the published version of the manuscript.

Conflicts of Interest

None declared.

References

1. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med 2020 Dec 25;54(24):1451-1462 [FREE Full text] [doi: 10.1136/bjsports-2020-102955] [Medline: 33239350]
2. Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA, et al. The physical activity guidelines for Americans. JAMA 2018 Nov 20;320(19):2020-2028 [FREE Full text] [doi: 10.1001/jama.2018.14854] [Medline: 30418471]
3. World Health Organization. WHO Guidelines on Physical Activity and Sedentary Behaviour: Web Annex: Evidence Profiles. Geneva: World Health Organization; 2020.
4. Yang Y, Dixon-Suen S, Dugué PA, Hodge A, Lynch B, English D. Physical activity and sedentary behaviour over adulthood in relation to all-cause and cause-specific mortality: a systematic review of analytic strategies and study findings. Int J Epidemiol 2022 May 09;51(2):641-667. [doi: 10.1093/ije/dyab181] [Medline: 34480556]
5. Barrera J. Actividad física como estilo de vida saludable: criterios básicos. Revista Médica de Risaralda 2003 Nov;9(2):43.
6. Aune D, Giovannucci E, Boffetta P, Fadnes LT, Keum N, Norat T, et al. Fruit and vegetable intake and the risk of cardiovascular disease, total cancer and all-cause mortality-a systematic review and dose-response meta-analysis of prospective studies. Int J Epidemiol 2017 Jun 01;46(3):1029-1056 [FREE Full text] [doi: 10.1093/ije/dyw319] [Medline: 28338764]
7. Frutas y verduras – esenciales en tu dieta Año Internacional de las Frutas y Verduras, 2021. Documento de antecedentes. Rome, Italy: Food and Agriculture Organization; 2020.
8. Hooper L, Abdelhamid A, Bunn D, Brown T, Summerbell C, Skeaff C. Effects of total fat intake on body weight. Cochrane Database Syst Rev 2015 Aug 07(8):CD011834. [doi: 10.1002/14651858.CD011834] [Medline: 26250104]
9. Nishida C, Uauy R. WHO Scientific Update on health consequences of trans fatty acids: introduction. Eur J Clin Nutr 2009 May 06;63 Suppl 2(S2):S1-S4. [doi: 10.1038/ejcn.2009.13] [Medline: 19424215]
10. World Health Organization. Diet Nutrition and the Prevention of Chronic Diseases Report of a Joint WHO/FAO Expert Consultation. Geneva: World Health Organization; 2003.
11. World Health Organization. Food-based dietary guidelines in the WHO European Region. WHO Regional Office for Europe. 2003. URL: https://apps.who.int/iris/handle/10665/107490 [accessed 2022-05-23]
12. World Health Organization. Guideline: Sugars Intake for Adults and Children. Geneva: World Health Organization; 2015.
13. Williamson C, Kelly P, Tomasone JR, Bauman A, Mutrie N, Niven A, et al. A modified Delphi study to enhance and gain international consensus on the Physical Activity Messaging Framework (PAMF) and Checklist (PAMC). Int J Behav Nutr Phys Act 2021 Aug 19;18(1):108 [FREE Full text] [doi: 10.1186/s12966-021-01182-z] [Medline: 34412638]
14. Williamson C, Baker G, Mutrie N, Niven A, Kelly P. Get the message? A scoping review of physical activity messaging. Int J Behav Nutr Phys Act 2020 Apr 15;17(1):51 [FREE Full text] [doi: 10.1186/s12966-020-00954-3] [Medline: 32295613]
15. Conroy DE, Hojjatinia S, Lagoa CM, Yang C, Lanza ST, Smyth JM. Personalized models of physical activity responses to text message micro-interventions: a proof-of-concept application of control systems engineering methods. Psychol Sport Exerc 2019 Mar;41:172-180 [FREE Full text] [doi: 10.1016/j.psychsport.2018.06.011] [Medline: 30853855]
16. Latimer AE, Brawley LR, Bassett RL. A systematic review of three approaches for constructing physical activity messages: what messages work and what improvements are needed? Int J Behav Nutr Phys Act 2010 May 11;7(1):36 [FREE Full text] [doi: 10.1186/1479-5868-7-36] [Medline: 20459779]
17. Bassett RL, Ginis KAM. Risky business: the effects of an individualized health information intervention on health risk perceptions and leisure time physical activity among people with spinal cord injury. Disabil Health J 2011 Jul;4(3):165-176. [doi: 10.1016/j.dhjo.2010.12.001] [Medline: 21723523]
18. Bassett-Gunter RL, Martin Ginis KA, Latimer-Cheung AE. Do you want the good news or the bad news? Gain- versus loss-framed messages following health risk information: the effects on leisure time physical activity beliefs and cognitions. Health Psychol 2013 Dec;32(12):1188-1198. [doi: 10.1037/a0030126] [Medline: 23088175]
19. Lin I, Liao T. A survey of blockchain security issues and challenges. Int J Netw Secur 2017;19(5):653-659. [doi: 10.6633/IJNS.201709.19(5).01]
20. Kim JW. Analysis of blockchain ecosystem and suggestions for improvement. J Inform Commun Convergence Eng 2021 Mar 31;19(1):8-15. [doi: 10.6109/jicce.2021.19.1.8]
21. Monrat AA, Schelen O, Andersson K. A survey of blockchain from the perspectives of applications, challenges, and opportunities. IEEE Access 2019;7:117134-117151. [doi: 10.1109/access.2019.2936094]

22. Oyinloye DP, Teh JS, Jamil N, Alawida M. Blockchain consensus: an overview of alternative protocols. Symmetry 2021 Jul 27;13(8):1363. [doi: 10.3390/sym13081363]
23. Hussien HM, Yasin SM, Udzir NI, Ninggal MI, Salman S. Blockchain technology in the healthcare industry: trends and opportunities. J Industrial Inform Integration 2021 Jun;22:100217. [doi: 10.1016/j.jii.2021.100217]
24. Bukhari D. Blockchain technology: a bibliometric analysis. In: HCI International 2020 - Posters. Cham: Springer; 2020.
25. Lopez-Barreiro J, Alvarez-Sabucedo L, Garcia-Soidan JL, Santos-Gago JM. Use of blockchain technology in the domain of physical exercise, physical activity, sport, and active ageing: a systematic review. Int J Environ Res Public Health 2022 Jul 02;19(13):8129 [FREE Full text] [doi: 10.3390/ijerph19138129] [Medline: 35805788]
26. Buterin V. Ethereum White Paper. 2013. URL: https://bibbase.org/network/publication/buterin-ethereumwhitepaperanextgenerationsmartcontractdecentralizedapplicationplatform-2013 [accessed 2022-05-15]
27. Beniiche A. A study of blockchain oracles. arXiv 2020 [FREE Full text]
28. Cai W, Wang Z, Ernst JB, Hong Z, Feng C, Leung VC. Decentralized applications: the blockchain-empowered software system. IEEE Access 2018;6:53019-53033. [doi: 10.1109/access.2018.2870644]
29. Gómez-Díaz R, García-Rodríguez A. Bibliotecas, juegos y gamificación: una tendencia de presente con mucho futuro. ThinKEPI 2018 Apr 25;12:125. [doi: 10.3145/thinkepi.2018.13]
30. Miau S, Yang J. Bibliometrics-based evaluation of the Blockchain research trend: 2008 – March 2017. Technol Analysis Strategic Manage 2018 Jan 31;30(9):1029-1045. [doi: 10.1080/09537325.2018.1434138]
31. Alsalamah HA, Nasser S, Alsalamah S, Almohana AI, Alanazi A, Alrrshaid F. Wholesome coin: a pHealth solution to reduce high obesity rates in gulf cooperation council countries using cryptocurrency. Front Blockchain 2021 Jul 12;4. [doi: 10.3389/fbloc.2021.654539]
32. Frikha T, Chaari A, Chaabane F, Cheikhrouhou O, Zaguia A. Healthcare and fitness data management using the IoT-based blockchain platform. J Healthc Eng 2021 Jul 9;2021:9978863-9978812 [FREE Full text] [doi: 10.1155/2021/9978863] [Medline: 34336176]
33. Jamil F, Kahng HK, Kim S, Kim D. Towards secure fitness framework based on IoT-enabled blockchain network integrated with machine learning algorithms. Sensors (Basel) 2021 Feb 26;21(5):1640 [FREE Full text] [doi: 10.3390/s21051640] [Medline: 33652773]
34. Jamil F, Qayyum F, Alhelaly S, Javed F, Muthanna A. Intelligent microservice based on blockchain for healthcare applications. Comput Material Continua 2021;69(2):2513-2530. [doi: 10.32604/cmc.2021.018809]
35. Cao P, Zhu G, Zhang Q, Wang F, Liu Y, Mo R. Blockchain-enabled HMM model for sports performance prediction. IEEE Access 2021;9:40255-40262. [doi: 10.1109/access.2021.3064261]
36. Hong Y, Park DW. Big data and blockchain to improve performance of professional sports teams. ASM Sci J 2020;13(1):19-27.
37. Yu S. Application of blockchain-based sports health data collection system in the development of sports industry. Mobile Inform Syst 2021 Jun 10;2021:1-6. [doi: 10.1155/2021/4663147]
38. Ma F. Design of running training assistance system based on blockchain technology in wireless network. J Wireless Com Network 2021 Jan 31;2021(1). [doi: 10.1186/s13638-021-01897-4]
39. Shan Y, Mai Y. Research on sports fitness management based on blockchain and Internet of Things. J Wireless Com Network 2020 Oct 15;2020(1). [doi: 10.1186/s13638-020-01821-2]
40. Mulyati, Rahardja U, Hardini M, Al Nasir AL, Aini Q. Taekwondo sports test and training data management using blockchain. In: Proceedings of the 2020 Fifth International Conference on Informatics and Computing (ICIC). 2020 Presented at: 2020 Fifth International Conference on Informatics and Computing (ICIC); Nov 03-04, 2020; Gorontalo, Indonesia. [doi: 10.1109/icic50835.2020.9288598]
41. Khezr S, Benlamri R, Yassine A. Blockchain-based model for sharing activities of daily living in healthcare applications. In: Proceedings of the 2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech). 2020 Presented at: 2020 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech); Aug 17-22, 2020; Calgary, AB, Canada. [doi: 10.1109/dasc-picom-cbdcom-cyberscitech49142.2020.00109]
42. Rahman MA, Hossain MS, Loukas G, Hassanain E, Rahman SS, Alhamid MF, et al. Blockchain-based mobile edge computing framework for secure therapy applications. IEEE Access 2018;6:72469-72478. [doi: 10.1109/access.2018.2881246]
43. Rahman M, Abualsaud K, Barnes S, Rashid M, Abdullah S. A natural user interface and blockchain-based in-home smart health monitoring system. In: Proceedings of the 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT). 2020 Presented at: 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT); Feb 02-05, 2020; Doha, Qatar. [doi: 10.1109/iciot48696.2020.9089613]
44. Rupasinghe T, Burstein F, Rudolph C, Strange S. Towards a blockchain based fall prediction model for aged care. In: Proceedings of the Australasian Computer Science Week Multiconference. 2019 Presented at: ACSW 2019: Australasian Computer Science Week 2019; Jan 29 - 31, 2019; Sydney NSW Australia. [doi: 10.1145/3290688.3290736]

45. Silva CA, Aquino GS, Melo SR, Egídio DJ. A fog computing-based architecture for medical records management. Wireless Commun Mobile Computing 2019 Feb 27;2019:1-16. [doi: 10.1155/2019/1968960]
46. Spinsante S, Poli A, Mongay Batalla J, Krawiec P, Dobre C, Bǎjenaru L, et al. Clinically-validated technologies for assisted living. J Ambient Intell Human Comput 2021 Aug 16;14(3):2095-2116. [doi: 10.1007/s12652-021-03419-y]
47. Velmovitsky PE, Miranda PA, Vaillancourt H, Donovska T, Teague J, Morita PP. A blockchain-based consent platform for active assisted living: modeling study and conceptual framework. J Med Internet Res 2020 Dec 04;22(12):e20832 [FREE Full text] [doi: 10.2196/20832] [Medline: 33275111]
48. Momma H, Kawakami R, Honda T, Sawada SS. Muscle-strengthening activities are associated with lower risk and mortality in major non-communicable diseases: a systematic review and meta-analysis of cohort studies. Br J Sports Med 2022 Jul 28;56(13):755-763 [FREE Full text] [doi: 10.1136/bjsports-2021-105061] [Medline: 35228201]
49. Leenders M, Sluijs I, Ros MM, Boshuizen HC, Siersema PD, Ferrari P, et al. Fruit and vegetable consumption and mortality: European prospective investigation into cancer and nutrition. Am J Epidemiol 2013 Aug 15;178(4):590-602. [doi: 10.1093/aje/kwt006] [Medline: 23599238]
50. Chowdhury MA, Hossain N, Kashem MA, Shahid MA, Alam A. Immune response in COVID-19: a review. J Infect Public Health 2020 Nov;13(11):1619-1629 [FREE Full text] [doi: 10.1016/j.jiph.2020.07.001] [Medline: 32718895]
51. Klika B, Jordan C. High-intensity circuit training using body weight: maximum results with minimal investment. ACSM's Health Fitness J 2013;17(3):8-13. [doi: 10.1249/FIT.0b013e31828cb1e8]
52. Jayedi A, Gohari A, Shab-Bidar S. Daily step count and all-cause mortality: a dose-response meta-analysis of prospective cohort studies. Sports Med 2022 Jan 21;52(1):89-99. [doi: 10.1007/s40279-021-01536-4] [Medline: 34417979]
53. Lopez-Barreiro J, Hernandez-Lucas P, Garcia-Soidan JL, Romo-Perez V. Effects of an eccentric training protocol using gliding discs on balance and lower body strength in healthy adults. J Clin Med 2021 Dec 19;10(24):5965 [FREE Full text] [doi: 10.3390/jcm10245965] [Medline: 34945261]
54. Kaleido. URL: https://www.kaleido.io/ [accessed 2023-04-10]
55. Santos-Gago JM, Ramos-Merino M, Alvarez-Sabucedo LM. Identification of free and WHO-compliant handwashing moments using low cost wrist-worn wearables. IEEE Access 2021;9:133574-133593. [doi: 10.1109/access.2021.3115434]
56. Noh JH, Kwon HY. A study on smart city security policy based on blockchain in 5G Age. In: Proceedings of the 2019 International Conference on Platform Technology and Service (PlatCon). 2019 Presented at: 2019 International Conference on Platform Technology and Service (PlatCon); Jan 28-30, 2019; Jeju, Korea (South). [doi: 10.1109/platcon.2019.8669406]

Abbreviations

BC: blockchain
dApp: decentralized application
FAO: Food and Agriculture Organization
GDPR: General Data Protection Regulation
HIIT: high-intensity interval training
NFT: nonfungible token
OECD: Organization for Economic Co-operation and Development
PA: physical activity
PAG: physical activity guidelines for Americans
POA: proof of authority
POW: proof of work
SC: smart contract
WHO: World Health Organization

Edited by A Mavragani; submitted 07.11.22; peer-reviewed by N Mathew, P Moriggl; comments to author 26.01.23; revised version received 14.02.23; accepted 03.03.23; published 19.04.23.

Please cite as:
Lopez-Barreiro J, Alvarez-Sabucedo L, Garcia-Soidan JL, Santos-Gago JM
Creation of a Holistic Platform for Health Boosting Using a Blockchain-Based Approach: Development Study
Interact J Med Res 2023;12:e44135
URL: https://www.i-jmr.org/2023/1/e44135
doi:10.2196/44135
PMID:37074766

Original Paper

The Use of Digital Technology for COVID-19 Detection and Response Management in Indonesia: Mixed Methods Study

Dewi Nur Aisyah[1,2*], PhD; Alfiano Fawwaz Lokopessy[1,3*], BSc; Maryan Naman[4*], MSc; Haniena Diva[1*], MPH; Logan Manikam[2,4*], PhD; Wiku Adisasmito[1,3*], PhD; Zisis Kozlakidis[5*], PhD

[1]Indonesia One Health University Network, Depok, Indonesia

[2]Department of Epidemiology and Public Health, Institute of Epidemiology and Health Care, University College London, London, United Kingdom

[3]Faculty of Public Health, Universitas Indonesia, Depok, Indonesia

[4]Aceso Global Health Consultants Pte Limited, Singapore, Singapore

[5]International Agency for Research on Cancer World Health Organization, Lyon, France

[*]all authors contributed equally

Corresponding Author:
Logan Manikam, PhD
Department of Epidemiology and Public Health
Institute of Epidemiology and Health Care
University College London
1-19 Torrington Place
London, WC1E 7HB
United Kingdom
Phone: 44 (0)207 679 2000
Email: logan.manikam.10@ucl.ac.uk

Abstract

Background: The COVID-19 pandemic has triggered a greater use of digital technologies as part of the health care response in many countries, including Indonesia. It is the world's fourth-most populous nation and Southeast Asia's most populous country, with considerable public health pressures.

Objective: The aim of our study is to identify and review the use of digital health technologies in COVID-19 detection and response management in Indonesia.

Methods: We conducted a literature review of publicly accessible information in technical and scientific journals, as well as news articles from September 2020 to August 2022 to identify the use case examples of digital technologies in COVID-19 detection and response management in Indonesia.

Results: The results are presented in 3 groups, namely (1) big data, artificial intelligence, and machine learning (technologies for the collection or processing of data); (2) health care system technologies (acting at the public health level); and (3) COVID-19 screening, population treatment, and prevention population treatment (acting at the individual patient level). Some of these technologies are the result of government-academia-private sector collaborations during the pandemic, which represent a novel, multisectoral practice in Indonesia within the public health care ecosystem. A small number of the identified technologies pre-existed the pandemic but were upgraded and adapted for current needs.

Conclusions: Digital technologies were developed in Indonesia during the pandemic, with a direct impact on supporting COVID-19 management, detection, response, and treatment. They addressed different areas of the technological spectrum and with different levels of adoption, ranging from local to regional to national. The indirect impact of this wave of technological creation and use is a strong foundation for fostering future multisectoral collaboration within the national health care system of Indonesia.

(Interact J Med Res 2023;12:e41308) doi:10.2196/41308

KEYWORDS

COVID-19; Indonesia; digital technology; digital innovation; digital health; response management; robot innovation; decontamination

Introduction

For the past 2 decades, Indonesia has faced various outbreaks of emerging and re-emerging infectious diseases, such as measles [1], SARS [2], MERS [3], H5N1 [4], H1N1 [5], and, most recently, COVID-19 [6]. These have challenged individuals, health care systems, and infrastructures on how to best prevent wider community transmission, how to treat patients effectively, and how to suppress cases until finally the disease outbreak can be controlled. The COVID-19 pandemic has created a global emergency that requires many different parties to collaborate in a coordinated and systematic manner to implement health policies and community actions. As with previous crises, they can act as a catalyst and trigger new ideas and innovations [7-9]. As a result, multiple breakthroughs and scaled-up implementations in digital technology have also emerged during this pandemic [10]. Some of them have contributed to the surveillance, detection, or responses of positive cases and their direct contacts [11,12], which helps policy makers manage the pandemic, including in Indonesia. Some are pre-existing technologies, such as remote consultations, whose use is being enhanced or modified for the pandemic [13-15].

Several recently published studies have presented the use of digital technologies during the COVID-19 pandemic in several countries. For example, Whitelaw et al [16] grouped technologies used in more than 10 countries into different functions for pandemic planning and responses, such as tracking, screening for infection, contact tracing, quarantine and self-isolation, and clinical management. Another study in Saudi Arabia also found the use of digital technologies at different stages of pandemic responses, namely at digital screening, surveillance, contact notification, and follow-up [17]. Although there might not exist an international consensus on the grouping of these technologies currently, these models remain useful in being able to navigate and study the field as well as the impact of implementing such technologies.

More specifically, in Indonesia, the use for many of these digital technologies has been promoted and enhanced by the government throughout the pandemic, through various innovation and research programs [18] and by the incorporation of digital technologies as part of routine data collection activities, supporting evidence-based policy making [19]. The surge in the number of COVID-19 cases and deaths has resulted in the tightening of barrier measures (eg, masks, personal protective equipment) and population movement restrictions (eg, curfew in large urban centers). These have placed a strategic focus on digital technology use in responding to the pandemic, such as through the recent government partnerships with telemedicine apps to provide free remote medical consultation for faster responses and easing potential crowding at hospitals [20]. During the pandemic, Indonesia, the most populous nation of Southeast Asia, experienced waves with corresponding sharp rises in cases and deaths (eg, 60% of positive cases increased in the week of June 19-29, 2021, alone [7] as compared to the immediately prior week). Altogether, Indonesia had recorded over 2.3 million cumulative cases and over 61,000 cumulative deaths as of July 5, 2021 [8], which led to government restrictions on people's mobility in its most populous islands of Java and Bali from July 3 to 20, 2021 [10]. This imposition of population movement restrictions was repeated on several occasions throughout 2021 and 2022. Drawing from the use of digital technologies in other countries, as well as the authors' expertise on the implementation of digital health technologies within Indonesia [21], this paper aims to identify and classify the use of digital technologies for COVID-19 detection and response management in Indonesia. By combining multiple sources in English and Indonesian, in scientific peer-reviewed literature, as well as information released by governmental departments, the authors believe that they can provide an exhaustive and detailed narrative review of the emerging digital health landscape.

Methods

Search Strategy and Selection Criteria

We considered any studies that reported the use of mobile apps or digital technologies or both that support COVID-19 pandemic control in Indonesia. We conducted a systematic literature search using OVID Embase, OVID Medline, and PubMed databases, as well as the Google Scholar search engine, using the terms "digital" or "technology*" or "robotic" or "tracing" or "dashboard" or "telemedicine" AND "COVID-19" or "coronavirus disease" or "SARS-Cov-2" AND "Indonesia." To expand the literature search, we also conducted manual searches through publicly accessible regional and national official announcements, press releases, and published data within Indonesia. We included studies that were published in English after January 2020 (ie, from the time of the first confirmed positive SARS-CoV-2 case in Indonesia to the time of the authoring of this paper) up to August 2022.

The inclusion criteria were (1) usage/practice related to the COVID-19 pandemic detection and response in Indonesia and (2) the use of digital technology and digitization of services directly related to the pandemic needs. We excluded studies that were not relevant to COVID-19 pandemic control and response, as well as studies that were not implemented in Indonesia. There were no exclusions in relation to the type of technology creator or the type of technology user (eg, public, private, or mixed consortia). All identified manuscripts were reviewed independently by 2 authors, and all those that referred to or presented specific information in relation to the implementation of a digital health technology for COVID-19 in Indonesia were included. The resulting list was confirmed by independent review by a third author. The review process followed Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.

Thematic Validation

Due to the limited number of identified manuscripts and to reach thematic validation, the authors interviewed (noncompensated) additionally 10 individuals from key stakeholder organizations (ie, 2 individuals from each of the institutions involved in digital technology use for COVID-19, namely the Ministry of Health, the COVID-19 National Taskforce, Telkom Indonesia, the Indonesian Red Cross, and the DKI Jakarta Local Government). This allowed for independent thematic validation, as derived from the initial data collection round. Subsequently, the authors collected, studied, and organized all the information obtained to determine (1) key practices when digital technologies were used as well as (2) any lessons learned. Preliminary thematic groups were generated and were linked to interview texts using traditional content analysis. Emerging themes were discussed and presented at team meetings. During these meetings, discordant classifications were discussed until a consensus was reached.

Ethics Statement

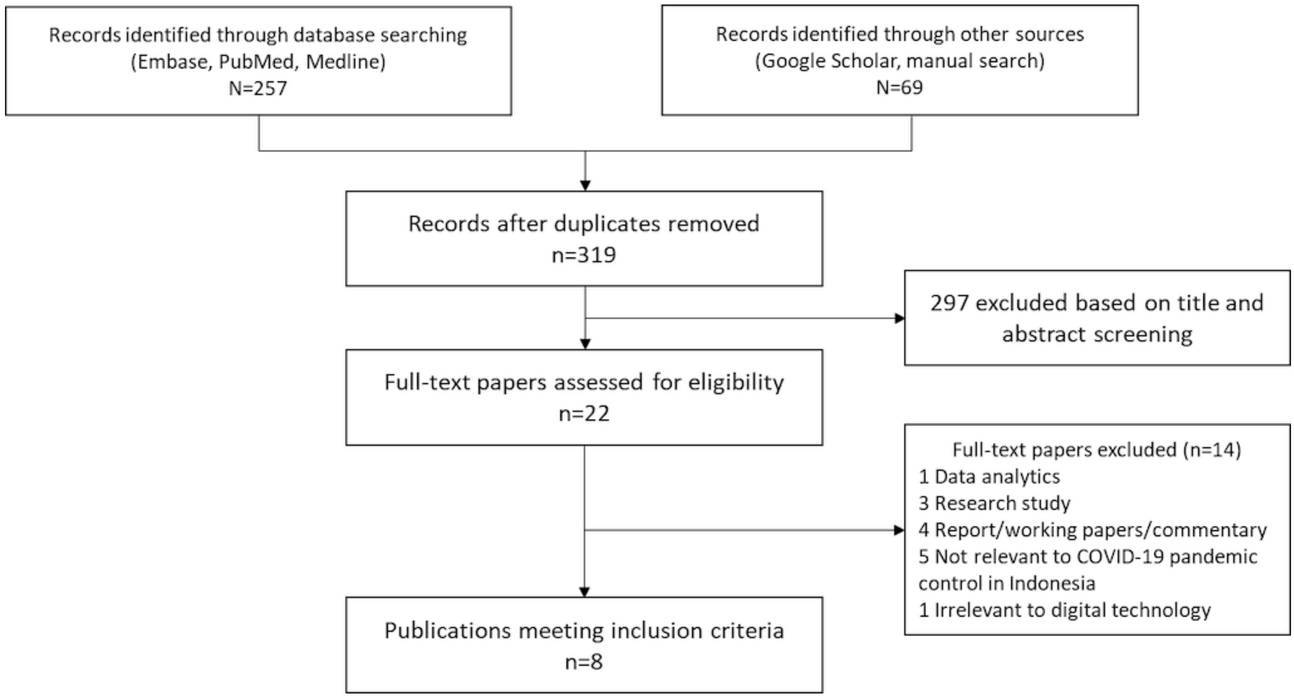
There was no patient involvement in this review. All interviews were conducted anonymously. Although participants were not signing a separate consent form, consent was obtained by completion of the interview. Thus, a waiver was obtained from the International Agency for Research on Cancer Ethics Committee (reference number 22-11362).

Results

Paper Screening

After excluding duplicates, we retrieved 319 papers from 3 databases and 1 from the gray literature. Most papers had simple mentions of the need to implement technological solutions within Indonesia, but few contained actual examples of doing so. Of the 319 papers, 22 (6.9%) met the criteria for full-text review. Finally, we identified 8 (36.4%) studies describing the use of digital technology to support the COVID-19 pandemic control and response in Indonesia (Figure 1). The list of included information and communications technology (ICT) tools is provided in Multimedia Appendix 1.

Figure 1. Paper screening following PRISMA guidelines. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.



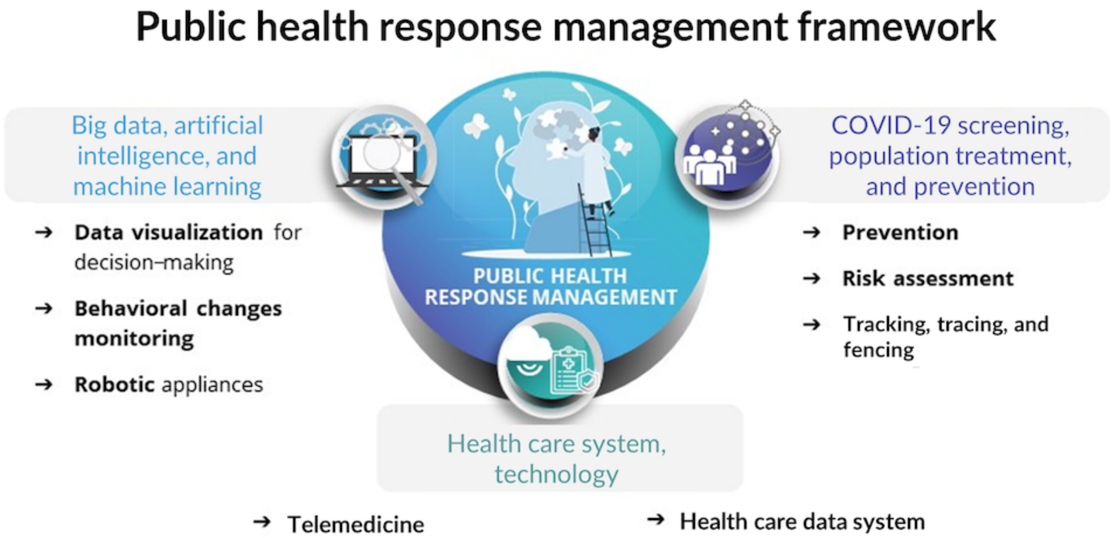
In Indonesia, as with many countries globally, a public health response and management system was built and implemented through COVID-19 detection, prevention, treatment, and monitoring. In total, 36 digital technologies were identified supporting the management aspects and 11 digital technologies identified supporting treatment, such as telemedicine apps providing free COVID-19 consultation and treatment for patients with mild symptoms and information for asymptomatic individuals. Each of the latter technologies facilitates a mix of telemedicine consultations and medicine delivery services for patients with COVID-19 [22]. Among the latter set of 11 technologies, 10 (90.9%) predated the pandemic, and their use was enhanced during the pandemic, while 1 (9.1%) was launched during the pandemic. Namely, those 11 digital apps are Alodok (PT Sumo Teknologi Solusi), GetWell (PT Telemedika Teknologi Indonesia), Good Doctor (Good Doctor Tech), GrabHealth (PT. Grab Indonesia), Halodoc (PT Media Dokter Investama), KlikDokter (PT Medika Komunika Teknologi), KlinikGo (PT Medika Nuswantara Digital), Link Sehat (PT Link Medis Sehat), Milvik Dokter (PT Milvik Indonesia), ProSehat (PT Atoma Medical), SehatQ (PT SehatQ Harsana Emedika), and YesDok (PT Yes Dok Indonesia).

Additionally, 9 robot innovation products were developed to support health workers treating hospitalized patients. Namely, the 9 robotics innovation products are Robot RAISA TIARA, Robot RAISA BCL, Robot Dekontaminasi (decontamination robot), Smart Syringe Pump, Autonomous UVC Mobile Robot, Robot Violeta, Smart Telemedicine Robot "Win-MTA," Service Robot, and Doctor Representative Robot. Their intended aims are to prevent within-hospital transmission or to alleviate health care workers' work burden. In addition, 2 health care data systems were developed assisting patients' treatment, namely SIRANAP (Sistem Informasi Rawat Inap) and Blood Plasma Donor. These 2 data systems help individuals with COVID-19 to find available hospital beds and blood plasma.

It is worth noting that the number of such available apps has grown quickly since the pandemic started in Indonesia. Overall, 1 new telemedicine app, 1 new health care data system app, 5 new mobile phone–based apps, and 1 big data/machine learning analytics platform were developed and launched during the pandemic. These digital technologies have been developed almost entirely as multisectoral government-university-private sector partnerships.

To study and classify these digital emerging technologies in a more systematic way, they were classified into 3 major user groups, as illustrated in Figure 2, namely (1) big data, artificial intelligence, and machine learning (ie, technologies for the collection, integration, ingestion, and processing of data, as well as robotic systems); (2) health care system technologies (ie, technologies acting at the individual level), and (3) COVID-19 screening, population treatment, and prevention (ie, technologies acting at the population level). These 3 major groups are further expanded next.

Figure 2. The use of digital technologies for detection and response management of the COVID-19 pandemic in Indonesia, divided into 3 main categories.



Big Data, Artificial Intelligence, and Machine Learning

Three types of uses were identified within this theme: (1) data visualization, (2) behavioral change monitoring, and (3) robotic appliances. Although the first 2 have primarily contributed to evidence-based policy making, the third supported health workers in treating patients.

In terms of data collation and visualization tools for decision-making support, Indonesia has 34 provinces that comprise districts and cities with decentralized local governments. Of these, 7 (20.6%) provinces developed websites and dashboards independently and customized them to their needs and local context to aggregate and manage COVID-19–related data, to keep the public and policy makers updated with the current situation, and to fact-check available information and provide relevant contact numbers for seeking medical services and treatment. Table 1 shows some examples of those websites and their respective dashboards.

For health protocol compliance/behavioral change monitoring, health protocol compliance–monitoring systems were all analyzed using the interoperable platform *Bersatu Lawan COVID-19 (BLC)*, which translates as "United Against COVID-19," that any government level can access, be it national, provincial, or district/city government. This system uses big data analysis that allows real-time monitoring of compliance to health protocols, such as mask wearing and keeping social distance, to inform policy makers on public behavioral changes [23]. Figure 3 depicts its function and interoperability with other COVID-19–related data supplied to Indonesian policy makers.

One of the flagship products of the BLC is the Health Protocol Compliance Monitoring System that has been supporting policy makers nationwide in observing the compliance of key public spaces. Through this data system, the Indonesian military (TNI), police (POLRI) personnel, and volunteer community ambassadors can submit reports from key public spaces, such as markets and train stations, on whether people have complied with wearing face masks and keeping a social distance of minimum 1 m. Figure 4 shows the dashboard and resulting reports created from these analyses made publicly accessible [24]. Since its launch in October 2020 and until December 2021, the BLC app has facilitated more than 211.3 million health protocol compliance–monitoring reports, with 721.4 million people monitored and 165,537,934 locations under observation in all of the 514 districts/cities in the 34 provinces in Indonesia.

In regard to robotic appliances, as the COVID-19 pandemic necessitated limited physical contact between health workers and patients, innovations emerged in robotic technologies by various government and higher education institutions assisting in the treating of patients with COVID-19 in various hospitals and institutions [17]. A study in China found a similar increase in the use of robotics to minimize physical contact and also found that robotics can help reduce the risks of health care workers getting infected by COVID-19 [25]. Additionally, robotic technologies may help in processing information, delivering food or medicine, carrying out temperature checks, and carrying out disinfection tasks. Table 2 compiles a list of the robotic innovations introduced in Indonesia in direct relation to clinical COVID-19 management [17].

Table 1. List of several COVID-19–related data collation and visualization tools.

Website or dashboard name	Functions
Pikobar (Pusat Informasi & Koordinasi COVID-19 di Provinsi Jawa Barat)	Presenting statistical updates on new positive cases, self-isolating people, hospitalization, suspected cases, probabilities, and contact tracing
Executive Information System Dinkes Provinsi DKI Jakarta	Presenting real-time information about isolation bed availability
Jakarta Tanggap COVID-19	Presenting information about COVID-19 cases in Jakarta
Pusat Informasi & Koordinasi Kota Depok Jawa Barat	Presenting information about COVID-19 cases in Depok and providing hotline services
COVID-19 NTB (Nusa Tenggara Barat)	Presenting information about COVID-19 cases in Depok and providing hotline services and fact-checking service
Sulsel Tanggap COVID-19	Presenting information about COVID-19 cases in Depok and providing hotline services and fact-checking service
Pusat Informasi COVID-19 Provinsi Maluku	Presenting information about COVID-19 cases in Depok and providing hotline services and fact-checking service

Figure 3. Functions and interoperability of BLC national data. BLC: Bersatu Lawan COVID-19; PPKM; Pemberlakuan Pembatasan Kegiatan Masyarakat (Community Mobility Restriction).

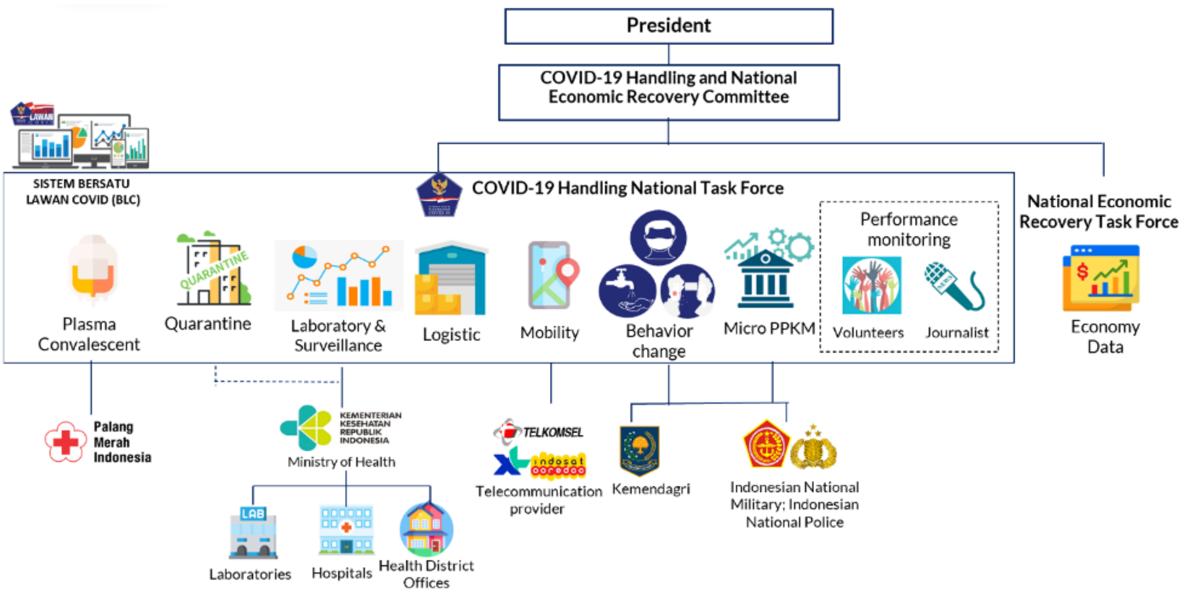


Figure 4. Dashboard for the Health Protocol Compliance Monitoring System.

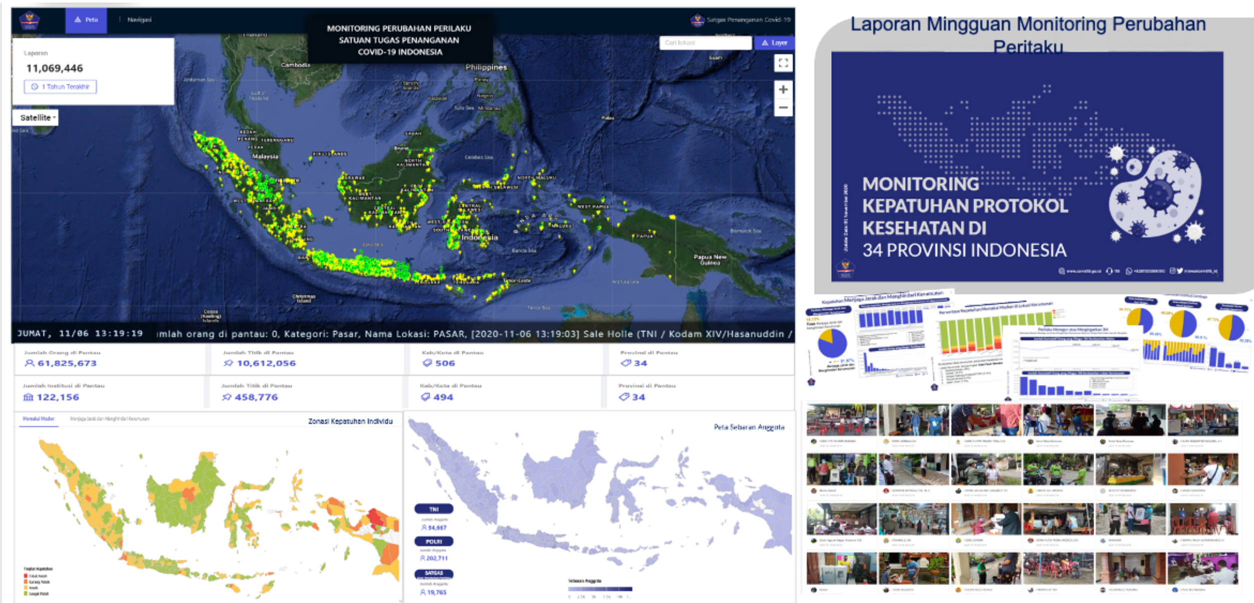


Table 2. List of robot innovation products related to COVID-19 management in Indonesia.

Robot name	Innovator	Functions	Other remarks	Status
Robot RAISA	Institut Teknologi Sepuluh Nopember Surabaya and Universitas Airlangga	Providing medical assistant and nurse-like functions	The robot can carry up to 5 kg of weight, has a camera, and can facilitate 2-way communication between health workers and patients.	Used in the University of Airlangga Hospital
Robot RAISA TIARA, Robot RAISA BCL[a]	Institut Teknologi Sepuluh Nopember Surabaya and Universitas Airlangga	Opening ICU[b] doors and checking the patient's body temperature, oxygen saturation, and heartbeat.	The robot can also remotely observe infusion drops and urine production from up to a 5 km distance and includes a 360° rotation function.	Prototype produced and used in at least 6 hospitals: Universitas Airlangga (UNAIR) Hospital, Dr Soetomo Hospital, Husada Utama Surabaya Hospital, Saiful Anwar Malang Hospital, Wisma Atlit, RSI Surabaya
Robot Dekontaminasi	Institut Teknologi Sepuluh Nopember Surabaya and Universitas Airlangga	Decontaminating used items and used personal protective equipment (PPE)	The robot is operated using a remote control.	Prototype produced and ready to be used
Smart Syringe Pump	Institut Teknologi Sepuluh Nopember Surabaya and Universitas Airlangga	Automatically filling medicine administered to the patient with a set schedule	A mobile app remotely operates the pump.	Prototype produced and ready to be used
Autonomous UVC Mobile Robot	Telkom University and Indonesia Institute of Sciences (LIPI)	Disinfecting and sterilizing isolation rooms for patients with COVID-19	The robot is equipped with UVC light to effectively kill coronavirus.	Prototype produced and tested and has been used in hospitals in West Java Province
Robot Violeta	Institut Teknologi Sepuluh Nopember Surabaya and Universitas Airlangga	Eliminating or decelerating the growth of pathogens by using UV 200-300 nm waves	Using remotely operated UV light from a 1-2 m distance, the robot takes 10-15 minutes to perform the sterilization task.	Used in the University of Airlangga Hospital
Smart Telemedicine Robot "Win-MTA"	Universitas Gadjah Mada and PT. Maetala Visionaire Tecnologia	Disinfecting objects using >39°C temperature, delivering medicine and prescriptions to patients, and sending patients' data (temperature, GPS location)	N/A[c]	Remains a prototype
Robot Pelayan	Universitas Gadjah Mada and Academic Hospital, Universitas Gadjah Mada	Automatically locating the patient's room and providing automated medicine and food delivery services to rooms	N/A	Remains a prototype
Doctor Representative Robot (Doper)	Telkom University	Facilitating medical and nutrition consultation without physical contact with health workers	N/A	Under development

[a]BLC: Bersatu Lawan COVID-19.

[b]ICU: intensive care unit.

[c]N/A: not applicable.

Health Care System Technology

Three types of uses were identified within this theme, namely telemedicine, digital support for quarantine, and health care data systems. All of them were developed with the aim to provide services for those self-isolating, seeking medical treatment, or seeking hospitalization.

Telemedicine and Quarantine Digital Support

Telemedicine is an example of digital technology already developed and used for health purposes before the pandemic within Indonesia, albeit sporadically. The World Health Organization (WHO) already discussed the potential benefits of telemedicine in overcoming distance barriers and speeding up health care delivery, while considering current high technological costs [26]. Portnoy et al [27] observed a rising use of telemedicine in the 2-3 years preceding the pandemic, and further growth during the pandemic, as quick medical advice could be provided to patients with mild symptoms, thus allowing for timely, accurate information dissemination and indirectly limiting community transmission. Further examples include apps allowing health care workers who are self-isolating to continue supporting patients remotely [28]. Moreover, research in China found this approach to be an effective solution to minimize infection risks to health workers by minimizing physical contact [29].

Despite these promising benefits, challenges, such as high cost [26] and integration with extant national health care systems, persist. In Indonesia, the surge in COVID-19 cases and the tighter population movement restrictions in July 202 brought to the fore a partnership between the government and 11 telemedicine mobile apps [22]. This umbrella partnership enabled the government to cover the cost of services and allowed patients to access medical advice and medications for free. Thus, the Indonesian government responded to the public health pressures rising from the surge in COVID-19–positive cases, while temporarily overcoming the cost challenge that is faced by telemedicine apps in other countries. It also allowed for limited integration with the national COVID-19 surveillance mechanism by providing access to COVID-19 testing. Table 3 compiles a list of the telemedicine mobile app partners of the Ministry of Health of Indonesia.

Table 3. List of telemedicine mobile app partners of the Ministry of Health.

Telemedicine mobile app	Year of launch	Number of users	Features
Alodokter	2014	40 million	Chat with doctors, health articles, consultation booking with doctors, online drug store, Alodokter insurance
GetWell	2021	>5000	Chat and video calls with doctors, personal health records, 24/7 panic button, health articles, integrated with PeduliLindungi app (government's COVID-19 and vaccination status app)
Good Doctor	2020	>1 million	Consultation with doctors, online drug store and delivery, doctor appointment-booking system, insurance claim
Halodoc	2016	20 million (2021)	Chat and video calls with doctors, see the doctor's experience and rating, a health store, booking a hospital doctor, getting a laboratory test, linking insurance
KlikDokter	2015	>1 million	Live chat with doctors, hospital appointment booking, pregnancy journey tracker (via the HalloBumil app), period tracker and calendar, heart and diabetes risk measurement, BMI calculator
KlinikGo	2022	>10	Online booking for hospital appointment
Link Sehat	2020	>10,000	Consultation with doctors, COVID-19–testing appointment, hospital schedule booking, online medical assistance, health articles
Milvik Dokter	2019	>10,000	Consultation with doctors, medicine and laboratory check
ProSehat	2015	>100,000	Online chat with doctors, home visits, getting a laboratory test, drive-through immunization, product delivery
SehatQ	2019	>500,000	Chat with doctors, buy drugs, pregnancy consultation and discussion forum, mental health, and other services
YesDok	2017	>100,000	Consultation with doctors (prediagnosis, first aid, education, medicine recommendation, and consultation playback), drug delivery

In addition to the telemedicine mobile apps partners of the Ministry of Health, many hospitals also provided telemedicine service using mobile apps, including the Cipto Mangunkusumo Hospital with the SiapDok app, the Siloam Hospital Group with the AIDO app, and the YARSI Hospital with the MAUDOK app. These telemedicine apps were organization specific, with the aim to help hospital patients arrange online appointments as well as receive health consultations with doctors.

The 2 most popular telemedicine apps in Indonesia are Halodoc, with monthly active users reaching 20 million [30], followed by Alodokter, with monthly active users being around 18 million [31]. According to another survey conducted by the Katadata Insight Center, during the COVID-19 pandemic, the most popular telemedicine apps were Halodoc (46.5%), followed by telemedicine provided by hospitals (41.7%) and Alodokter (35.7%) [32].

In principle, the features of teleconsultation with doctors in these apps are similar. The differentiating features are the user interface (UI) and user experience (UX) aspects, as well as the promotion(s) offered by the operating company, such as medicine delivery and cashback offers.

Health Care Data System (Hospital, Laboratory, and Tracing)

The surging cases and the rising needs for hospitalization have created unprecedented bed occupancy rates in hospitals in Indonesia, leading to difficulties in finding available beds for patients, especially those requiring urgent attention. The more than 12,000 new COVID-19–positive cases and more than 175,000 active cases recorded at the end of January 2021 made the government launch an additional health care system to check hospital bed availability, especially intensive care units (ICUs) in all 34 provinces [33]. By July 2021, as Indonesia faced a subsequent wave with a sharp rise in the number of cases and hospitalizations, an additional health care data system was launched for individuals in need of real-time access to information, such as access to oxygen [34] and to receive and donate convalescent plasma [35].

In addition to supporting COVID-19 case identification and laboratory test result integration across Indonesia, the Ministry of Health developed New All Records (NAR) TC-19 for all COVID-19 laboratory networks to input both polymerase chain reaction (PCR) and antigen test results. At the beginning of the pandemic, only the National Institute of Health and Research Development (NIHRD) had the capacity for COVID-19 testing [36]. Through extended collaboration with other ministries, institutions, nongovernmental organizations (NGOs), international donors, and the private sector, the number of COVID-19 laboratory networks that were using NAR increased to 936 across all 34 of Indonesia's provinces as of June 2022 [37]. SILACAK was also developed by the Ministry of Health to strengthen contact-tracing efforts in Indonesia [38]. Table 4 contains the 4 health care data systems, including hospital, laboratory, and contact-tracing data systems.

Table 4. List of health care and laboratory data systems.

Website or dashboard	Functions
SIRANAP[a]	The SIRANAP platform provides beds and ICU[b] availability data. Hospitals provide real-time updates through this platform every 3 hours, recently with additional volunteer support from IndoRelawan. This feature has already merged into the PeduliLindungi app.
Blood Plasma Donor	The system provided access to people who have recovered from COVID-19 and are eligible for convalescent plasma donation.
NAR[c] TC-19	Laboratory test data results are integrated into the system both for PCR[d] and the antigen test.
SILACAK	Data collection tools for health care workers are used to carry out contact tracing in the community.

[a]SIRANAP: Sistem Informasi Rawat Inap (translates as "hospitalization information system").

[b]ICU: intensive care unit.

[c]NAR: New All Records.

[d]PCR: polymerase chain reaction.

COVID-19 Screening and Population Treatment

This last group of findings compiles the use of digital technologies for prevention, risk assessment, and contact tracing. These 3 functions are often found combined within a single platform. The following list details the platforms offering these:

- PeduliLindungi, which means "to care for and to protect," is a smartphone-based app released by the Ministry of Communication and Information Technology of the Republic of Indonesia that the public can use for self-assessment, for example, to know the COVID-19 infection risk within their surroundings using the government's population and contact-tracing databases [39,40]. Additionally, the app is also synchronized with vaccination data, where people can check whether they are eligible for vaccination and register to receive one. The app also provides a list of nearby vaccine centers. People who have been fully vaccinated can also access their vaccine certificates through the app [41].
- Corona Likelihood Metric (CLM) is a COVID-19–screening mobile app that provides an online self-assessment form with the help of machine learning technology formulated by the Government of DKI Jakarta Province, in collaboration with the Harvard CLM Team and Klakklik. CLM can also recommend whether a citizen should take a COVID-19 test [39].
- Fight Covid-19 is a mobile app used by 1 of the local governments, namely Bangka Belitung Province, to trace the mobility of people from COVID-19 epicenter provinces, such as Jakarta to Bangka Belitung. The app is used to track the travel history of arrivals using location data collected from the phone GPS [40].
- Blue Pass is a device for contact tracing within an office setting. This device has been successfully used in Singapore, and a trial took place at the National Disaster Mitigation Body (BNPB), Indonesia. The device uses the GPS to record people who stand within a 3 m distance from another person. When one person tests positive for COVID-19, all recorded people who ever stood within the 3 m radius are notified as a form of contact tracing [42].
- Electronic Health Alert Card (e-HAC) is a mobile app being used to record people's international mobility to Indonesia and people's mobility within Indonesian provinces. All passengers of airplanes, ships, and trains are obliged to fill in their travel record data (destination and origin) to be able to enter the Indonesian border and travel domestically. The app is also connected with official clinics in Indonesia, where people can get COVID-19 tests before traveling; thus, it can record COVID-19 test results for domestic travel [43].
- 10 Rumah Aman is a mobile app developed by the Kantor Staf Presiden (KSP, or President Staff Office) and the Kementerian Komunikasi dan Informatika (Kemenkominfo, or the Ministry of Communication and Information Technology) to educate the community about the COVID-19 pandemic and recommend preventative actions, such as routine temperature checking. Several features are displayed in the app, such as "Check Body Temperature," "Become a Safe Warrior," "Information for Your Health," "Regarding COVID-19," "Check Your Health Here," "COVID-19 WhatsApp," and "Monitor Body Temperature Map."
- Provincial mobile apps (Pusat Informasi dan Koordinasi COVID-19 Jawa Barat [PIKOBAR], Sawarna, Pantau Pandemi Sulawesi Barat [Papa Sulbar]): A special mention should be made for some provinces that also created or adopted COVID-19 features on their mobile apps specifically for their local populations. For example,

PIKOBAR (or the West Java COVID-19 Information and Coordination Center) provides information about COVID-19 case distribution across West Java, information about the schedule and location for COVID-19 vaccinations, information about self-quarantine, COVID-19 logistic requests, and hospital and call center contact numbers. In Bandung City, the local government created Sawarna, a mobile app that helps the local community know the COVID-19 case distribution in Bandung. Another mobile app available is Papa Sulbar (or the Pandemic Monitor at West Sulawesi) that provides information about COVID-19 case distribution in West Sulawesi as well as the latest updates of COVID-19 pandemic control across the province.

- Mobile JKN is a mobile app developed by the Badan Penyelenggara Jaminan Sosial (BPJS, or the Social Health Insurance Administration Body) to facilitate access for BPJS participants based on a prepandemic beta version, which was subsequently further enhanced. Using Mobile JKN, individuals can get information about BPJS, such as checking membership, paying dues, checking health facilities, and requesting reprinting of membership cards. During the COVID-19 pandemic, mobile JKN also adopted a COVID-19 self-screening feature.

Several countries have adopted population screening apps to aid in the control of the pandemic waves as well as to function as a reference and warning point for individual users (eg, if the latter were colocated in time and space with known individuals with COVID-19). Such examples include the Corona Warn app in Germany [44], the CovidSafe app in Australia [45], and Covid Tracker in Ireland [46]. However, the adoption rates of such apps were altogether lower than originally anticipated (eg, the government-endorsed CovidSafe app in Australia was installed by 21.6% of the population, and that is 1 of the highest adoption rates observed), primarily due to concerns about personal data security.

The data visualization tools described in this study still exist and are being used to provide updated information about COVID-19 cases at national and local levels. The Health Protocol Compliance Monitoring System is currently still being used; however, the number of reports has decreased. Telemedicine apps are still being used, and health care data systems, such as SIRANAP, have been integrated into PeduliLindungi. Blood Plasma Donor is no longer active since WHO did not recommend blood plasma convalescent transfers for patients with COVID-19 since December 2021 [47]. Regarding COVID-19 population screening and treatment, (1) PeduliLindungi has been downloaded by more than 90 million of the Indonesian population and will become a citizen's health app adopted by the Ministry of Health [48]; (2) the CLM is still available on the DKI Jakarta local government website, although the usage is low; (3) Fight Covid is still used in Bangka Belitung Province, although the usage has decreased; (4) Blue Pass has been implemented at several places, such as the BNPB, Bintan resorts, and other resorts or tourist attractions. However, Blue Pass is no longer widely used compared to the first launch at the beginning of 2021; and (5) e-HAC is no longer used, although its function has been integrated into PeduliLindungi.

Discussion

Principal Findings

History has shown that major crises can often trigger new ideas and innovations [49]. In this context, the digital technologies that have been developed for COVID-19 in Indonesia and were identified in this review represent a leap forward for Indonesian digital health innovation. The pandemic has afforded the opportunity for the largest number of health technologies ever (almost 50 different technologies in total) to be introduced into the Indonesian health care ecosystem.

As a developing country, Indonesia can learn more from other developed countries in Asia, such as South Korea using the MERS-CoV outbreak in 2015 as its turning point to advance digital health use and innovation within its health care services [23]. Our findings demonstrate an opportunity for these technologies to impact many different areas of the Indonesian health care services, as the digital health technologies introduced cover a wide number of applications (from decision-making support and encompassing health system technologies to robotics to individual patient monitoring and tracing). Several international studies have also highlighted the need for developing countries to accelerate digital innovation, given the gaps in research and innovation, in digital health [49,50] particularly in promoting learning systems that foster ongoing collaborations between government, industry (private sector), academia, and community, sometimes called the "quadruple helix of innovation" [51]. There have been several efforts from global health organizations, such as the WHO Access to COVID-19 Tools (ACT) Accelerator Diagnostics Partnership. This initiative focused on bringing high-quality rapid tests, training professionals, and establishing testing for over 500 million people in low- and middle-income countries [52]. As this review demonstrates, the majority of health care technologies introduced were multisectoral with a wide potential reach, with the exception of 3 local provincial apps and 1 institutional one (BPJS) that were more limited in their offerings. Thus, the pandemic has created a precedent for further multisectoral development of health care innovation with a potential national rollout.

However, this work also identified some of the challenges Indonesia is facing to advance the use and innovation of digital technologies as follows: First, the data privacy challenge emerged alongside the invention and use of many digital technologies. This is not an exclusive issue to Indonesia, as it has previously been reported regionally for South Korea, Singapore, Taiwan, and Hong Kong [23]. South Korea and Taiwan have been using electronic wristbands to prevent people from violating self-isolation by using location-tracking systems. In Singapore, TraceTogether is dedicated as a contact-tracing platform using Bluetooth, while in Taiwan, a similar platform uses a digital fencing system. The usage of these tracking or fencing tools raises questions regarding the protection of privacy. However, the Singaporean government has anticipated debates toward the importance of privacy and data protection [53]. The system will not store data, not even geolocation data, and the users' phone numbers and personal identification data are not exchanged at any point. In Indonesia, at an individual user level, the government has protected the users' privacy by not storing geolocation data in the local app, in addition to not exchanging any user information and disallowing permission options to access users' data [54]. At a population level, a further concern relates to the analysis and publication of the gathered data, since COVID-19 infections are being publicly reported extensively, potentially risking the leak of patients' personal data [23]. In response to this issue, Indonesia is improving its information technology regulation to ensure users' safety and privacy [25]. Government-led apps, such as PeduliLindungi and others have set an example and updated their user agreement and privacy policies in parallel and in line with the governmental initiative to ensure users' trust in using the apps safely [41]. To protect data and privacy related to COVID-19, the Ministry of Health worked together with the National Cyber and Crypto Agency (BSSN), an agency under the Government of Indonesia with relevant expertise. Thus, although the existing data safety and privacy solutions might not be in their final form and might still require further updating in the future, the implemented solutions so far are functional.

Second, several implementation challenges remain prevalent in Indonesia's use of digital technologies. These include (1) users' adoption of mobile apps, (2) digital literacy and disparities of technology use among provinces, (3) data analysis that is often hampered by multisectoral coordination, and (4) the need to invest in human resources to foster innovation in the mid- and longer term. The initial low user adoption for apps triggered the government's issuance of regulation to mandatorily use specific apps, such as in the case of e-HAC [55]. In doing so, the debate enhanced relating to data privacy, data collection, and how much an app provider or a local or national government should know about its citizens.

Regarding the second challenge identified, technological disparity (as in a combined effect of digital literacy and disparity of access to new technologies) was existing prepandemic and perhaps has been consolidated further during the COVID-19 pandemic, with certain population groups benefiting from multiple digital health technologies (eg, in the capital Jakarta), while rural populations have fewer options available to them. This aspect was mentioned in all but 1 of the manuscripts considered in this paper where innovations are mainly promoted by major provinces in Java and the western part of Indonesia, exposing years of inequalities in education and wealth between the islands in the world's largest archipelagic state [56]. Finally, regarding challenges (3) and (4), these have been major persistent issues prepandemic and these have come to the fore during the COVID-19 handling in Indonesia [57], with the country still working to develop frameworks that will foster multisectoral coordination for optimum data sharing and analysis. It is likely to remain high in the policy-making agenda, as the evidence from the pandemic demonstrated the usefulness of data sharing, informing policy makers across different ministerial bodies, at national and subnational levels. Lastly, a need to upskill human resources (eg, develop digital literacy of front-line health care staff) will not only foster the current use of digital technologies but also become an important step in maintaining digital innovation beyond the duration of the pandemic. This outlook aligns with objective 4, point 78 on the WHO Global Strategy on Human Resources for Health Workforce 2030, which is to strengthen human resources for health information systems and build the human capital required to operate them [58].

Strengths and Limitations

The strength of this study is that it used a systematic approach based on PRISMA guidelines to perform an extensive literature search. The study is the first digital health care technology investigation for Indonesia and will likely set a precedent for similar future investigations. Furthermore, we identified and included the gray literature to ensure that as many as possible digital apps were captured.

However, there are also limitations to this study. Despite the intensive literature search, only a limited number of papers discussing digital technology implementation in Indonesia were identified. In particular, robotic technologies were much less mature than mobile-based platforms, and as such, their relative level of readiness (and overall impact) was difficult to estimate. Furthermore, most of the available information was commonly available in Indonesian languages, thus limiting the direct comparability of the information to other apps internationally. Finally, the relative percentage of the installation of these apps over the entire population of Indonesia is only known for few of the apps, and thus we were unable to provide a complete picture regarding the level of usage, frequency of usage, and overall level of reliance.

Conclusion

This review presented the use of digital health technologies for COVID-19 detection and response management in Indonesia. The results were grouped into 3 broad use types for ease of analysis, namely (1) big data, artificial intelligence, and machine learning; (2) health care system technologies; and (3) COVID-19 screening, population treatment, and prevention. The introduction of these digital technologies represents the single-largest introduction of digital technologies within the Indonesian health care ecosystem. Additionally, almost all the technologies were the result of multisectoral coordination among government bodies at a national and subnational scale, along with higher education institutions, research institutions, and the private sector. Thus, the case of Indonesia may provide a blueprint for the introduction of digital health care technologies for several other resource-restricted settings. The introduction of these technologies demonstrated the potential benefit of big data for informing public health policy making during health emergencies, as initiated and led by the National Task Force for COVID-19 Mitigation Acceleration.

This work also acknowledges the many challenges that remain, such as data privacy, disparity of technological access, the need for further multisectoral coordination, and the ability to support such innovation by appropriately skilled staff. Therefore, although it is important to identify the benefits of the implemented digital health technologies, it also remains critical to maintain the multisectoral cooperation frameworks (among government bodies, academia, and the private sector) for the longer term, both for addressing population health needs and for advancing Indonesia's digital health care transformation.

Conflicts of Interest

Where authors are identified as personnel of the International Agency for Research on Cancer or the World Health Organization (WHO), the authors alone are responsible for the views expressed in this paper, and they do not necessarily represent the decisions, policy, or views of the International Agency for Research on Cancer or WHO.

Multimedia Appendix 1
Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist.
[DOCX File , 32 KB - ijmr_v12i1e41308_app1.docx]

References

1. Sitepu F, Depari E, Mudatsir M, Harapan H. Being unvaccinated and contact with measles cases as the risk factors of measles outbreak, North Sumatera, Indonesia. Clin Epidemiol Glob Health 2020 Mar;8(1):239-243 [FREE Full text] [doi: 10.1016/j.cegh.2019.08.006]
2. Kandun IN. Emerging diseases in Indonesia: control and challenges. Trop Med Health 2006;34(4):141-147. [doi: 10.2149/tmh.34.141]
3. Setianingsih TY, Wiyatno A, Hartono TS, Hindawati E, Rosamarlina, Dewantari AK, et al. Detection of multiple viral sequences in the respiratory tract samples of suspected Middle East respiratory syndrome coronavirus patients in Jakarta, Indonesia 2015-2016. Int J Infect Dis 2019 Sep;86:102-107 [FREE Full text] [doi: 10.1016/j.ijid.2019.06.022] [Medline: 31238156]
4. Cumulative number of confirmed human cases for avian influenza A(H5N1) reported to WHO, 2003-2021. World Health Organization. 2021 Apr 15. URL: https://www.who.int/publications/m/item/cumulative-number-of-confirmed-human-cases-for-avian-influenza-a(h5n1)-reported-to-who-2003-2021-15-april-2021 [accessed 2023-01-10]
5. Adisasmito W, Budayanti SN, Aisyah DN, Gallo Cassarino T, Rudge JW, Watson SJ, et al. Phylogenetic characterisation of circulating, clinical influenza isolates from Bali, Indonesia: preliminary report from the BaliMEI project. BMC Infect Dis 2017 Aug 23;17(1):583 [FREE Full text] [doi: 10.1186/s12879-017-2684-2] [Medline: 28830452]
6. Aisyah DN, Mayadewi CA, Diva H, Kozlakidis Z, Siswanto, Adisasmito W. A spatial-temporal description of the SARS-CoV-2 infections in Indonesia during the first six months of outbreak. PLoS One 2020 Dec 22;15(12):e0243703 [FREE Full text] [doi: 10.1371/journal.pone.0243703] [Medline: 33351801]
7. Weekly epidemiological update on COVID-19 - 29 June 2021. World Health Organization. 2021 Jun 29. URL: https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---29-june-2021 [accessed 2021-07-06]
8. WHO coronavirus (COVID-19) dashboard. World Health Organization. URL: https://covid19.who.int/table [accessed 2021-07-06]
9. Kozlakidis Z, Catchpoole D. Healthcare innovation: will COVID-19 be a transformative experience? Innov Digit Health Diagn Biomark 2021;1(2):25-26 [FREE Full text] [doi: 10.36401/IDDB-21-X2]
10. Suhenda D, Janti N. Emergency mobility restrictions now in full effect. The Jakarta Post. 2021. URL: https://www.thejakartapost.com/news/2021/07/04/emergency-mobility-restrictions-now-in-full-effect.html [accessed 2021-07-06]
11. Abbas R, Michael K. COVID-19 contact trace app deployments: learnings from Australia and Singapore. IEEE Consumer Electron Mag 2020 Sep 1;9(5):65-70. [doi: 10.1109/mce.2020.3002490]
12. Kerr CC, Mistry D, Stuart RM, Rosenfeld K, Hart GR, Núñez RC, et al. Controlling COVID-19 via test-trace-quarantine. Nat Commun 2021 May 20;12(1):2993 [FREE Full text] [doi: 10.1038/s41467-021-23276-9] [Medline: 34017008]
13. Araujo KM, Denadai R. Digital clinician-patient consultation for dermatology care in stressful COVID-19 environment. J Patient Exp 2020 Aug 14;7(4):426-427 [FREE Full text] [doi: 10.1177/2374373520942376] [Medline: 33062851]
14. Hanna MG, Reuter VE, Ardon O, Kim D, Sirintrapun SJ, Schüffler PJ, et al. Validation of a digital pathology system including remote review during the COVID-19 pandemic. Mod Pathol 2020 Nov;33(11):2115-2127 [FREE Full text] [doi: 10.1038/s41379-020-0601-5] [Medline: 32572154]
15. Ben-Arye E, Paller CJ, Lopez AM, White S, Pendleton E, Kienle GS, et al. The Society for Integrative Oncology Practice Recommendations for online consultation and treatment during the COVID-19 pandemic. Support Care Cancer 2021 Oct 14;29(10):6155-6165 [FREE Full text] [doi: 10.1007/s00520-021-06205-w] [Medline: 33852088]
16. Whitelaw S, Mamas M, Topol E, Van Spall H. Applications of digital technology in COVID-19 pandemic planning and response. The Lancet Digital Health 2020 Aug;2(8):e435-e440. [doi: 10.1016/S2589-7500(20)30142-4]
17. Algaissi AA, Alharbi NK, Hassanain M, Hashem AM. Preparedness and response to COVID-19 in Saudi Arabia: building on MERS experience. J Infect Public Health 2020 Jun;13(6):834-838 [FREE Full text] [doi: 10.1016/j.jiph.2020.04.016] [Medline: 32451260]

18. Tim Inovasi Kemenristek/BRIN. Tanggap Hadapi COVID-19: Katalog Inovasi Karya Peneliti dan Perekayasa Konsorsium Riset dan Inovasi COVID-19 untuk Mengatasi Pandemi. Ministry of Research and Technology of the Republic of Indonesia. 2020. URL: https://persi.or.id/wp-content/uploads/2020/06/katalog_seri18.pdf [accessed 2023-02-07]
19. Biro Pers, Media, dan Informasi Sekretariat Presiden. Presiden: Lakukan Pelacakan Lebih Agresif dengan Gunakan Teknologi Telekomunikasi. Berita Terkini. 2020 Jun 04. URL: https://covid19.go.id/p/berita/presiden-lakukan-pelacakan-lebih-agresif-dengan-gunakan-teknologi-telekomunikasi [accessed 2021-07-06]
20. Rokom. Kemenkes Fasilitasi Konsultasi, Obat Gratis bagi Pasien COVID-19 di Jakarta via Fasilitas Telemedicine. Sehat Negeriku. 2021 Jul 6. URL: https://sehatnegeriku.kemkes.go.id/baca/berita-utama/20210705/1338034/kemenkes-fasilitasi-konsultasi-obat-gratis-bagi-pasien-covid-19-di-jakarta-via-fasilitas-telemedicine/ [accessed 2023-01-10]
21. Aisyah DN, Ahmad RA, Artama WT, Adisasmito W, Diva H, Hayward AC, et al. Knowledge, attitudes, and behaviors on utilizing mobile health technology for TB in Indonesia: a qualitative pilot study. Front Public Health 2020 Oct 6;8:531514 [FREE Full text] [doi: 10.3389/fpubh.2020.531514] [Medline: 33123508]
22. Telemedicine Platforms Provide Free Services for COVID-19 Self-Isolation Patients. Cabinet Secretariat of the Republic of Indonesia Gov't. URL: https://setkab.go.id/en/govt-announces-procedures-for-telemedicine-services-for-covid-19-self-isolation-patients/ [accessed 2021-10-22]
23. Adisasmito W, Suwandono A, Trihono GA, Aisyah DN, Solikha D. Studi Komparasi Pembelajaran Penangan COVID-19 Indonesia-Korea Selatan. Ministry of National Development Planning, Republic of Indonesia. 2021 Apr 26. URL: https://covid19.go.id/p/hasil-kajian/studi-komparasi-pembelajaran-penanganan-covid-19-indonesia-korea-selatan [accessed 2021-07-12]
24. Regulasi di Indonesia. URL: https://covid19.go.id/ [accessed 2023-01-10]
25. Yang G, Lv H, Zhang Z, Yang L, Deng J, You S, et al. Keep healthcare workers safe: application of teleoperated robot in isolation ward for COVID-19 prevention and control. Chin J Mech Eng 2020 Jun 09;33(1):47. [doi: 10.1186/s10033-020-00464-0]
26. Telemedicine: opportunities and developments in Member States: report on the second global survey on eHealth. World Health Organization. 2010. URL: https://apps.who.int/iris/handle/10665/44497 [accessed 2021-07-06]
27. Portnoy J, Waller M, Elliott T. Telemedicine in the era of COVID-19. J Allergy Clin Immunol Pract 2020 May;8(5):1489-1491 [FREE Full text] [doi: 10.1016/j.jaip.2020.03.008] [Medline: 32220575]
28. Hollander JE, Carr BG. Virtually perfect? Telemedicine for Covid-19. N Engl J Med 2020 Apr 30;382(18):1679-1681. [doi: 10.1056/nejmp2003539]
29. Ohannessian R, Duong TA, Odone A. Global telemedicine implementation and integration within health systems to fight the COVID-19 pandemic: a call to action. JMIR Public Health Surveill 2020 Apr 02;6(2):e18810 [FREE Full text] [doi: 10.2196/18810] [Medline: 32238336]
30. Lima Tahun Berinovasi, Halodoc Terus Fokus Jawab Tantangan Kesehatan di Indonesia. Halodoc. 2021 Apr 23. URL: https://www.halodoc.com/media/lima-tahun-berinovasi-halodoc-terus-fokus-jawab-tantangan-kesehatan-di-indonesia [accessed 2022-09-13]
31. About us. Alodokter. 2020. URL: https://www.alodokter.com/about [accessed 2022-09-13]
32. Annur CM. Layanan Telemedicine yang Paling Banyak Digunakan di Indonesia, Apa Saja? Databoks. 2022 Apr 07. URL: https://databoks.katadata.co.id/datapublish/2022/04/07/layanan-telemedicine-yang-paling-banyak-digunakan-di-indonesia-apa-saja [accessed 2022-09-13]
33. Adlin E. Cara Cek Ketersediaan Tempat Tidur di Rumah Sakit via Online, Coba Unduh SIRANAP. Ayo Jakarta. 2021 Jan 26. URL: https://www.ayojakarta.com/read/2021/01/26/30149/cara-cek-ketersediaan-tempat-tidur-di-rumah-sakit-via-online-coba-unduh-siranap [accessed 2021-07-12]
34. Perjuangan Pasien Covid-19 Dapatkan Tabung Oksigen. Newsline Metro TV. 2021 Jul 12. URL: https://www.metrotvnews.com/play/kELC8aRD-perjuangan-pasien-covid-19-dapatkan-tabung-oksigen [accessed 2021-07-15]
35. Hapsari MA. Sulitnya Mencari Donor Plasma Konvalesen, Banyak Penyintas Takut ke RS. Kompas.com. 2021 Jun 24. URL: https://megapolitan.kompas.com/read/2021/06/24/06040011/sulitnya-mencari-donor-plasma-konvalesen-banyak-penyintas-takut-ke-rs [accessed 2021-07-15]
36. Aisyah DN, Mayadewi CA, Budiharsana M, Solikha DA, Ali PB, Igusti G, et al. Building on health security capacities in Indonesia: Lessons learned from the COVID-19 pandemic responses and challenges. Zoonoses Public Health 2022 Sep 26;69(6):757-767 [FREE Full text] [doi: 10.1111/zph.12976] [Medline: 35618675]
37. Daftar Laboratorium Jejaring Pemeriksa COVID-19 yang Input ke Dalam NAR Selama 3 Minggu Terakhir. Badan Litbangkes. 2022 Jun 20. URL: https://www.litbang.kemkes.go.id/laboratorium-pemeriksa-covid-19/ [accessed 2022-11-20]
38. Wisnuwardani DP. Bantu Daerah Deteksi COVID-19, Satgas dan Kemenkes RI Luncurkan Program Penguatan Tracing di 10 Provinsi. Liputan 6. 2020 Nov 03. URL: https://www.liputan6.com/health/read/4398377/bantu-daerah-deteksi-covid-19-satgas-dan-kemenkes-ri-luncurkan-program-penguatan-tracing-di-10-provinsi [accessed 2022-11-20]
39. Hanggara A. Kenapa Kamu Harus Menggunakan Corona Likelihood Metric. Jakarta Smart City. 2020. URL: https://smartcity.jakarta.go.id/blog/514/kenapa-kamu-harus-menggunakan-corona-likelihood-metric [accessed 2021-07-12]

40. Wibowo A. Setiap Pendatang di Babel Dilacak Pergerakannya Pakai Aplikasi Fight COVID-19. Berita Terkini. 2020 Apr 24. URL: https://covid19.go.id/p/berita/setiap-pendatang-di-babel-dilacak-pergerakannya-pakai-aplikasi-fight-covid-19 [accessed 2021-07-12]
41. Amanda S. Daftar Fitur Aplikasi PeduliLindungi untuk Atasi Pandemi COVID-19. Tirto.id. 2021 Feb 03. URL: https://tirto.id/daftar-fitur-aplikasi-pedulilindungi-untuk-atasi-pandemi-covid-19-f9U6 [accessed 2021-07-12]
42. Jati R. BNPB Uji coba Blue Pass pada Pegawai untuk Tracing Covid-19. National Disaster Mitigation Body, Republic of Indonesia. 2021 Jan 12. URL: https://bnpb.go.id/berita/BNPB%20Uji%20coba%20Blue%20Pass%20pada%20Pegawai%20untuk%20Tracing%20Covid-19 [accessed 2021-07-12]
43. Panduan Penggunaan Aplikasi E-HAC. Ministry of Health, Republic of Indonesia. URL: https://infeksiemerging.kemkes.go.id/download/Panduan_Pengguna_Aplikasi_E-HAC.pdf [accessed 2021-07-12]
44. Horstmann KT, Buecker S, Krasko J, Kritzler S, Terwiel S. Who does or does not use the 'Corona-Warn-App' and why? Eur J Public Health 2021 Feb 01;31(1):49-51 [FREE Full text] [doi: 10.1093/eurpub/ckaa239] [Medline: 33340328]
45. Abbas H, Xu X, Sun C. Role of COVIDsafe app and control measures in Australia in combating COVID-19 pandemic. TG 2021 Jul 19;15(4):708-719. [doi: 10.1108/TG-01-2021-0004]
46. Banks J, Doherty CP. CoVID Tracker Ireland: What is its Function? Ir Med J 2022 Jan 20;115(1):513. [Medline: 35279047]
47. WHO recommends against the use of convalescent plasma to treat COVID-19. World Health Organization. 2021 Dec 07. URL: https://www.who.int/news/item/07-12-2021-who-recommends-against-the-use-of-convalescent-plasma-to-treat-covid-19 [accessed 2022-09-16]
48. Banks J, Doherty CP. Blueprint for Digital Health Transformation Strategy 2024. Ministry of Health of the Republic of Indonesia. URL: https://dto.kemkes.go.id/ENG-Blueprint-for-Digital-Health-Transformation-Strategy-Indonesia%202024.pdf [accessed 2023-01-10]
49. Iyawa GE, Herselman M, Botha A. A scoping review of digital health innovation ecosystems in developed and developing countries. 2017 Presented at: InIST-Africa Week Conference (IST-Africa); May 31-June 2, 2017; Windhoek, Namibia p. 1-10. [doi: 10.23919/istafrica.2017.8102325]
50. Nielsen P. Digital innovation: a research agenda for information systems research in developing countries. In: Choudrie J, Islam MS, Wahid F, Bass JM, Priyatma JE, editors. Information and Communication Technologies for Development. Cham: Springer International; 2017.
51. Roman M, Varga H, Cvijanovic V, Reid A. Quadruple Helix Models for Sustainable Regional Innovation: Engaging and Facilitating Civil Society Participation. Economies 2020 Jun 08;8(2):48. [doi: 10.3390/economies8020048]
52. Benda A, Zerajic L, Ankita A, Cleary E, Park Y, Pandey S. COVID-19 Testing and Diagnostics: A Review of Commercialized Technologies for Cost, Convenience and Quality of Tests. Sensors (Basel) 2021 Oct 01;21(19) [FREE Full text] [doi: 10.3390/s21196581] [Medline: 34640901]
53. Goggin G. COVID-19 apps in Singapore and Australia: reimagining healthy nations with digital technology. Media International Australia 2020 Aug 14;177(1):61-75 [FREE Full text] [doi: 10.1177/1329878X20949770]
54. Kebijakan Kerahasiaan: PeduliLindungi berkomitmen terhadap privasi Anda. Ministry of Communication and Informatics of the Republic of Indonesia. 2022 Aug 08. URL: https://www.pedulilindungi.id/kebijakan-privasi-data#1-cara-kerja-pedulilindungi [accessed 2021-10-23]
55. Pakai Aplikasi EHAC Untuk Validasi Surat Hasil Tes COVID-19, Proses Keberangkatan Cepat Dan Lancar. PT Angkasa Pura Indonesia. 2020 Jan 08. URL: https://www.angkasapura2.co.id/id/news/event/info/453-pakai-aplikasi-ehac-untuk-validasi-surat-hasil-tes-covid-19-proses-keberangkatan-cepat-dan-lancar [accessed 2023-02-11]
56. Booth A. Africa in Asia? The development challenges facing Eastern Indonesia and East Timor. Oxford Development Studies 2004 Mar 01;32(1):19-35. [doi: 10.1080/1360081042000184101]
57. Aisyah DN, Mayadewi CA, Igusti G, Manikam L, Adisasmito W, Kozlakidis Z. Laboratory Readiness and Response for SARS-Cov-2 in Indonesia. Front Public Health 2021;9:705031 [FREE Full text] [doi: 10.3389/fpubh.2021.705031] [Medline: 34350153]
58. Global strategy on Human Resources for Health: Workforce 2030. World Health Organization. 2016. URL: https://apps.who.int/iris/bitstream/handle/10665/250368/9789241511131-eng.pdf;sequence=1 [accessed 2021-10-25]

Abbreviations

BLC: Bersatu Lawan COVID-19
BNPB: National Disaster Mitigation Body
BPJS: Badan Penyelenggara Jaminan Sosial
CLM: Corona Likelihood Metric
e-HAC: Electronic Health Alert Card
ICU: intensive care unit
NAR: New All Records
Papa Sulbar: Pantau Pandemi Sulawesi Barat
PCR: polymerase chain reaction

PIKOBAR: Pusat Informasi dan Koordinasi COVID-19 Jawa Barat
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SIRANAP: Sistem Informasi Rawat Inap
WHO: World Health Organization

Edited by A Mavragani; submitted 22.07.22; peer-reviewed by S Pandey, KS Sahu; comments to author 12.08.22; revised version received 02.12.22; accepted 09.01.23; published 28.02.23.

Please cite as:
Nur Aisyah D, Lokopessy AF, Naman M, Diva H, Manikam L, Adisasmito W, Kozlakidis Z
The Use of Digital Technology for COVID-19 Detection and Response Management in Indonesia: Mixed Methods Study
Interact J Med Res 2023;12:e41308
URL: https://www.i-jmr.org/2023/1/e41308
doi:10.2196/41308
PMID:36623206

Original Paper

Impact of Social Isolation, Physician-Patient Communication, and Self-perception on the Mental Health of Patients With Cancer and Cancer Survivors: National Survey Analysis

Avishek Choudhury[1], PhD

Industrial and Management Systems Engineering, Benjamin M Statler College of Engineering and Mineral Resources, West Virginia University, Morgantown, WV, United States

Corresponding Author:
Avishek Choudhury, PhD
Industrial and Management Systems Engineering
Benjamin M Statler College of Engineering and Mineral Resources
West Virginia University
1374 Evansdale Drive
321 Engineering Sciences Building
Morgantown, WV, 26506
United States
Phone: 1 304 293 9431
Email: avishek.choudhury@mail.wvu.edu

Abstract

Background: Cancer is perceived as a life-threatening, fear-inducing, and stigmatized disease. Most patients with cancer and cancer survivors commonly experience social isolation, negative self-perception, and psychological distress. The heavy toll that cancer takes on patients continues even after treatment. It is common for many patients with cancer to feel uncertain about their future. Some undergo anxiety, loneliness, and fear of getting cancer again.

Objective: This study examined the impact of social isolation, self-perception, and physician-patient communication on the mental health of patients with cancer and cancer survivors. The study also explored the impact of social isolation and physician-patient communication on self-perception.

Methods: This retrospective study used restricted data from the 2021 Health Information National Trends Survey (HINTS), which collected data from January 11, 2021, to August 20, 2021. We used the partial least squares structural equation modeling (PLS-SEM) method for data analysis. We checked for quadratic effects among all the paths connecting social isolation, poor physician-patient communication, mental health (measured using the 4-item Patient Health Questionnaire [PHQ-4]), and negative self-perception. The model was controlled for confounding factors such as respondents' annual income, education level, and age. Bias-corrected and accelerated (BCA) bootstrap methods were used to estimate nonparametric CIs. Statistical significance was tested at 95% CI (2-tailed). We also conducted a multigroup analysis in which we created 2 groups. Group A consisted of newly diagnosed patients with cancer who were undergoing cancer treatment during the survey or had received cancer treatment within the last 12 months (receipt of cancer treatment during the COVID-19 pandemic). Group B consisted of respondents who had received cancer treatment between 5 and 10 years previously (receipt of cancer treatment before the COVID-19 pandemic).

Results: The analysis indicated that social isolation had a quadratic effect on mental health, with higher levels of social isolation associated with worse mental health outcomes up to a certain point. Self-perception positively affected mental health, with higher self-perception associated with better mental health outcomes. In addition, physician-patient communication significantly indirectly affected mental health via self-perception.

Conclusions: The findings of this study provide important insights into the factors that affect the mental health of patients with cancer. Our results suggest that social isolation, negative self-perception, and communication with care providers are significantly related to mental health in patients with cancer.

(Interact J Med Res 2023;12:e45382) doi:10.2196/45382

KEYWORDS

cancer communication; cancer stigma; mental health; social isolation; cancer survivorship; patient-centeredness

Introduction

Background

Cancer is perceived as a life-threatening, fear-inducing, and stigmatized disease [1-3]. Cancer diagnosis and treatment require a longitudinal and systematic approach involving a multidisciplinary care team, including pathologists, radiologists, oncologists, nurses, and social workers. The members of a care team often perform tasks at broadly two levels: (1) clinical activities and (2) nonclinical activities. Most of the efforts and resources of the care team are invested in augmenting the clinical activities that directly improve cancer detection. The nonclinical tasks involve verbal and nonverbal communication with the patients and other team members, which needs further development.

Textbox 1 presents a simplified version of the overall cancer care process (from diagnosis to treatment to aftercare) from a patient's perspective. It should be noted that the simplified version provides a broad understanding of the journey of a typical patient with cancer. The process might differ across different health care establishments. Once clinically diagnosed with cancer, a patient is likely to undergo complex emotional experiences and face challenges in handling the bad news, selecting treatment options, dealing with the social isolation and stigma, and, most importantly, performing all the patient tasks (eg, comprehending diagnosis, traveling, following treatment protocols, communicating with the care team, handling self-care activities, managing finances, seeking help from family, and making arrangements to support dependents) throughout the treatment process. Some even believe that the cancer treatment is worse than the ailment [2]. The heavy toll that cancer takes on patients continues even after treatment. It is common for many patients with cancer to feel uncertain about their future. Some undergo anxiety, loneliness, and fear of getting cancer again. They even experience fatigue, difficulty sleeping, persistent pain from neuropathy, and emotional distress. The struggles worsen with age [4] and low health literacy [5].

Textbox 1. A simplified version of the cancer care pathway in sequential order from a patient's perspective.

- Cancer suspicion and diagnosis: at this stage, the patient feels discomfort and visits their care provider for a medical checkup. The care provider prescribes cancer diagnostic tests.
- Receiving the cancer *bad news*: this is the moment when the patient, after waiting days or weeks to get the diagnosis results, visits the clinic to receive the cancer *bad news*.
- Comprehending the diagnosis, care options, and next steps: on the same day, after receiving the bad news, the patient must understand their diagnosis, cancer severity, treatment options, and next steps. It should be noted that the patient is undergoing negative emotions from the diagnosis.
- Communicating with the oncologist: on the same day or shortly afterward, the patient, along with the family (if any), must speak with the oncologist to discuss the treatment in detail.
- Communicating with the care team, including social worker: depending on the patient's needs, they must speak with other care team members about the support they might need during and after treatment.
- Scheduling appointments for the treatment: at this stage, based on availability, treatment appointments are scheduled.
- Receiving the treatment: this is the period (several months) during which the patient must travel, receive the cancer treatment, and adhere to any clinical recommendation. This is when patients gradually become isolated from society and their usual day-to-day activities.
- Recovering from the treatment: this is when the patient recovers physically and mentally from the often painful treatment process.
- Trying to get back to normal life: this is when the patient voluntarily engages with other cancer survivors or patients with cancer on dedicated digital venues to share their journey. Although still scared of getting cancer again, they try to gradually return to their normal life.

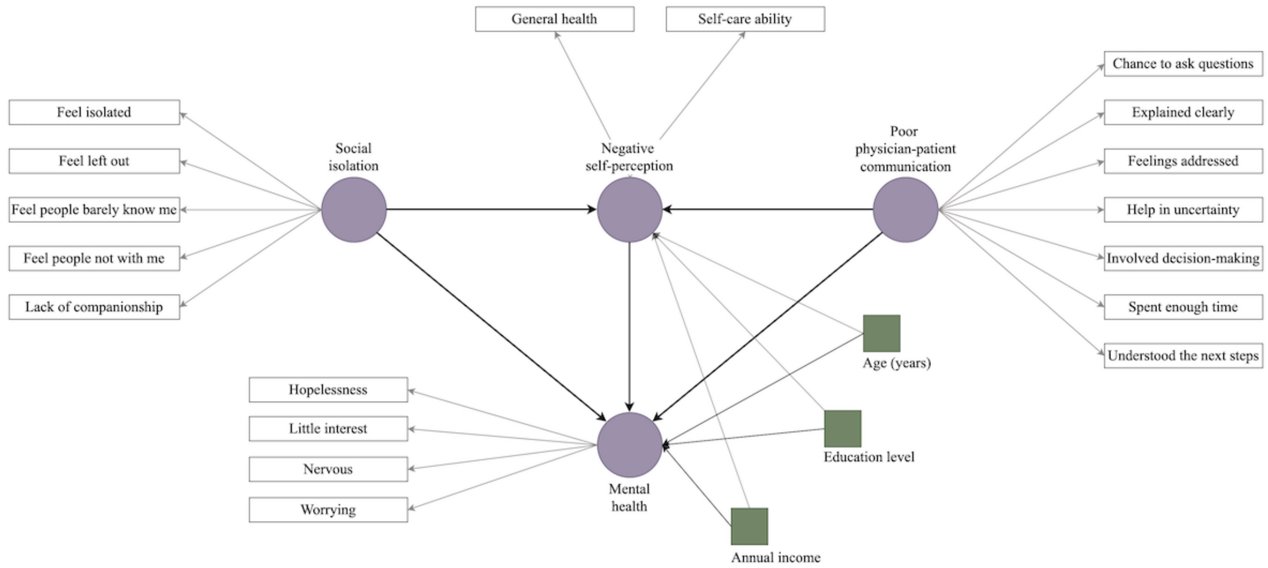
Study Hypotheses

Most patients with cancer and cancer survivors commonly experience social isolation, negative self-perception, and psychological distress [6-9]. Prior studies have explored how cancer induces these challenges [10-13]; for example, studies have been conducted to capture the negative impact of pain and exhaustion associated with cancer treatments (chemotherapy and radiation) on a patient's mental health [14-16]. However, there is a lack of evidence capturing the impact of social isolation and self-perception on psychological distress in patients with cancer. Therefore, we explore the association between social isolation and mental health, hypothesizing that increased social isolation will hinder the mental health of patients with cancer (hypothesis 1).

The potential impact of social isolation among patients with cancer and cancer survivors can extend beyond mental health concerns to the point where it can distort their self-perception [17-19]. When patients with cancer experience social isolation for an extended time for any given reason, be it a disrupted lifestyle or limited physical capability, they start developing negative perceptions about themselves, particularly negative perceptions of their general health and self-care ability. Besides, cancer treatment can often lead to physical changes, such as hair loss, weight changes, or scarring. These changes can be difficult for patients to adjust to and can exacerbate negative self-perceptions. The extent to which social isolation contributes to this negative self-perception is not yet studied. To address this gap, we explore the association between social isolation and negative self-perception of patients with cancer and cancer survivors, hypothesizing that increasing social isolation will encourage (increase) negative self-perception (hypothesis 2).

Effective physician-patient communication is essential in cancer care for several reasons. It ensures patient satisfaction, facilitates shared decision-making, and helps patients with cancer to understand their diagnosis, treatment options, and prognosis. Overall, effective communication can relieve patients from some of the mental and emotional burdens of the care process and help to make the process more patient centered. By contrast, poor communication may lead to confusion and misunderstandings, contributing to patient anxiety and uncertainty. Patients may feel that their concerns are not heard or addressed, leading to frustration and mistrust. However, there is a lack of evidence confirming the potential impact of communication on the mental health of patients with cancer. Therefore, this study explores the association between patients' poor communication with the care provider and their mental health, hypothesizing that poor physician-patient communication negatively affects their mental health (hypothesis 3). We also explore the impact of poor communication on patients' negative self-perception and hypothesize that poor physician-patient communication will increase the negative self-perception of patients (hypothesis 4). Figure 1 illustrates the interactions explored in this study. Our study will contribute to the existing body of knowledge by providing a more in-depth understanding of the role of social isolation, self-perception, and physician-patient communication in the mental health of patients with cancer.

Figure 1. Conceptual framework illustrating the impact of social isolation and poor physician-patient communication on mental health (measured using the 4-item Patient Health Questionnaire for Anxiety and Depression [PHQ-4]) and negative self-perception of patients with cancer.



Methods

Ethics Approval and Data Source

The study was approved by the institutional review board of West Virginia University, Morgantown, West Virginia, United States (2212691613). The 2021 Health Information National Trends Survey (HINTS) deidentified data were obtained and analyzed after approval from the National Cancer Institute (NCI). HINTS is a nationally representative survey of adults in the United States that aims to assess attitudes, behaviors, and knowledge related to cancer and cancer prevention [20].

Survey Instruments and Latent Constructs

In a 2021 pilot program, the NCI administered this survey to oversample cancer survivors using 3 cancer registries from the Surveillance, Epidemiology, and End Results (SEER) program. The pilot program, called HINTS-SEER, was designed to provide a larger sample of cancer survivors for HINTS analyses. HINTS-SEER data were collected from January 11, 2021, to August 20, 2021 [21]. According to HINTS data, 1234 respondents completed the survey. The HINTS service considers a questionnaire to be complete if at least 80% of the questions in each section of the survey are answered [21]. For our research, we handled missing data using the pairwise deletion method.

We used 18 observed variables from the HINTS-SEER survey to feed the proposed conceptual model. Questions from these 18 items were grouped to form 4 latent reflective constructs (Table 1). All 4 latent constructs' convergent, reliability, and discriminant validity were validated. The respondents' *mental health* was measured using the 4-item Patient Health Questionnaire (PHQ-4), which is a brief self-report measure of depression and anxiety used to assess the presence and severity of these conditions [22]. The PHQ-4 consists of 2 questions, 2 of which are designed to assess symptoms of depression (2-item Patient Health Questionnaire [PHQ-2]) and 2 of which are designed to assess symptoms of anxiety (2-item Generalized Anxiety Disorder scale [GAD-2]). Each question asks the respondent to rate the frequency of specific symptoms over the past 2 weeks using a 4-point Likert scale ranging from *not at all* to *nearly every day* [22,23].

The construct of *social isolation* was measured using questions from the Patient-Reported Outcomes Measurement Information System (PROMIS) social isolation instrument [24]. This is a self-report measure used to assess the extent to which an individual experiences feelings of loneliness and social disconnection [24]. The PROMIS social isolation instrument consists of 4 items designed to assess both the quantity and quality of an individual's social interactions and the perceived support they receive from their social network. Each item on the PROMIS social isolation instrument asks the respondent to rate the frequency of specific feelings and experiences related to social isolation using a 5-point Likert scale ranging from *never* to *always* [24]. We added another question (*How often do you feel you lack companionship?*) to improve its convergence.

Seven other questions were combined to determine the quality of physician-patient communication (PPC). The PPC scale is a tool used to assess the quality of communication between patients and health care providers [25]. The scale consists of 7 questions that ask the respondent to rate the extent to which they agree or disagree with statements about their communication with their health care provider. Each question is rated on a 4-point Likert scale ranging from *always* to *never*.

Similarly, questions regarding patients' self-health perception and ability to manage self-care were combined to measure their self-perception. In the context of this study, self-perception is a multidimensional construct that refers to an individual's beliefs, attitudes, and evaluations of themselves and their abilities. In the context of health, self-perception may include an individual's beliefs and attitudes about their health, their ability to manage their health and well-being, and their perceived control over their health outcomes. General health perception, which refers to an individual's overall perceptions of their health, is an essential aspect of self-perception in health. Research has shown that an individual's general health perception is related to a range of health behaviors, including adherence to treatment regimens, engagement in health-promoting behaviors, and use of health care services. Perception of self-care, which refers to an individual's beliefs and attitudes about their ability to take care of themselves, is also an important aspect of self-perception in the context of health. Individuals who perceive themselves as able to manage their health and well-being effectively may be more likely to engage in health-promoting behaviors and seek appropriate health care services when needed. Combining measures of general health perception and perception of self-care can provide a more comprehensive understanding of an individual's self-perception in the context of their health.

Table 1. Reliability and validity of the latent constructs.

Construct	Cronbach α (>.70[a])	Composite reliability (>0.70[a])		AVE[b] (>0.50[c])
		ρ_a	ρ_c	
Mental health (PHQ-4[d])	.88	0.89	0.88	0.65
Social isolation	.89	0.90	0.89	0.63
Poor physician-patient communication	.92	0.92	0.92	0.61
Negative self-perception	.72	0.72	0.72	0.57

[a]Adequate fit.

[b]AVE: average variance extracted.

[c]Acceptable fit.

[d]PHQ-4: 4-item Patient Health Questionnaire.

Structural Equation Modeling

We used partial least squares structural equation modeling (PLS-SEM) to explore the proposed conceptual framework [26]. This method allows the simultaneous estimation of multiple and interrelated dependent relationships between variables and latent constructs [26]. We used the bootstrapping method with 5000 subsamples and controlled for possible confounding factors such as respondents' annual income, education level, and age. Bias-corrected and accelerated (BCA) bootstrap methods were used to estimate nonparametric CIs. Statistical significance was tested at 95% CI (2-tailed).

During the recent COVID-19 pandemic, the emotional distress and the number of deaths were significant factors responsible for increased mental health problems and social isolation across the globe. Many were scared and uncertain about the impact of SARS-CoV-2 on them or their families. This emotional distress was even more ingrained among patients with cancer. Many cancer treatments and consultations were delayed because of the unmanageable workload in the health care industry. The government-imposed lockdowns worldwide also contributed to the social isolation of many individuals, and patients with cancer were not an exception. Given these circumstances, that is, increased mental health problems, social isolation, and overwhelmed health care industry, it is acceptable to assume that the association among social isolation, mental health, physician-patient communication, and self-perception in patients with cancer would be significantly different during the pandemic than during other times. Therefore, we conducted a multigroup analysis (MGA) to test this assumption.

In structural equation modeling (SEM), MGA is a statistical technique used to compare a structural model's fit across different groups or subpopulations [27]. The MGA allows researchers to test whether the same model fits equally well across other groups or whether there are significant differences in the relationships among variables between 2 groups [27]. In this MGA, we created 2 groups. Group A consisted of newly diagnosed patients with cancer who were undergoing cancer treatment during the survey or had received cancer treatment within the last 12 months, that is, receipt of cancer treatment during the COVID-19 pandemic. Group B consisted of respondents who had received cancer treatment between 5 and 10 years previously, that is, receipt of cancer treatment before the COVID-19 pandemic.

Finally, as an additional analysis, we tested for possible curvilinear effects. We checked for quadratic effects (QEs) [28] among all the paths connecting *social isolation*, *poor physician-patient communication*, *mental health* (measured using PHQ-4), and *negative self-perception*.

Results

Overview

Table 2 presents the statistics regarding the sociodemographic variables of the participants. Questions from these 18 items were grouped to form four latent reflective constructs as shown in Table 1: (1) social isolation, (2) negative self-perception, (3) poor physician-patient communication, and (4) mental health. Confirmatory factor analysis was performed to analyze their psychometric properties. All factor loadings were noted to be >0.50. The model fit was evaluated on the standardized root mean square residual (SRMR), an absolute measure of fit that is indicative of the standardized difference between the observed correlation and the predicted correlation. SRMR <0.080 is considered a good fit (observed=0.046). The constructs' reliability and validity were determined using Cronbach α, composite reliability (ρ_a and ρ_c), and the average variance extracted (AVE). The discriminant validity was measured using the heterotrait-monotrait (HTMT) ratio. All HTMT ratios were <0.85, indicating reliable discriminant validity. In addition, we checked for multicollinearity using the variance inflation factor (VIF) and did not find any evidence of multicollinearity. All VIF values were substantially <2.5, ranging between 1.03 and 1.7.

Table 2. Participant characteristics and demographics.

	Received cancer treatment, n (%)				
	Newly diagnosed	Cancer survivors			
	During survey, n (%)	<1 year previously, n (%)	Between 1 and 5 years previously, n (%)	Between 5 and 10 years previously, n (%)	>10 years previously, n (%)
Sex					
Male (n=495)	62 (12.5)	30 (6.1)	119 (24)	109 (22)	174 (35.2)
Female (n=580)	67 (11.6)	19 (3.3)	115 (19.8)	131 (22.6)	248 (42.8)
Age group (years)					
18 to 34 (n=8)	0 (0)	0 (0)	2 (25)	3 (37.5)	3 (37.5)
35 to 49 (n=31)	3 (9.7)	1 (3.2)	10 (32.3)	9 (29)	8 (25.8)
50 to 64 (n=224)	26 (11.6)	5 (2.2)	55 (24.6)	60 (26.8)	78 (34.8)
65 to 75 (n=373)	40 (10.7)	22 (5.9)	91 (24.4)	78 (20.9)	142 (38.1)
≥75 (n=432)	58 (13.4)	21 (4.9)	75 (17.4)	88 (20.4)	190 (44)
Education level					
Less than high school (n=30)	3 (10)	2 (6.7)	6 (20)	8 (26.7)	11 (36.7)
High school graduate (n=130)	22 (16.9)	7 (5.4)	25 (19.2)	34 (26.2)	42 (32.3)
College (n=288)	37 (12.8)	13 (4.5)	66 (22.9)	72 (25)	100 (34.7)
Bachelor’s degree (n=297)	35 (11.8)	10 (3.4)	67 (22.6)	53 (17.8)	132 (44.4)
Postbaccalaureate degree (n=327)	32 (9.8)	17 (5.2)	70 (21.4)	73 (22.3)	135 (41.3)
Employment status					
Employed full time (n=204)	14 (6.9)	6 (2.9)	53 (26)	66 (32.4)	65 (31.9)
Employed part time (n=60)	5 (8.3)	2 (3.3)	15 (25)	11 (18.3)	27 (45)
Homemaker (n=40)	6 (15)	3 (7.5)	4 (10)	12 (30)	15 (37.5)
Student (n=2)	0 (0)	0 (0)	1 (50)	1 (50)	0 (0)
Retired (n=690)	89 (13)	32 (4.6)	139 (20.1)	131 (19)	299 (43.3)
Disabled (n=38)	10 (26.3)	4 (10.5)	9 (23.7)	10 (26.3)	5 (13.2)
Unemployed <1 year (n=10)	1 (10)	0 (0)	4 (40)	2 (20)	3 (30)
Unemployed >1 year (n=10)	1 (10)	0 (0)	3 (30)	3 (30)	3 (30)
Other (n=10)	2 (20)	1 (10)	4 (40)	1 (10)	2 (20)
Race					
Non-Hispanic White (n=815)	99 (12.1)	37 (4.5)	180 (22.1)	185 (22.7)	314 (38.5)
Non-Hispanic Black (n=14)	2 (14.3)	0 (0)	2 (14.3)	2 (14.3)	8 (57.1)
Hispanic (n=115)	13 (11.3)	4 (3.5)	24 (20.9)	24 (20.9)	50 (43.5)
Non-Hispanic Asian (n=61)	5 (8.2)	3 (4.9)	15 (24.6)	13 (21.3)	25 (41)
Non-Hispanic other, n=10	3 (30)	1 (10)	1 (10)	3 (30)	2 (20)
Annual household income (US $)					
<9999 (n=21)	3 (14.3)	3 (14.3)	2 (9.5)	1 (4.8)	12 (57.1)
10,000 to 14,999 (n=38)	5 (13.2)	3 (7.9)	7 (18.4)	9 (23.7)	14 (36.8)
15,000 to 19,999 (n=28)	5 (17.9)	1 (3.6)	6 (21.4)	9 (32.1)	7 (25)
20,000 to 34,999 (n=123)	25 (20.3)	3 (2.4)	18 (14.6)	31 (25.2)	46 (37.4)
35,000 to 49,999 (n=120)	9 (7.5)	8 (6.7)	27 (22.5)	26 (21.7)	50 (41.7)
50,000 to 74,999 (n=177)	25 (14.1)	10 (5.6)	55 (31.1)	34 (19.2)	53 (29.9)
75,000 to 99,000 (n=178)	19 (10.7)	5 (2.8)	35 (19.7)	40 (22.5)	79 (44.4)

	Received cancer treatment, n (%)				
	Newly diagnosed	Cancer survivors			
	During survey, n (%)	<1 year previously, n (%)	Between 1 and 5 years previously, n (%)	Between 5 and 10 years previously, n (%)	>10 years previously, n (%)
100,000 to 199,000 (n=272)	32 (11.8)	12 (4.4)	57 (21)	57 (21)	114 (41.9)
≥200,000 (n=130)	9 (6.9)	5 (3.8)	29 (22.3)	35 (26.9)	52 (40)

Failure to Reject Hypotheses

Table 3 shows significant direct, indirect, and total effects of social isolation on mental health, failing to reject hypothesis 1. In other words, increasing social isolation will negatively affect mental health. The finding of a significant indirect effect of social isolation on mental health via negative self-perception, as indicated by a negative coefficient (−0.08) and *P*<.001, suggests that social isolation may influence mental health in patients with cancer through its effect on negative self-perception. The negative coefficient for the indirect effect indicates that higher levels of social isolation are associated with reduced mental health outcomes through their impact on negative self-perception. This finding suggests that social isolation may have a risk effect or harmful effect on mental health in patients with cancer, potentially by increasing negative self-perception.

In addition, the QE was also significant between these 2 constructs. The significant QE indicates a statistically significant curvilinear relationship between social isolation and mental health in patients with cancer. The quadratic term ($-0.136x^2$) represents the curvilinear relationship between the 2 variables. The negative coefficient indicates that the relationship is stronger at very high or very low levels of social isolation and weaker at moderate levels. The linear term (−0.321) represents the overall trend in the relationship between the 2 variables, with a negative coefficient indicating that mental health decreases as social isolation increases. It is important to remember that this equation represents the overall trend in the relationship between social isolation and mental health, but individual patients may not necessarily follow this trend. In other words, the QE suggests a threshold or optimal level of social isolation associated with better mental health outcomes. This optimal level may differ for individuals, depending on their personalities, coping strategies, and support networks; for instance, some patients with cancer may find that a certain degree of social isolation allows them to focus on their needs, engage in self-reflection, and develop a sense of independence and self-reliance, which can help to reduce mental distress. By contrast, too much social isolation may lead to feelings of loneliness, helplessness, and despair, which can negatively affect mental health.

The QE implies that interventions to reduce social isolation in patients with cancer should consider the complex and curvilinear relationship between social isolation and mental health. This may involve tailoring interventions to individual needs, providing different levels and types of social support, and encouraging patients to develop a sense of control and agency over their social relationships.

We observed significant direct and total effects of social isolation and negative self-perception. A significant direct effect of social isolation on negative self-perception with a coefficient of 0.36 would indicate that an increase in social isolation is associated with an increase in negative self-perception and vice versa. Therefore, we fail to reject hypothesis 2.

We did not observe any significant direct effect of poor communication on patient mental health; however, specific indirect and total effects were significant, failing to reject hypothesis 3. The finding of a significant indirect effect of poor physician-patient communication on the mental health of patients with cancer via negative self-perception, as indicated by a negative coefficient (−0.03) and a statistically significant *P* value, suggests that poor physician-patient communication may influence the mental health of patients with cancer through its effect on negative self-perception. The negative coefficient for the indirect effect implies that higher levels of communication are associated with better mental health outcomes through their effect on negative self-perception. This finding suggests that poor physician-patient communication may protect mental health in patients with cancer, potentially by reducing negative self-perception. poor physician-patient communication also had a significant direct and total effect on negative self-perception, implying that patients who experience inadequate communication with their care team will develop negative self-perception. Therefore, we fail to reject hypothesis 4.

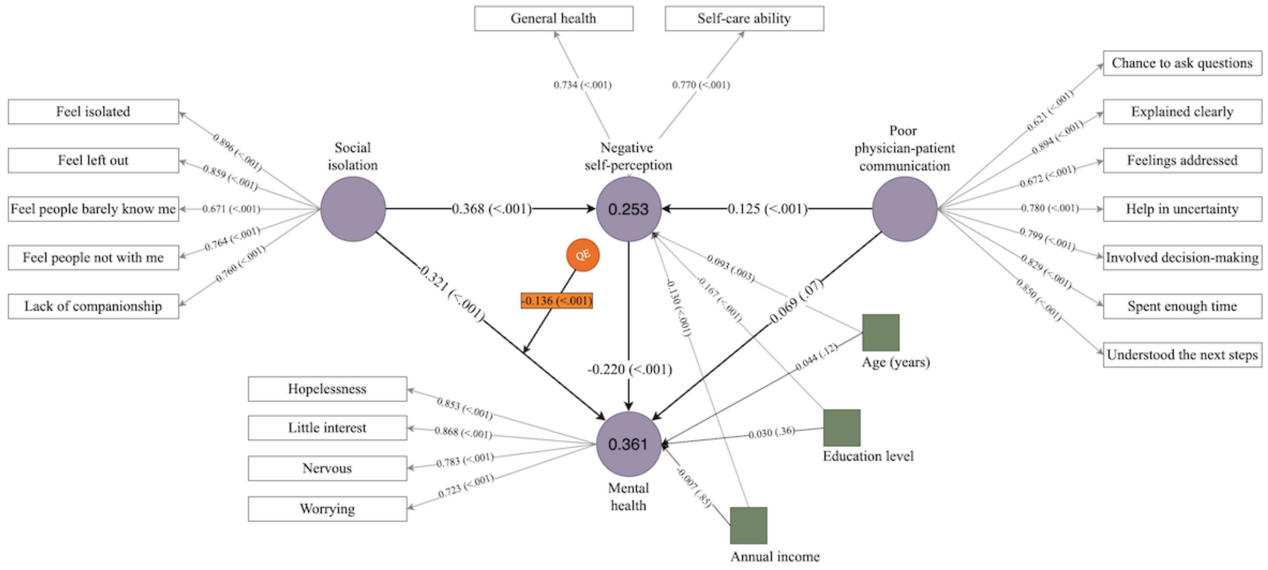
In addition, the model identified a significant direct effect of negative self-perception on mental health, where increasing negative self-perception would hinder a patient's mental health. The control variables, including age, annual income, and education, had no significant effect on mental health, but the indirect and total effects were significant. The direct effects of the control variables on negative self-perception were also significant. Patients with higher education and income had better self-perception (better perception of general health and better ability for self-care). Contrastingly, negative self-perception was found to increase with age. Figure 2 illustrates the final model.

Table 3. Standardized direct, indirect, and total effects.

Paths	Estimate (β)	Standardized mean estimate	2-tailed *t* test	*P* value
Direct effects				
Social isolation → negative self-perception	.36	0.03	9.86	<.001
Social isolation → mental health	−.32	0.04	6.67	<.001
QE[a] (social isolation) → mental health	−.13	0.03	3.64	<.001
Poor communication with care provider → negative self-perception	.12	0.03	3.25	.001
Poor communication with care provider → mental health	−.07	0.03	1.82	.07
Negative self-perception → mental health	−.22	0.04	5.16	<.001
Age → mental health	.04	0.02	1.53	.12
Age → negative self-perception	.09	0.03	3.03	.003
Annual income → mental health	−.01	0.04	0.18	.86
Annual income → negative self-perception	−.13	0.03	3.54	<.001
Education level → mental health	.03	0.03	0.91	.36
Education level → negative self-perception	−.16	0.03	4.85	<.001
Specific indirect effects				
Poor communication with care provider → negative self-perception → mental health	−.03	0.01	2.60	.009
Social isolation → negative self-perception → mental health	−.08	0.02	4.72	<.001
Age → negative self-perception → mental health	−.02	0.01	2.55	.01
Education level → negative self-perception → mental health	.03	0.01	3.41	.001
Annual income → negative self-perception → mental health	.02	0.01	2.86	.004
Total effects				
Social isolation → mental health	−.40	0.47	8.79	<.001
Social isolation → negative self-perception	.36	0.03	9.86	<.001
QE (social isolation) → mental health	−.13	0.03	3.64	<.001
Poor communication with care provider → mental health	−.10	0.04	2.57	.01
Poor communication with care provider → negative self-perception	.13	0.04	3.26	.001
Negative self-perception → mental health	−.22	0.04	5.16	<.001
Age → mental health	.02	0.02	0.80	.42
Age → negative self-perception	.09	0.03	3.02	.003
Annual income → mental health	.02	0.03	0.60	.54
Annual income → negative self-perception	−.13	0.03	3.54	<.001
Education level → mental health	.06	0.03	2.05	.04
Education level → negative self-perception	−.16	0.03	4.85	<.001

[a]QE: quadratic effect.

Figure 2. Structural framework illustrating direct and indirect relationships among social isolation, mental health (measured using the 4-item Patient Health Questionnaire for Anxiety and Depression [PHQ-4]), poor physician-patient communication, and negative self-perception. The model shows the path coefficient and its significance. The observed factors indicated by square boxes are the control variables, and the values on the 2 endogenous latent constructs are the R2 values. QE: quadratic effect.



MGA Results

Table 4 shows the conceptual framework across two groups: (1) group A consisted of newly diagnosed patients with cancer who were undergoing cancer treatment during the survey or had received cancer treatment within the last 12 months, that is, receipt of cancer treatment during the COVID-19 pandemic; and (2) group B consisted of respondents who had received cancer treatment between 5 and 10 years previously, that is, receipt of cancer treatment before the COVID-19 pandemic. We noted that the direct effect of negative self-perception on mental health was significantly higher among patients treated before the COVID-19 pandemic (group B). Similarly, the indirect effect of social isolation on mental health was significantly higher among patients who had received cancer treatment before the COVID-19 pandemic. We did not observe any significant differences in other effects.

Table 4. A multigroup analysis of patients with cancer receiving treatment during and before the COVID-19 pandemic.

Paths	Estimate (β; group A-B)	*t* test (group A-B)	*P* value (group A-B)
Direct effects			
Social isolation → negative self-perception	.01	0.09	.92
Social isolation → mental health	.11	0.92	.36
Poor communication with care provider → negative self-perception	−.13	1.24	.24
Poor communication with care provider → mental health	.05	0.56	.58
Negative self-perception → mental health	−.21	2.23	.03
Age → mental health	.03	0.35	.72
Age → negative self-perception	−.07	0.79	.43
Annual income → mental health	.03	0.34	.73
Annual income → negative self-perception	.03	0.33	.74
Education level → mental health	−.02	0.22	.83
Education level → negative self-perception	.09	1.02	.31
Specific indirect effects			
Poor communication with care provider → negative self-perception → mental health	−.01	0.18	.86
Social isolation → negative self-perception → mental health	−.06	2.01	.04
Age → negative self-perception → mental health	−.01	0.69	.49
Education level → negative self-perception → mental health	.01	0.67	.50
Annual income → negative self-perception → mental health	.02	1.22	.22

Discussion

Principal Findings

The findings of this study provide important insights into the factors that affect the mental health of patients with cancer. Our results suggest that social isolation, negative self-perception, and communication with care providers are significantly related to mental health in patients with cancer.

It is common for patients with cancer to experience mental health challenges such as anxiety, depression, and distress, which can be further exacerbated by social isolation. Several studies have assessed the impact of social isolation on mental health; for instance, a 2021 study reported a significant association between social isolation and loneliness in patients with cancer during the COVID-19 pandemic [29]. The study also acknowledged the correlation between loneliness and depressive symptoms, including suicidal ideation [29]. Another study acknowledged the positive correlation between social isolation and mental health (symptoms of anxiety and depression) in patients with breast cancer [30]. Supporting existing evidence, our study observed a significant direct effect, where increased social isolation was responsible for worsening the mental health of patients with cancer. Adding to the body of knowledge, we also found a significant indirect effect of social isolation on mental health, that is, a significant mediation effect of negative self-perception. It is worth noting that most of these findings are consistent with previous research, but discrepancies may arise because of differences in the type of cancer or the stage of the illness.

Another novelty of our study is the quadratic (curvilinear) effect of social isolation on the mental health of patients with cancer. This finding, as indicated by a statistically significant coefficient of −0.136, suggests that the relationship between the 2 variables is not necessarily a simple linear relationship; rather, it follows a more complex pattern. The negative coefficient (−0.136) for the quadratic term indicates that the relationship between social isolation and mental health is stronger at very high or very low levels of social isolation and weaker at moderate levels. This finding suggests that there may be a threshold level of social isolation beyond which mental health outcomes deteriorate more rapidly. Therefore, minimizing social isolation might not be optimal for reducing mental distress among patients with cancer. Instead, attempts should be made to provide moderate isolation to the patients where they stay, with adequate time for self-reflection and access to social activities.

The linear term (−0.321) represents the overall trend in the relationship between social isolation and mental health, with a negative coefficient indicating that mental health decreases as social isolation increases. This finding is consistent with the idea that social isolation is generally associated with negative mental health outcomes. However, the magnitude of this effect may vary, depending on the level of social isolation. Overall, these findings suggest that social isolation is an important factor to consider in the mental health of patients with cancer and that interventions to reduce social isolation may be an effective way to improve mental health outcomes in this population. Further research is needed to understand the nature of the relationship between social isolation and mental health in patients with cancer, as well as the potential moderating or mediating factors that may influence this relationship. It is also important to consider the implications of these findings for clinical practice; for example, health care providers may need to pay particular attention to the social isolation levels of patients with cancer and address any potential issues that may arise. This may involve providing support and resources to help patients maintain social connections or referring patients to social workers or other professionals who can provide additional support.

The finding of a significant effect of negative self-perception on mental health, as indicated by a coefficient of −0.22 and a statistically significant *P* value, suggests that self-perception is a crucial determinant of mental health in patients with cancer. The negative coefficient for the effect of negative self-perception on mental health indicates that lower levels of negative self-perception are associated with better mental health outcomes. This finding suggests that self-perception may have a protective effect on mental health in patients with cancer, potentially by providing individuals with a more positive view of themselves and their abilities. Consistent with our findings, a 2021 study reported a significant association between self-perception and mental health in patients with cancer [18]. Another study in 2017 found a relationship between negative self-perception and deteriorated physical and mental health [31]. This longitudinal study observed 100 patients with cancer and followed the relationship between self-perception and mental health for a year [31].

In our study, physician-patient communication was found to be positively associated with mental health, suggesting that patients with cancer who reported higher levels of satisfaction with their communication with their physicians had higher levels of mental health. This is consistent with research showing that effective physician-patient communication is important in promoting mental health [32]. However, our study is the first to explore this relationship explicitly for patients with cancer. Effective communication can expedite patient recovery [33], not necessarily as a direct effect but possibly via indirect routes such as establishing physician-patient trust, understanding, and agreement. Ultimately, these proximal outcomes lead to overall well-being through improved access to care, better patient knowledge, shared decision-making, management of emotions, and patient empowerment. The positive impact of physician-patient communication on patients' self-perception, as observed in our study, has also been noted by others in a different context. A 2022 study reported the positive impact of effective physician-patient communication on patients' perception of safety and security, thereby augmenting their self-perception [34]. Another study in 2020 dealing with 250 patients with hypertension reported a significant impact of effective physician-patient communication on patients' perception of self-care ability and satisfaction as well as pharmaceutical adherence in patients with hypertension [35]. These findings support the notion that improving physician-patient communication may enhance self-perception in patients with cancer.

Our study observed that the magnitude of the indirect impact of social isolation on mental health, when mediated by negative self-perception, was significantly greater in patients with cancer who were treated before the COVID-19 pandemic than in those who received treatment during the pandemic. This implies that negative self-perception played a much stronger role in determining the mental health of patients with cancer before the pandemic. It is also important to note that the relationships between these variables may vary based on patient characteristics such as age, literacy, and income; for example, older patients with cancer may be more vulnerable to social isolation because of a higher prevalence of physical limitations and a smaller social network. By contrast, younger patients with cancer may be more vulnerable to social isolation because of a lack of experience with illness and a greater reliance on social support.

Similarly, patients with cancer with lower literacy levels may have difficulty understanding medical information and experience poorer communication with their health care providers. This may affect their mental health outcomes. Finally, patients with cancer with lower incomes may have limited access to health care and may experience financial stress, which can affect their mental health.

Implications and Limitations

Our findings have important implications for the overall treatment of patients with cancer. They suggest that interventions addressing social isolation, improving self-perception, and enhancing physician-patient communication may improve the mental health of patients with cancer. These interventions could include support groups, cognitive behavioral therapy, and training for health care providers in effective communication skills. Future research should examine the effectiveness of these interventions in different subgroups of patients with cancer, such as those with different ages, literacy levels, and income levels.

It is important to acknowledge the limitations of this study. One potential limitation is that the data were collected through self-report measures subject to biases such as social desirability and self-presentation. In addition, the results may not be generalizable to the larger population of patients with cancer. It would be valuable to replicate this study with a larger sample to further examine the relationships among the variables. Furthermore, the study was conducted during a particularly fluid period when social isolation was the norm because of the COVID-19 pandemic. It is important to consider whether the curvilinear relationship between social isolation and mental health observed in our study will hold true once the pandemic has fully resolved. Although we acknowledge that the long-term relevance of our findings in nonpandemic contexts is uncertain, the curvilinear relationship between social isolation and mental health observed in our study is an important finding that may have implications beyond the current pandemic. Future research should explore the impact of social isolation on mental health in patients with cancer in nonpandemic contexts, which may help to elucidate further the complex relationship between social isolation and mental health. Despite the uncertainties surrounding the long-term relevance of our findings, we believe that our study provides valuable insights into the impact of social isolation on the mental health of patients with cancer, particularly during times of heightened social isolation, such as during the COVID-19 pandemic.

Data Availability

The study does not contain identifiable patient data. Consent from the individuals involved in this study was not required. The author cannot grant requests for any underlying data because the data were acquired from the Health Information National Trends Survey (HINTS) under the Microdata Dissemination Policy. Individuals and organizations wishing to access the data can make a request directly to the HINTS service.

Conflicts of Interest

None declared.

References

1. Agustina E, Dodd RH, Waller J, Vrinten C. Understanding middle-aged and older adults' first associations with the word "cancer": a mixed methods study in England. Psychooncology 2018 Jan;27(1):309-315 [FREE Full text] [doi: 10.1002/pon.4569] [Medline: 29047225]
2. Quaife SL, Winstanley K, Robb KA, Simon AE, Ramirez AJ, Forbes LJ, et al. Socioeconomic inequalities in attitudes towards cancer: an international cancer benchmarking partnership study. Eur J Cancer Prev 2015 May;24(3):253-260 [FREE Full text] [doi: 10.1097/CEJ.0000000000000140] [Medline: 25734238]
3. Vrinten C, Gallagher A, Waller J, Marlow LA. Cancer stigma and cancer screening attendance: a population based survey in England. BMC Cancer 2019 Jun 11;19(1):566 [FREE Full text] [doi: 10.1186/s12885-019-5787-x] [Medline: 31185949]
4. Schroyen S, Marquet M, Jerusalem G, Dardenne B, Van den Akker M, Buntinx F, et al. The link between self-perceptions of aging, cancer view and physical and mental health of older people with cancer: a cross-sectional study. J Geriatr Oncol 2017 Jan;8(1):64-68. [doi: 10.1016/j.jgo.2016.09.001] [Medline: 27659547]
5. Tripathi L, Datta SS, Agrawal SK, Chatterjee S, Ahmed R. Stigma perceived by women following surgery for breast cancer. Indian J Med Paediatr Oncol 2017 Apr;38(2):146-152 [FREE Full text] [doi: 10.4103/ijmpo.ijmpo_74_16] [Medline: 28900322]

6. Phelan SM, Griffin JM, Jackson GL, Zafar SY, Hellerstedt W, Stahre M, et al. Stigma, perceived blame, self-blame, and depressive symptoms in men with colorectal cancer. Psychooncology 2013 Jan;22(1):65-73 [FREE Full text] [doi: 10.1002/pon.2048] [Medline: 21954081]
7. Mitchell AJ, Ferguson DW, Gill J, Paul J, Symonds P. Depression and anxiety in long-term cancer survivors compared with spouses and healthy controls: a systematic review and meta-analysis. Lancet Oncol 2013 Jul;14(8):721-732. [doi: 10.1016/S1470-2045(13)70244-4] [Medline: 23759376]
8. Moore S, Leung B, Bates A, Ho C. Social isolation: impact on treatment and survival in patients with advanced cancer. J Clin Oncol 2018;36(34_suppl):156 [FREE Full text] [doi: 10.1200/JCO.2018.36.34_suppl.156]
9. Pahl DA, Wieder MS, Steinberg DM. Social isolation and connection in adolescents with cancer and survivors of childhood cancer: a systematic review. J Adolesc 2021 Feb;87:15-27. [doi: 10.1016/j.adolescence.2020.12.010] [Medline: 33450464]
10. Smith GL, Lopez-Olivo MA, Advani PG, Ning MS, Geng Y, Giordano SH, et al. Financial burdens of cancer treatment: a systematic review of risk factors and outcomes. J Natl Compr Canc Netw 2019 Oct 01;17(10):1184-1192 [FREE Full text] [doi: 10.6004/jnccn.2019.7305] [Medline: 31590147]
11. Short PF, Moran JR, Punekar R. Medical expenditures of adult cancer survivors aged <65 years in the United States. Cancer 2011 Jun 15;117(12):2791-2800 [FREE Full text] [doi: 10.1002/cncr.25835] [Medline: 21656757]
12. Quach T, Nuru-Jeter A, Morris P, Allen L, Shema SJ, Winters JK, et al. Experiences and perceptions of medical discrimination among a multiethnic sample of breast cancer patients in the greater San Francisco Bay area, California. Am J Public Health 2012 May;102(5):1027-1034. [doi: 10.2105/AJPH.2011.300554] [Medline: 22420791]
13. Caston NE, Waters AR, Williams CP, Biddell C, Spees L, Gallagher K, et al. Patient-reported discrimination among limited-resourced cancer survivors: a brief report. J Psychosoc Oncol (forthcoming) 2022 Dec 15:1-12. [doi: 10.1080/07347332.2022.2154186] [Medline: 36519613]
14. Bamonti PM, Moye J, Naik AD. Pain is associated with continuing depression in cancer survivors. Psychol Health Med 2018 Dec;23(10):1182-1195 [FREE Full text] [doi: 10.1080/13548506.2018.1476723] [Medline: 29901408]
15. Li XM, Xiao WH, Yang P, Zhao HX. Psychological distress and cancer pain: results from a controlled cross-sectional survey in China. Sci Rep 2017 Jan 11;7:39397 [FREE Full text] [doi: 10.1038/srep39397] [Medline: 28074915]
16. Galloway SK, Meadors P, Boselli D, Walsh D. Anxiety, depression, pain, and social support in a large representative cancer population. J Clin Oncol 2019 Nov 01;37(31_suppl):76 [FREE Full text] [doi: 10.1200/jco.2019.37.31_suppl.76]
17. Yılmaz M, Dissiz G, Usluoğlu AK, Iriz S, Demir F, Alacacioglu A. Cancer-related stigma and depression in cancer patients in a middle-income country. Asia Pac J Oncol Nurs 2019 Oct 23;7(1):95-102 [FREE Full text] [doi: 10.4103/apjon.apjon_45_19] [Medline: 31879690]
18. Niveau N, New B, Beaudoin M. How should self-esteem be considered in cancer patients? Front Psychol 2021 Oct 28;12:763900 [FREE Full text] [doi: 10.3389/fpsyg.2021.763900] [Medline: 34777169]
19. Foltz AT. The influence of cancer on self-concept and life quality. Semin Oncol Nurs 1987 Nov;3(4):303-312. [doi: 10.1016/s0749-2081(87)80023-2] [Medline: 3321276]
20. Health Information National Trends Survey. National Cancer Institute. 2023. URL: https://hints.cancer.gov [accessed 2023-03-11]
21. Westat. Health information national trends survey 5 (HINTS 5): cycle 4 methodology report. Health Information National Trends Survey, National Cancer Institute. 2020 Dec. URL: https://hints.cancer.gov/docs/methodologyreports/HINTS5_Cycle4_MethodologyReport.pdf [accessed 2023-02-12]
22. Hesse BW, Greenberg AJ, Peterson EB, Chou WY. The health information national trends survey (HINTS): a resource for consumer engagement and health communication research. Stud Health Technol Inform 2017;240:330-346. [Medline: 28972526]
23. Kroenke K, Spitzer RL, Williams JB, Löwe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. Psychosomatics 2009 Nov;50(6):613-621. [doi: 10.1176/appi.psy.50.6.613] [Medline: 19996233]
24. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, PROMIS Cooperative Group. The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol 2010 Nov;63(11):1179-1194 [FREE Full text] [doi: 10.1016/j.jclinepi.2010.04.011] [Medline: 20685078]
25. Moser RP, Trivedi N, Murray A, Jensen RE, Willis G, Blake KD. Patient-Centered Communication (PCC) scale: psychometric analysis and validation of a health survey measure. PLoS One 2022 Dec 30;17(12):e0279725 [FREE Full text] [doi: 10.1371/journal.pone.0279725] [Medline: 36584146]
26. Hair JF, Ringle CM, Sarstedt M. PLS-SEM: indeed a silver bullet. J Mark Theory Pract 2011 Apr;19(2):139-152 [FREE Full text] [doi: 10.2753/mtp1069-6679190202]
27. Matthews L. Applying multigroup analysis in PLS-SEM: a step-by-step process. In: Latan H, Noonan R, editors. Partial Least Squares Path Modeling: Basic Concepts, Methodological Issues and Applications. Cham, Switzerland: Springer; Aug 14, 2017:219-243.
28. Henseler J, Fassott G, Dijkstra TK, Wilson B. Analysing quadratic effects of formative constructs by means of variance-based structural equation modelling. Eur J Inf Syst 2011 Sep 06;21(1):99-112 [FREE Full text] [doi: 10.1057/ejis.2011.36]

29. Miaskowski C, Paul SM, Snowberg K, Abbott M, Borno HT, Chang SM, et al. Loneliness and symptom burden in oncology patients during the COVID-19 pandemic. Cancer 2021 Sep 01;127(17):3246-3253 [FREE Full text] [doi: 10.1002/cncr.33603] [Medline: 33905528]
30. Liu B, Wu X, Shi L, Li H, Wu D, Lai X, et al. Correlations of social isolation and anxiety and depression symptoms among patients with breast cancer of Heilongjiang province in China: the mediating role of social support. Nurs Open 2021 Jul;8(4):1981-1989 [FREE Full text] [doi: 10.1002/nop2.876] [Medline: 33939294]
31. Schroyen S, Missotten P, Jerusalem G, Van den Akker M, Buntinx F, Adam S. Association between self-perception of aging, view of cancer and health of older patients in oncology: a one-year longitudinal study. BMC Cancer 2017 Sep 02;17(1):614 [FREE Full text] [doi: 10.1186/s12885-017-3607-8] [Medline: 28865449]
32. Choi BM, Obeng-Kusi M, Axon DR. Association between patient-provider communication and self-perceived mental health in us adults with cancer: real-world evidence through medical expenditure panel survey. Diseases 2022 Oct 15;10(4):88 [FREE Full text] [doi: 10.3390/diseases10040088] [Medline: 36278587]
33. King A, Hoppe RB. "Best practice" for patient-centered communication: a narrative review. J Grad Med Educ 2013 Sep;5(3):385-393 [FREE Full text] [doi: 10.4300/JGME-D-13-00072.1] [Medline: 24404300]
34. Kulińska J, Rypicz Ł, Zatońska K. The impact of effective communication on perceptions of patient safety-a prospective study in selected polish hospitals. Int J Environ Res Public Health 2022 Jul 27;19(15):9174 [FREE Full text] [doi: 10.3390/ijerph19159174] [Medline: 35954529]
35. Świątoniowska-Lonc N, Polański J, Tański W, Jankowska-Polańska B. Impact of satisfaction with physician-patient communication on self-care and adherence in patients with hypertension: cross-sectional study. BMC Health Serv Res 2020 Nov 16;20(1):1046 [FREE Full text] [doi: 10.1186/s12913-020-05912-0] [Medline: 33198739]

Abbreviations

AVE: average variance extracted
BCA: bias-corrected and accelerated
GAD-2: 2-item Generalized Anxiety Disorder scale
HINTS: Health Information National Trends Survey
HTMT: heterotrait-monotrait
MGA: multigroup analysis
NCI: National Cancer Institute
PHQ-2: 2-item Patient Health Questionnaire
PHQ-4: 4-item Patient Health Questionnaire
PLS-SEM: partial least squares structural equation modeling
PPC: physician-patient communication
PROMIS: Patient-Reported Outcomes Measurement Information System
QE: quadratic effect
SEER: Surveillance, Epidemiology, and End Results
SEM: structural equation modeling
SRMR: standardized root mean square residual
VIF: variance inflation factor

Edited by A Mavragani; submitted 28.12.22; peer-reviewed by M Katz, M Thomson; comments to author 21.02.23; revised version received 11.03.23; accepted 11.03.23; published 07.04.23.

Please cite as:
Choudhury A
Impact of Social Isolation, Physician-Patient Communication, and Self-perception on the Mental Health of Patients With Cancer and Cancer Survivors: National Survey Analysis
Interact J Med Res 2023;12:e45382
URL: https://www.i-jmr.org/2023/1/e45382
doi:10.2196/45382
PMID:37027201

Original Paper

Effects of Antidepressants on COVID-19 Outcomes: Retrospective Study on Large-Scale Electronic Health Record Data

Md Mahmudur Rahman[1*], MSc; Atqiya Munawara Mahi[1*], BSc; Rachel Melamed[2], PhD; Mohammad Arif Ul Alam[1,3], PhD

[1]The Richard A Miner School of Computer & Information Sciences, University of Massachusetts, Lowell, MA, United States

[2]Department of Biological Sciences, University of Massachusetts Lowell, Lowell, MA, United States

[3]Department of Medicine, University of Massachusetts Chan Medical School, Worcester, MA, United States

[*]these authors contributed equally

Corresponding Author:
Mohammad Arif Ul Alam, PhD
The Richard A Miner School of Computer & Information Sciences
University of Massachusetts
220 Pawtucket St
Lowell, MA, 01854
United States
Phone: 1 978 934 1971
Email: mohammadariful_alam@uml.edu

Abstract

Background: Antidepressants exert an anticholinergic effect in varying degrees, and various classes of antidepressants can produce a different effect on immune function. While the early use of antidepressants has a notional effect on COVID-19 outcomes, the relationship between the risk of COVID-19 severity and the use of antidepressants has not been properly investigated previously owing to the high costs involved with clinical trials. Large-scale observational data and recent advancements in statistical analysis provide ample opportunity to virtualize a clinical trial to discover the detrimental effects of the early use of antidepressants.

Objective: We primarily aimed to investigate electronic health records for causal effect estimation and use the data for discovering the causal effects of early antidepressant use on COVID-19 outcomes. As a secondary aim, we developed methods for validating our causal effect estimation pipeline.

Methods: We used the National COVID Cohort Collaborative (N3C), a database aggregating health history for over 12 million people in the United States, including over 5 million with a positive COVID-19 test. We selected 241,952 COVID-19–positive patients (age >13 years) with at least 1 year of medical history. The study included a 18,584-dimensional covariate vector for each person and 16 different antidepressants. We used propensity score weighting based on the logistic regression method to estimate causal effects on the entire data. Then, we used the Node2Vec embedding method to encode SNOMED-CT (Systematized Nomenclature of Medicine-Clinical Terms) medical codes and applied random forest regression to estimate causal effects. We used both methods to estimate causal effects of antidepressants on COVID-19 outcomes. We also selected few negatively effective conditions for COVID-19 outcomes and estimated their effects using our proposed methods to validate their efficacy.

Results: The average treatment effect (ATE) of using any one of the antidepressants was −0.076 (95% CI −0.082 to −0.069; *P*<.001) with the propensity score weighting method. For the method using SNOMED-CT medical embedding, the ATE of using any one of the antidepressants was −0.423 (95% CI −0.382 to −0.463; *P*<.001).

Conclusions: We applied multiple causal inference methods with novel application of health embeddings to investigate the effects of antidepressants on COVID-19 outcomes. Additionally, we proposed a novel drug effect analysis–based evaluation technique to justify the efficacy of the proposed method. This study offers causal inference methods on large-scale electronic health record data to discover the effects of common antidepressants on COVID-19 hospitalization or a worse outcome. We found that common antidepressants may increase the risk of COVID-19 complications and uncovered a pattern where certain antidepressants were associated with a lower risk of hospitalization. While discovering the detrimental effects of these drugs on outcomes could guide preventive care, identification of beneficial effects would allow us to propose drug repurposing for COVID-19 treatment.

(Interact J Med Res 2023;12:e39455) doi:10.2196/39455

KEYWORDS

causal inference; treatment effect; drug effect; COVID-19 outcomes; COVID-19 severity; drug repurposing; COVID-19; depression; mental health; data mining; electronic health record; machine learning; antidepressant; causal inference method

Introduction

The COVID-19 outbreak [1], which was declared a pandemic in 2020 [2], is a devastating health crisis that needs new preventive strategies and treatments. One characteristic distinguishing this pandemic from others is the remarkable heterogeneity of outcomes among infected people. While some patients have mild illness, 18% have moderate or severe outcomes [3,4]. Worse outcomes have been associated with several risk factors, including age [5], sex [6-8], socioeconomic background, and comorbidities, such as obesity [9-11], chronic obstructive pulmonary disease [12-15], type 2 diabetes [16,17], and hypertension [18-20]. Yet, these risk factors do not fully explain the variation in outcomes. Some drugs may change the course of COVID-19 [21,22]. Discovering these either beneficial or harmful effects could improve medical care. For instance, certain cyclooxygenase inhibitors, which are common anti-inflammatory drugs, have been associated with worse outcomes, suggesting that some pain relievers should be avoided in COVID-19 patients [23]. On the other hand, discovering medications associated with improved outcomes can help us identify new therapies. From the early stages of the outbreak, a number of drugs have been proposed for repurposing, including hydroxychloroquine, which was notorious, and remdesivir, a broad-spectrum antiviral, which was successful [24]. Because the SARS-CoV-2 virus targets the renin-angiotensin-aldosterone system through its interaction with the ACE2 receptor, previous investigations have used the cohort study method to investigate infections and outcomes in people taking ACE inhibitors or angiotensin receptor blockers [25]. Those results indicated that a protective effect could be identified from retrospective analyses of people on these medications.

Building on these encouraging findings, we aimed to discover whether other classes of medications could impact outcomes. We focused on antidepressants, which are common drugs used by over 13% of adults in the United States [26]. Antidepressants have been linked to unexpected effects on diverse inflammatory and cardiovascular outcomes [27]. Use of antidepressants has been associated with an increased risk of hospital mortality [28], possibly due to their cardiovascular effects [27]. In a small study in France, protective effects on severe COVID-19 outcomes were found [29]. In another study, Clelland et al [30] showed a significant protective association between antidepressant use and COVID-19. Hoertel et al [31] showed a protective effect of fluoxetine or fluvoxamine on COVID-19 mortality. In a separate study, Hoertel et al [29] used a Cox regression model to investigate the association between antidepressant intake and the risk of intubation and COVID-19 mortality. Because of the popularity of antidepressants and their previous associations with complications that are relevant to COVID-19 outcomes [32,33], we investigated the possible effects of antidepressants on COVID-19 using a large population in the United States.

While previous work has assessed the association of medication use with COVID-19 severity [34], including the studies mentioned above, the work was limited by both a small population size and minimal adjustment for confounding. For instance, Israel et al [34] estimated the effects of drugs on COVID-19 using a case-control method. They matched COVID-19 cases against a control cohort of COVID-19–negative people on a set of 5 to 12 selected confounders. This limited adjustment for confounding is typical of previous studies, which have not controlled for possible confounders, including history of high cholesterol and other recorded medical care. Factors, such as socioeconomic status, can influence disease risk and influence levels of medical care, and such factors could create confounding. It is possible that people with more medical care, including those diagnosed and in treatment for other medical conditions, are at reduced risk for complications. If all confounders are known and well measured, then such an approach could work well. However, given the lack of knowledge about the risk factors for COVID-19, approaches controlling for a minimal set of possible confounders are vulnerable to residual confounding. It is crucial to critically assess the methods used for estimating these results.

Here, we made use of the National COVID Cohort Collaborative (N3C), a database aggregating health history for over 12 million people in the United States, including over 5 million with a positive COVID-19 test. In addition to health history for each COVID-19 patient, this data set provides a severity score for each patient, based on the World Health Organization COVID-19 severity scale [3]. To estimate the effects of antidepressants on COVID-19 outcomes, we considered 16 antidepressants from the N3C data enclave, with each used by more than 5000 patients. To estimate the effects of these drugs on outcomes, we applied causal inference methods, including a novel application of health embeddings. Our methods build on approaches like the case-control or cohort study to estimate effects of exposures on an outcome of interest. While measurement of true causal effects requires a randomized trial, such approaches are expensive and unlikely to be performed for every common drug. Instead, causal inference methods aim to emulate a clinical trial using observational health data. While the resulting inferences cannot be conclusively deemed causal, they represent our best possible estimate using nonexperimental data. Therefore, unlike previous work, we used causal inference methods to rigorously adjust for confounding. We used both a well-established method (high-dimensional propensity score [35] weighting) and a relatively less common method based on embeddings of medical codes [36]. The contributions of this work include both the estimates of the effects of antidepressants, and the rigorous assessment and comparison of methods for causal effect inference.

Methods

Data Sources

Our analysis made use of the N3C resource, which aggregates data on over 12 million people in the United States across dozens of sites of care. This population includes over 5 million people with COVID-19. Data sources were united using the OMOP Common Data Model, which allowed common concept identifiers to be created and a common format to be achieved across diverse data sources. An application under the Data Use Request system allowed us to access the deidentified version of the data. These data create a comprehensive portrait of the health history of millions of people, with loss of only exact dates and exact locations for each person.

From the set of COVID-19–positive people, we obtained their subsequent severity score previously calculated based on the World Health Organization index [3]. For each person, this is the most severe encounter in their medical history, based on a 5-level scoring system. The levels are mild, mild with emergency department visit, moderate with hospitalization, severe with hospitalization, and hospital mortality. Because of the small size of the population with a severe condition or mortality, we grouped together all hospitalized patients (around 20% of the positive population) to identify how antidepressant use affects hospitalization. People missing a severity score were considered as nonhospitalized, since any record of hospitalization is likely to have been noted in the health record. Therefore, our focus was on identifying causal effects on the presence of the hospitalization outcome. Notably, this measure has been previously used to assess demographic factors associated with COVID-19 outcomes [3].

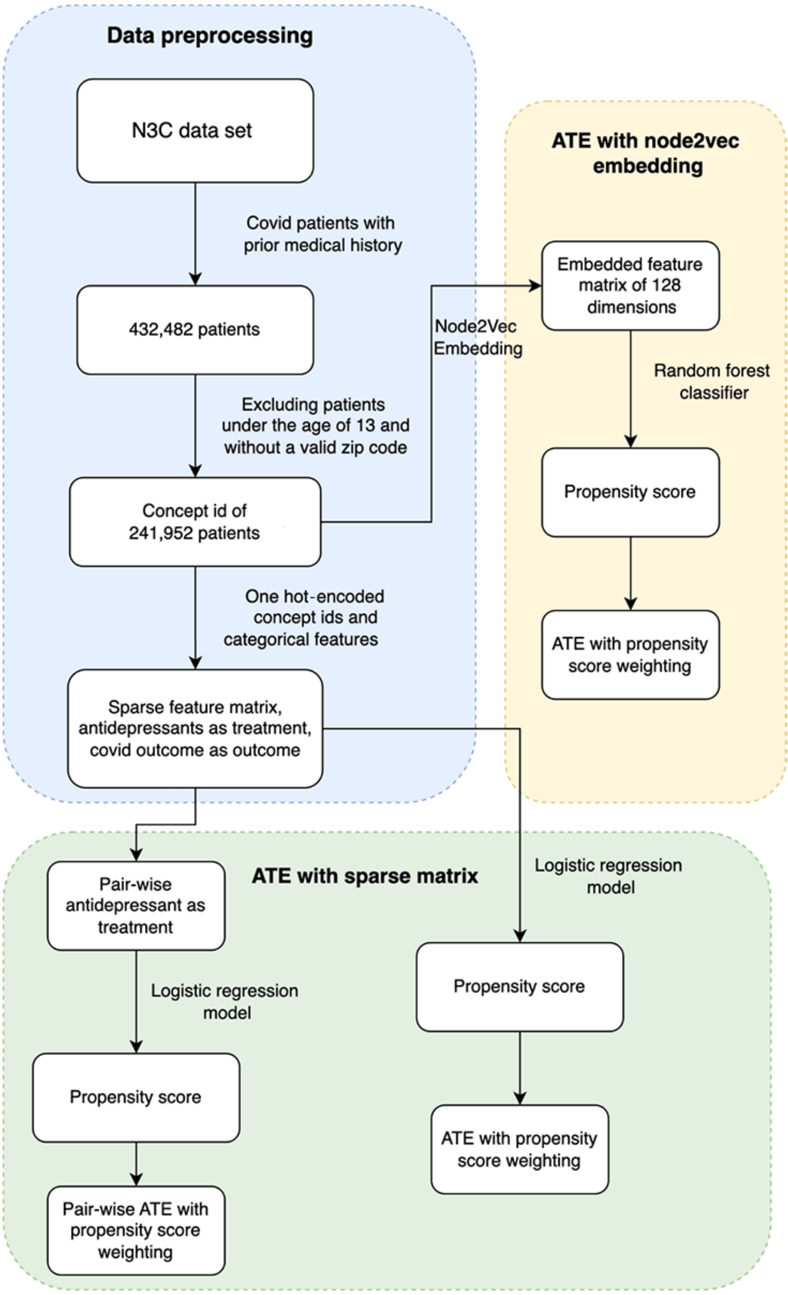
Study Population

We identified 16 common antidepressants using the OMOP concept relation data. First, we obtained all concepts of the type "ingredient" that are descendants of the ATC class "antidepressants" (OMOP concept ID 21604686). Then, we obtained all drugs that contained these ingredients and obtained all instances of use of these drugs, using the condition_era table in N3C. We retained all ingredients used by more than 5000 people to create a set of 16 antidepressants.

Antidepressants are divided into 5 classes based on which neurotransmitter they affect. Among the 16 antidepressants we considered for our study, fluoxetine, paroxetine, sertraline, citalopram, and escitalopram are classified as selective serotonin reuptake inhibitors (SSRIs); duloxetine, venlafaxine, and desvenlafaxine are classified as serotonin and norepinephrine reuptake inhibitors (SNRIs); trazodone, mirtazapine, vortioxetine, vilazodone, and bupropion are classified as atypical antidepressants; and nortriptyline, amitriptyline, and doxepin are classified as tricyclic antidepressants [37]. Monoamine oxidase inhibitors are a type of antidepressant that can cause potentially serious side effects, and they are rarely prescribed by doctors nowadays [37]. Moreover, our data set had no data points that involved antidepressants from the monoamine oxidase inhibitor class. Hence, we ignored this class in our study.

Among the COVID-19–positive population, we further restricted our analysis to those who had a medical history of at least 1 year. This is common in pharmacoepidemiology studies to obtain an adequate history of the study population. We further restricted our study to those with an age of over 13 years and with a valid zip code. Eventually, we identified 241,952 individuals taking one or more antidepressants (Figure 1).

Figure 1. Overview of causal inference estimates and methods for obtaining them. ATE: average treatment effect; N3C: National COVID Cohort Collaborative.



Ethical Considerations

In this study, we used deidentified observational data from the N3C [38], and this is not considered human subject research. N3C has approved the data for secondary use without the need for institutional review board approval and has approved this study for publication/submission.

Causal Inference Analysis

We performed multiple analyses to assess the causal effect while controlling for all measured confounders. First, we obtained all health history and demographics preceding the positive COVID-19 diagnosis for each patient in the data set. This set of variables consisted of all possible diagnosis codes (from the "condition_era" table), as well as age, gender, race, ethnicity, and the 3-digit zip code (from the "person" table).

We identified a number of treatment effects of interest. We were interested in estimating the effect of taking each antidepressant versus not taking each antidepressant (nonuser analysis), and additionally the difference in effects for each pair of antidepressants (active comparator analysis). Each treatment effect of interest identifies a pair of populations: the treated cohort (all COVID-19–positive people who are taking the drug of interest) and the comparison cohort (all COVID-19–positive people who are either not taking the drug of interest in the nonuser analysis or are taking another drug in the active comparator analysis).

The average treatment effect (ATE) was defined as the mean difference in outcomes between the two cohorts. If h_i represents the hospitalization outcome for the *i*th person, the formula for the ATE is as follows:

One major issue with the use of an observational data set instead of performing a randomized controlled trial is the risk of having selection bias in the experimental setup. We expect the pair of cohorts to differ in terms of health history and demographics, which can confound an unadjusted estimate of the ATE. Therefore, it is important to adjust for these differences to obtain an unbiased effect estimate. A common method to adjust for confounding is propensity score weighting, which creates a weighted pseudopopulation where treated comparator populations are balanced for possible confounders [39]. Lee et al [40] elaborately explained how machine learning models improve the performance of propensity score weighting. Pan et al [41] presented some references on how classification and regression models provide an improved version of propensity score weighting. Moreover, we included all medical history data of the patients as features, which resulted in high-dimensional input feature vectors (18,584-dimensional). Processing high-dimensional data is computationally expensive and is not feasible in almost all existing methods other than sparse logistic regression [42]. Considering this, here, we implemented propensity score weighting using sparse logistic regression and random forest. Both methods share the goal of representing possible confounders, and we performed 2 representations to avoid sensitivity to misspecification of the model for propensity. The propensity score represents the probability of each person falling into the treated or comparator cohorts, given their history and demographics, as follows:

p(treated | person i's health history, demographics) = pi (2)

We estimated this propensity score using 2 different and complementary methods, and then, we used this score to weight each person's overall contribution to the estimate of the ATE as follows:

For each causal effect of interest, we estimated the propensity for treatment using 2 different ways of encoding health history (Figure 1). First, we estimated propensity for treatment by performing a high-dimensional regularized logistic regression, which was fitted to model p_i separately for each causal effect of interest. For this analysis, we encoded each of the health history and demographic variables using the one-hot encoding scheme. Therefore, we modeled all previous diagnoses and treatments, creating a resulting 18,584-dimensional covariate vector for each person. Second, we used an embedding representation of patient health status at the time of the COVID-19–positive test. The embedding representation was precomputed by Pattisapu et al [36], using the Node2Vec method to encode SNOMED-CT (Systematized Nomenclature of Medicine-Clinical Terms) medical concepts to the embedded vector space. For each of the 18,584 health history codes, we matched the code to its 128-dimensional pretrained embedding vector e_c. Then, for the *i*th person, given their list of previous medical codes $\{codes_i\}$, we created an overall representation of patient health by averaging these vectors as follows:

Then, we modeled the propensity for treatment given the vector of the patient health state. Logistic regression was not feasible for this large nonsparse data, so we used random forest, which is also a popular tool for estimating the propensity score. Here, the ATE was assessed on the full data set. We chose these methods because the first (high-dimensional propensity score) is the more standard method and the second (embedding) can potentially account for poorly measured confounders [43]. By performing both types of causal inferences, we can evaluate the sensitivity of our results to specifications of the propensity model. These approaches have the potential to adjust for confounding, unlike previous methods [29-31].

Obtaining CIs

We obtained CIs using the bootstrap method. Specifically, we sampled with replacement to obtain our pair of cohorts. For each sample, we estimated the propensity weights and used this to estimate the overall causal effect. This process was repeated 100 times to create 100 estimates, providing the CIs.

Assessing Our Results Using Negative Controls

The practice of using negative control outcomes, which are outcomes thought not to be causally affected by an exposure, is intended to form a point of comparison for our causal effects of interest. We selected negative control outcomes using literature on known causal effects of antidepressants, selecting some common outcomes that are not likely to be the result of antidepressant use. We selected the following: fracture of bone (SNOMED-CT code 125605004), asthma (SNOMED-CT code 195967001), chronic kidney disease (SNOMED-CT code 709044004), disorder of nail (SNOMED-CT code 17790008), and eczema (SNOMED-CT code 43116000). For each negative outcome, we estimated the causal effect in the same way as for our outcome of interest (hospitalization with COVID-19).

Results

Topics of Interest

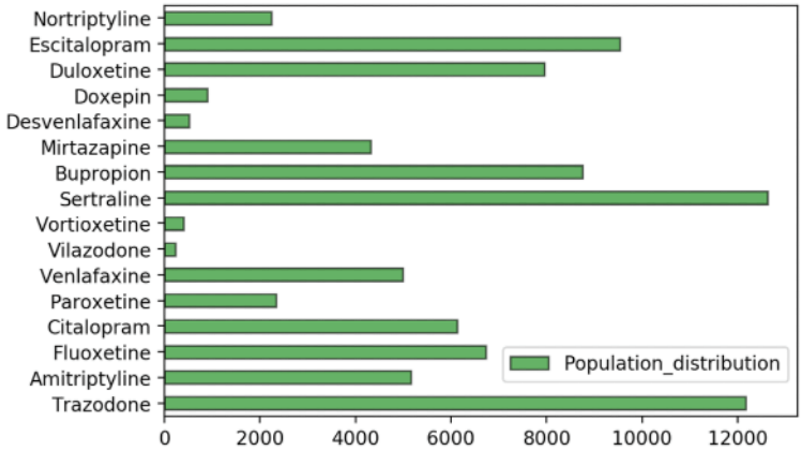
Our main results addressed 2 topics of interest. First, we were interested in discovering new effects of drugs on COVID-19 outcomes, as measured by the severity score. Second, we wanted to evaluate our methods in order to contribute to the causal inference literature.

Causal Effects of Interest

In order to discover the actionable unknown effects of drugs on COVID-19 trajectory, we focused on a set of causal effects of interest. We were interested in the effect of each common antidepressant on COVID-19 hospitalization outcome. Figure 2 shows the frequency of prescription of each antidepressant in the population. We followed the approach of emulating randomized trials using observational data [44]. We created target randomized trials to follow the user versus nonuser design and to follow the active comparator design. In the user versus nonuser design, we emulated a trial where people are randomized to either using an antidepressant or not using an antidepressant. In the active comparator design, the target trial compared people taking one antidepressant versus another. Each such target trial defined 2 populations of interest: the treated and comparison populations. Then, for each causal effect of interest, we used multiple methods to estimate the relationship. Therefore, we performed one effect estimate for each antidepressant in a user versus nonuser design, and one estimate for each pair of antidepressants and comparison of each antidepressant to another in an active comparator design.

We obtained the population of people with a positive polymerase chain reaction test from the N3C data and obtained the score calculating the severity of their COVID-19 outcomes. We further identified those people with a history of taking antidepressants before their positive test. Using these data, we identified the treated and comparison cohorts for each effect of interest. To emulate a randomized trial, we must adjust for any medical history that may create a biased association between the treatment and outcome. We adjusted for all medical history data before the positive COVID-19 test using the propensity score weighting method to obtain the adjusted ATE (see Methods). We calculated CIs by creating 100 bootstrap samples of the data set (see Methods). We used 2 methods to encode medical history in order to calculate the propensity score: high-dimensional sparse representation of history, and representation by medical code embeddings. These 2 methods share the goal of representing possible confounders, but we intended for these 2 complementary representations to enable critical assessment of the methods and their effect estimates.

Figure 2. Population distribution for each antidepressant. This represents the number of patients who took each of the antidepressants as treatment.



Causal Effect Estimates Indicate a Significant Impact of Antidepressants on Hospitalization

The results indicated a significantly worse outcome (higher rates of hospitalization among users) (Figure 3). In order to assess whether these results are specific to hospitalization outcomes or rather some difference in overall sickness between the 2 cohorts, we selected a set of negative control outcomes. Good negative controls are those that may be associated with confounding variables, such as overall sickness, but are not associated with the exposure of interest [45,46]. We used information on the known side effects of antidepressants to select 5 negative control outcomes (fracture, asthma, chronic kidney disease, nail disorder, and eczema) that were intended to represent diverse medical states unrelated to antidepressant use. While antidepressants showed significant associations with the outcome of interest (hospitalization), all negative control outcomes had no significant association with antidepressants (Figure 4).

Figure 3. Average rate of nonhospitalization in the user versus nonuser design. ATE: average treatment effect.

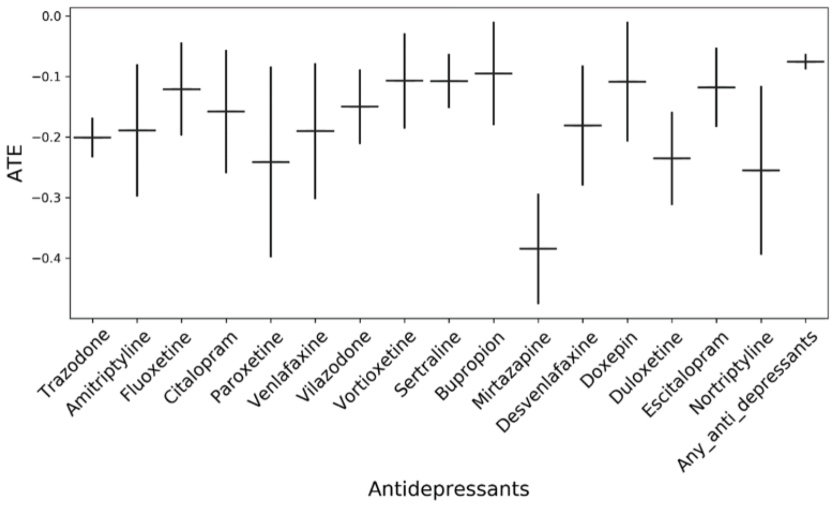
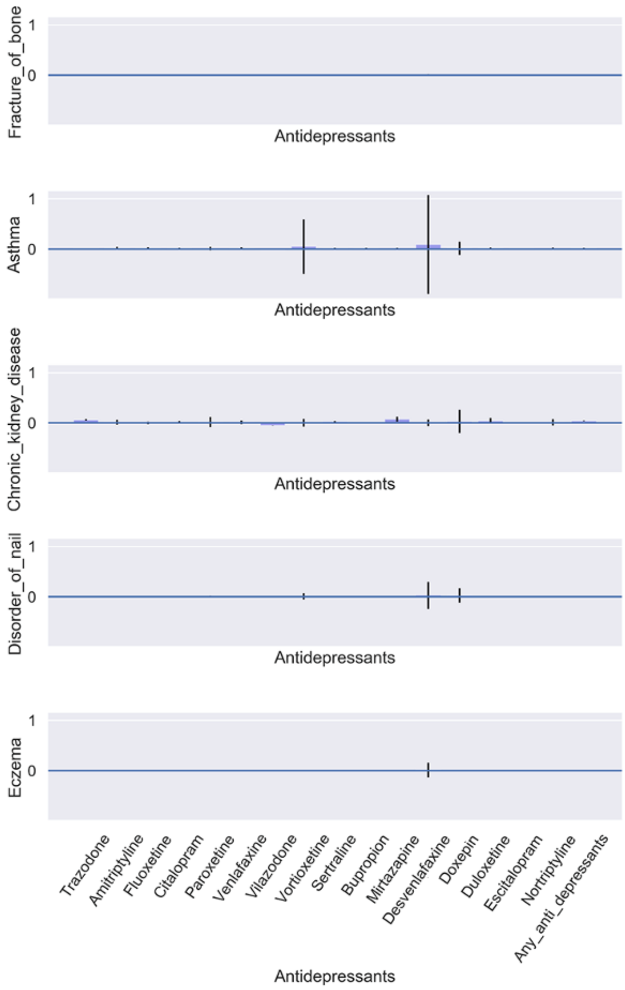


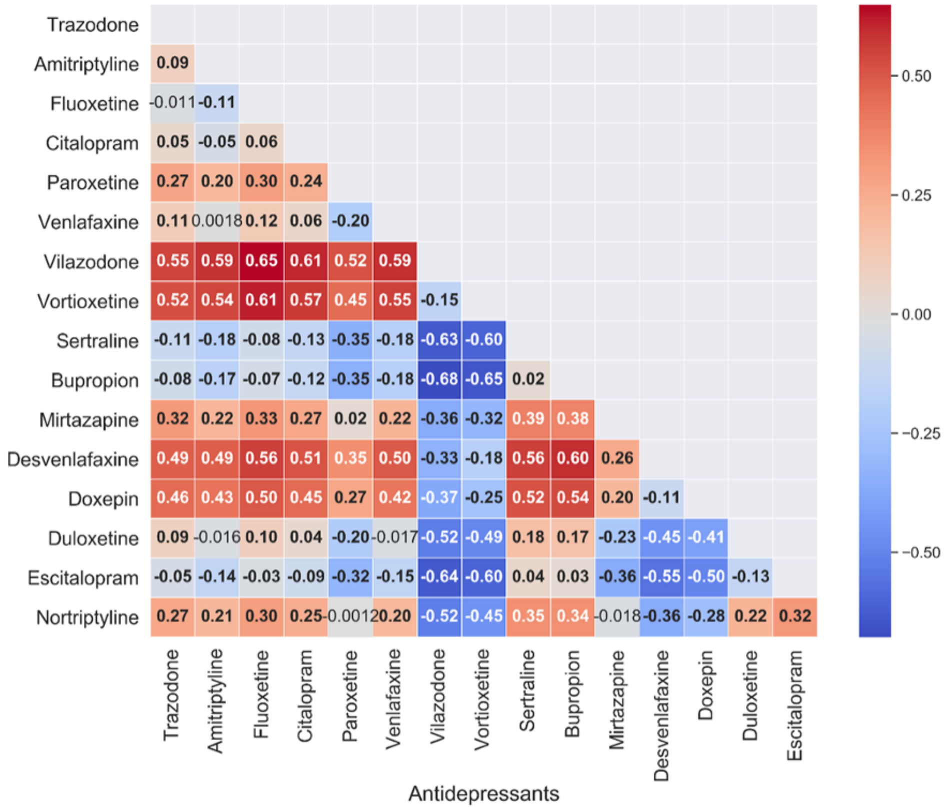
Figure 4. Negative control average treatment effects.



Active Comparator Design Suggests Differences in Outcomes Between Antidepressants

While all antidepressants appeared to increase the risk of hospitalization, we also performed a head-to-head active comparator analysis to assess diversity in the effects. Vilazodone and vortioxetine, as compared to the other antidepressants, appeared to confer some protection against hospitalization. This may be due to uncharacterized cardiovascular effects, which have been described in certain contexts (Figure 5).

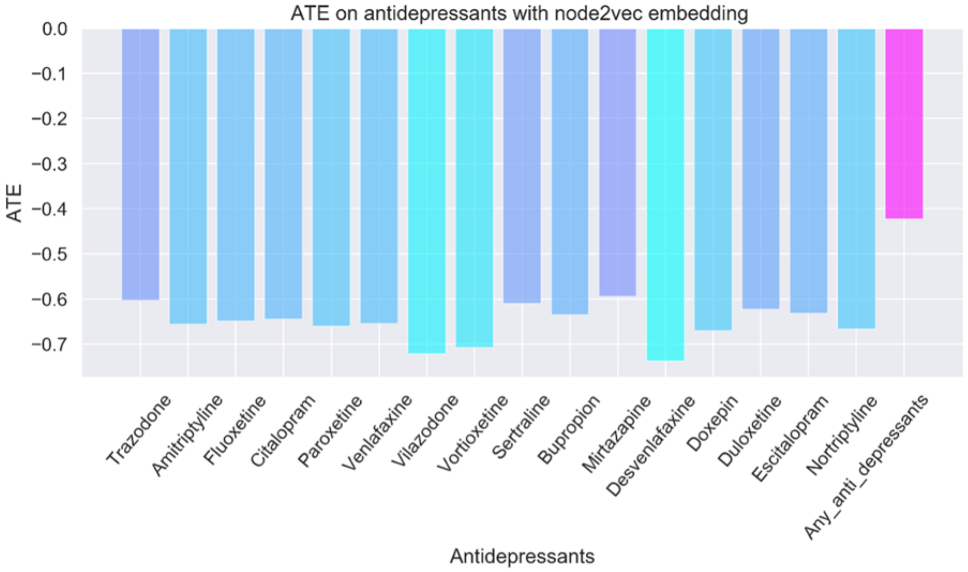
Figure 5. Results of the active comparator design comparing each antidepressant to others. The average treatment effect is shown in each box of the treatment (row) versus comparator (column), with bold numbers indicating a CI not overlapping the null effect.



Comparing the High-Dimensional Propensity Score Against the Embedding-Based Propensity Score

The high-dimensional propensity score required a sparse encoding that did not on its own capture any meaning of medical codes. That is, each of the 10,000 codes modeled in the propensity score must be modeled using a one-hot vector, and all context for each code is lost. This means that 2 medical codes for very similar conditions, for example, fracture of the right leg and fracture of the left leg, are encoded no more similarly than 2 medical codes for unrelated conditions. In order to examine the impact of a coding system that retains the meaning of the medical codes, we made use of medical embeddings that exist for the SNOMED-CT concept coding system [36]. While embeddings have been explored previously for matching [43], the use of embeddings to create propensity score weights has not been reported to our knowledge. We created an embedding representation of patient health history using the average of all medical codes in a person's history. Then, as with the high-dimensional sparse representation of health history, we calculated propensity weights and obtained the ATE. The ATEs were much more extreme using this method (Figure 6), and the negative control outcomes similarly had a biased result (not shown). This contrasts with the high-dimensional propensity weighting method, where the negative control outcomes, as expected, retain a null effect.

Figure 6. Average treatment effect using embedding representation of patient history. ATE: average treatment effect.



Discussion

Principal Findings

This study aimed to apply causal inference methods to discover whether taking any common antidepressants is associated with poor COVID-19 outcomes, and to compare different methods for assessing these effects. Our findings suggest that common antidepressants may increase the risk of COVID-19 complications. Additionally, in our analysis of the effect of each common antidepressant as compared to that of other antidepressants, we uncovered a pattern where certain antidepressants were associated with a lower risk of hospitalization. We also experimented with multiple methods for encoding health history to uncover causal effects. Because health data involve thousands of medical codes, representing each of these codes using a sparse representation can create a very large regression problem for propensity weighting. In addition, this representation of medical history does not make use of knowledge about the meaning of these medical codes. In order to make use of this information, we alternatively encoded medical history using embedding vectors created using the Node2Vec method. These embeddings have been extensively evaluated as an efficient representation of medical knowledge [36]. We found that both methods had a consistent direction of effect, but the effect estimates were more extreme using embeddings rather than the sparse encoding.

Our findings about the effects of antidepressants, if replicated in other data sets, could suggest that providers should change their uses of antidepressants to improve COVID-19 outcomes among high-risk groups. Among prior studies, Lenze et al [47] performed a randomized controlled trial of 152 patients and showed a significantly reduced risk of worse clinical outcomes in patients with symptomatic COVID-19 who were administered fluoxetine than in those who were administered placebo. Oskotsky et al [48] performed an observational study on COVID-19 mortality and implemented a propensity score matching method only on the exposure of some specific SSRIs (fluoxetine or fluvoxamine). Moreover, several clinical and preclinical studies found an association between fluoxetine intake and a lower risk of intubation or mortality [29,49,50]. However, these findings do not answer if other SSRI or non-SSRI antidepressants are as effective as fluoxetine or fluvoxamine. In contrast, our study considered 16 antidepressants of 4 types, including SSRI, SNRI, atypical, and tricyclic antidepressants. The results indicated significant associations of citalopram, escitalopram, venlafaxine, desvenlafaxine, mirtazapine, doxepin, and vilazodone with a reduced risk of worse COVID-19 outcomes, apart from fluoxetine. Some other prior studies assessing the association between antidepressant intake and COVID-19 severity have used limited adjustment for confounders or small populations [29-31]. Our study has made extensive efforts to adjust for confounding. The results support and are consistent with the findings of prior preliminary studies. Further, our study showed that the association between antidepressant intake and a reduced risk of COVID-19 mortality, intubation, or a worse outcome is not only for fluoxetine or SSRIs, and other antidepressants from several classes have similar effects on COVID-19 outcomes. Our consistent results from 2 causal inference methods support that these methods can be used to mine possible effects from large health record data. As the methods are not specific to antidepressant advances, these findings suggest that the N3C data set could be used with these methods to investigate other factors impacting COVID-19 outcomes, including other drugs and other medical procedures and treatments.

Limitations

In a nonrandomized setting, it is not possible to be certain that the results are free of residual confounding. Although our method carefully considered all medical history data, thus accounting for all measured confounding, unmeasured confounding could still bias the results. To mitigate this risk, we have undertaken an analysis using negative control outcomes. One possible example of unmeasured confounding is if people taking antidepressants generally have poor health. Poor health is not directly recorded in the N3C. However, in this case, we would expect an association between antidepressants and increased prevalence of many other diseases, such as the set of negative control outcomes we selected for analysis. Because we did not find any association of antidepressants with these negative control outcomes, our results do not appear to be due to this type of confounding. Other limitations include the short duration of observation for our data set, as we only used 3 years of data to estimate confounding. Some confounders may be recorded only more distantly in health history, but this time window is commonly used in observational data analysis. We also did not use the duration or dose of antidepressants; therefore, our results represent the impact of any use of antidepressants on disease outcomes.

Another caveat concerns the embedding effect estimates. This method estimated treatment effects that were much more extreme than those in the more traditional encoding of health history. Under a conservative interpretation, we believe this is more likely due to the shortcomings of this approach, which makes it susceptible to bias, rather than being due to a true extreme causal effect. The bias in the results could be due to one of the following reasons. First, the embedding vectors do not precisely represent important confounders. As the vectors are only 128-dimensional, some information about specific medical codes that may be crucial confounders may be lost. Second, the method for calculating propensity weights based on medical embeddings must be improved. This may involve developing other ways to represent a patient's health history given a set of embeddings. Third, these embeddings are not designed to represent a patient's state before drug prescription, and performance may be improved by applying medical embeddings specifically designed to represent the confounding relationship between health history and drug prescription [43]. Further experimentation is needed to assess how best to use embedding vectors for causal inference.

Conclusions

In this study, we investigated how antidepressants affected COVID-19 outcomes, using causal inference methods. In addition to standard propensity score analysis, we implemented a novel application of health embeddings. To support the effectiveness of the suggested strategy, we also offered a novel drug effect analysis–based evaluation tool. This study used causal inference techniques on large electronic health record data to identify how commonly prescribed antidepressants affect hospitalization for COVID-19 or a worse outcome. The research suggested a pattern in which some antidepressants are connected to a decreased risk of hospitalization. Because the risk profile of antidepressants is well known, our findings can be used to provide justification for investment in future large-scale clinical trials to find the best treatment for depression in those with COVID-19 at high risk of poor outcomes. Future work can build on our methods to identify more factors influencing COVID-19 outcomes to help predict who is at high risk and to suggest interventions.

Data Availability

All the analyses were conducted using the National COVID Cohort Collaborative (N3C) database, which is available to researchers and investigators conducting COVID-19–related research, subject to certain eligibility criteria and data use agreements. Access to the data is granted through a secure online platform that requires registration and approval by the N3C Data Enclave Governance Committee. Researchers who are interested in using the cohort we generated from original N3C data, are asked to get N3C approval first and then contact the corresponding author, MAUA (mohammadariful_alam@uml.edu). For more information about accessing the N3C data and the eligibility criteria, please visit the N3C website [51].

Conflicts of Interest

None declared.

References

1. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 2020 Feb 20;382(8):727-733. [doi: 10.1056/nejmoa2001017]
2. Responding to community spread of COVID-19. World Health Organization. URL: https://www.who.int/publications/i/item/responding-to-community-spread-of-covid-19 [accessed 2023-03-09]
3. Bennett TD, Moffitt RA, Hajagos JG, Amor B, Anand A, Bissell MM, et al. The National COVID Cohort Collaborative: Clinical Characterization and Early Severity Prediction. medRxiv Preprint posted online on January 23, 2021 [FREE Full text] [doi: 10.1101/2021.01.12.21249511] [Medline: 33469592]
4. Wingert A, Pillay J, Gates M, Guitard S, Rahman S, Beck A, et al. Risk factors for severity of COVID-19: a rapid review to inform vaccine prioritisation in Canada. BMJ Open 2021 May 13;11(5):e044684 [FREE Full text] [doi: 10.1136/bmjopen-2020-044684] [Medline: 33986052]
5. Kang S, Jung SI. Age-Related Morbidity and Mortality among Patients with COVID-19. Infect Chemother 2020 Jun;52(2):154-164 [FREE Full text] [doi: 10.3947/ic.2020.52.2.154] [Medline: 32537961]
6. Biswas M, Rahaman S, Biswas TK, Haque Z, Ibrahim B. Association of Sex, Age, and Comorbidities with Mortality in COVID-19 Patients: A Systematic Review and Meta-Analysis. Intervirology 2020 Dec 09;64(1):1-12 [FREE Full text] [doi: 10.1159/000512592] [Medline: 33296901]
7. Mesas AE, Cavero-Redondo I, Álvarez-Bueno C, Sarriá Cabrera MA, Maffei de Andrade S, Sequí-Dominguez I, et al. Predictors of in-hospital COVID-19 mortality: A comprehensive systematic review and meta-analysis exploring differences

by age, sex and health conditions. PLoS One 2020 Nov 3;15(11):e0241742 [FREE Full text] [doi: 10.1371/journal.pone.0241742] [Medline: 33141836]
8. Goldstein JR, Lee RD. Demographic perspectives on the mortality of COVID-19 and other epidemics. Proc Natl Acad Sci U S A 2020 Sep 08;117(36):22035-22041 [FREE Full text] [doi: 10.1073/pnas.2006392117] [Medline: 32820077]
9. Gao F, Zheng KI, Wang XB, Sun QF, Pan KH, Wang TY, et al. Obesity Is a Risk Factor for Greater COVID-19 Severity. Diabetes Care 2020 Jul;43(7):e72-e74. [doi: 10.2337/dc20-0682] [Medline: 32409499]
10. Popkin BM, Du S, Green WD, Beck MA, Algaith T, Herbst CH, et al. Individuals with obesity and COVID-19: A global perspective on the epidemiology and biological relationships. Obes Rev 2020 Nov 26;21(11):e13128 [FREE Full text] [doi: 10.1111/obr.13128] [Medline: 32845580]
11. Kamyari N, Soltanian AR, Mahjub H, Moghimbeigi A. Diet, Nutrition, Obesity, and Their Implications for COVID-19 Mortality: Development of a Marginalized Two-Part Model for Semicontinuous Data. JMIR Public Health Surveill 2021 Jan 26;7(1):e22717 [FREE Full text] [doi: 10.2196/22717] [Medline: 33439850]
12. Zhao Q, Meng M, Kumar R, Wu Y, Huang J, Lian N, et al. The impact of COPD and smoking history on the severity of COVID-19: A systemic review and meta-analysis. J Med Virol 2020 Oct 17;92(10):1915-1921 [FREE Full text] [doi: 10.1002/jmv.25889] [Medline: 32293753]
13. Gerayeli FV, Milne S, Cheung C, Li X, Yang CWT, Tam A, et al. COPD and the risk of poor outcomes in COVID-19: A systematic review and meta-analysis. EClinicalMedicine 2021 Mar;33:100789 [FREE Full text] [doi: 10.1016/j.eclinm.2021.100789] [Medline: 33758801]
14. Leung JM, Niikura M, Yang CWT, Sin DD. COVID-19 and COPD. Eur Respir J 2020 Aug 13;56(2):2002108 [FREE Full text] [doi: 10.1183/13993003.02108-2020] [Medline: 32817205]
15. Tal-Singer R, Crapo JD. COPD at the Time of COVID-19: A COPD Foundation Perspective. Chronic Obstr Pulm Dis 2020 Apr;7(2):73-75 [FREE Full text] [doi: 10.15326/jcopdf.7.2.2020.0149] [Medline: 32324976]
16. Gamble A, Pham Q, Goyal S, Cafazzo JA. The Challenges of COVID-19 for People Living With Diabetes: Considerations for Digital Health. JMIR Diabetes 2020 May 15;5(2):e19581 [FREE Full text] [doi: 10.2196/19581] [Medline: 32392473]
17. Bhatti R, Khamis AH, Khatib S, Shiraz S, Matfin G. Clinical Characteristics and Outcomes of Patients With Diabetes Admitted for COVID-19 Treatment in Dubai: Single-Centre Cross-Sectional Study. JMIR Public Health Surveill 2020 Dec 07;6(4):e22471 [FREE Full text] [doi: 10.2196/22471] [Medline: 33284130]
18. Fang L, Karakiulakis G, Roth M. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? The Lancet Respiratory Medicine 2020 Apr;8(4):e21. [doi: 10.1016/s2213-2600(20)30116-8]
19. Schiffrin EL, Flack JM, Ito S, Muntner P, Webb RC. Hypertension and COVID-19. Am J Hypertens 2020 Apr 29;33(5):373-374 [FREE Full text] [doi: 10.1093/ajh/hpaa057] [Medline: 32251498]
20. Bonner C, Cvejic E, Ayre J, Isautier J, Semsarian C, Nickel B, et al. The Psychological Impact of Hypertension During COVID-19 Restrictions: Retrospective Case-Control Study. JMIRx Med 2021 Mar 30;2(1):e25610 [FREE Full text] [doi: 10.2196/25610] [Medline: 34076628]
21. Aygün İ, Kaya M, Alhajj R. Identifying side effects of commonly used drugs in the treatment of Covid 19. Sci Rep 2020 Dec 09;10(1):21508 [FREE Full text] [doi: 10.1038/s41598-020-78697-1] [Medline: 33299085]
22. FitzGerald GA. Misguided drug advice for COVID-19. Science 2020 Mar 27;367(6485):1434-1434. [doi: 10.1126/science.abb8034] [Medline: 32198292]
23. Reese J, Coleman B, Chan L, Blau H, Callahan TJ, Cappelletti L, et al. NSAID use and clinical outcomes in COVID-19 patients: a 38-center retrospective cohort study. Virol J 2022 May 15;19(1):84. [doi: 10.1186/s12985-022-01813-2]
24. Beigel J, Tomashek KM, Dodd LE, Mehta AK, Zingman BS, Kalil AC, et al. Remdesivir for the Treatment of Covid-19 — Final Report. N Engl J Med 2020 Nov 05;383(19):1813-1826. [doi: 10.1056/nejmoa2007764]
25. Hippisley-Cox J, Young D, Coupland C, Channon KM, Tan PS, Harrison DA, et al. Risk of severe COVID-19 disease with ACE inhibitors and angiotensin receptor blockers: cohort study including 8.3 million people. Heart 2020 Jul 31;106(19):1503-1511. [doi: 10.1136/heartjnl-2020-317393]
26. Brody DJ, Gu Q. Antidepressant Use Among Adults: United States, 2015-2018. CDC. 2020. URL: https://www.cdc.gov/nchs/products/databriefs/db377.htm [accessed 2023-03-09]
27. Smoller JW, Allison M, Cochrane BB, Curb JD, Perlis RH, Robinson JG, et al. Antidepressant use and risk of incident cardiovascular morbidity and mortality among postmenopausal women in the Women's Health Initiative study. Arch Intern Med 2009 Dec 14;169(22):2128-2139. [doi: 10.1001/archinternmed.2009.436] [Medline: 20008698]
28. Ghassemi M, Marshall J, Singh N, Stone DJ, Celi LA. Leveraging a critical care database: selective serotonin reuptake inhibitor use prior to ICU admission is associated with increased hospital mortality. Chest 2014 Apr;145(4):745-752 [FREE Full text] [doi: 10.1378/chest.13-1722] [Medline: 24371841]
29. Hoertel N, Sánchez-Rico M, Vernet R, Beeker N, Jannot A, Neuraz A, AP-HP / Universities / INSERM COVID-19 Research CollaborationAP-HP COVID CDR Initiative. Association between antidepressant use and reduced risk of intubation or death in hospitalized patients with COVID-19: results from an observational study. Mol Psychiatry 2021 Sep 04;26(9):5199-5212. [doi: 10.1038/s41380-021-01021-4] [Medline: 33536545]

30. Clelland CL, Ramiah K, Steinberg L, Clelland JD. Analysis of the impact of antidepressants and other medications on COVID-19 infection risk in a chronic psychiatric in-patient cohort. BJPsych Open 2021 Dec 03;8(1):e6 [FREE Full text] [doi: 10.1192/bjo.2021.1053] [Medline: 34859759]
31. Hoertel N. Do the Selective Serotonin Reuptake Inhibitor Antidepressants Fluoxetine and Fluvoxamine Reduce Mortality Among Patients With COVID-19? JAMA Netw Open 2021 Nov 01;4(11):e2136510 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.36510] [Medline: 34779851]
32. Sidik SM. Common antidepressant slashes risk of COVID death, study says. Nature 2021 Oct 29:Online ahead of print. [doi: 10.1038/d41586-021-02988-4] [Medline: 34716441]
33. Mas M, García-Vicente JA, Estrada-Gelonch A, Pérez-Mañá C, Papaseit E, Torrens M, et al. Antidepressant Drugs and COVID-19: A Review of Basic and Clinical Evidence. J Clin Med 2022 Jul 12;11(14):4038 [FREE Full text] [doi: 10.3390/jcm11144038] [Medline: 35887802]
34. Israel A, Schäffer AA, Cicurel A, Cheng K, Sinha S, Schiff E, et al. Identification of drugs associated with reduced severity of COVID-19 - a case-control study in a large population. Elife 2021 Jul 27;10:e68165 [FREE Full text] [doi: 10.7554/eLife.68165] [Medline: 34313216]
35. Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA. High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. Epidemiology 2009 Jul;20(4):512-522 [FREE Full text] [doi: 10.1097/EDE.0b013e3181a663cc] [Medline: 19487948]
36. Pattisapu N, Patil S, Palshikar G, Varma V. Medical Concept Normalization by Encoding Target Knowledge. In: Proceedings of the Machine Learning for Health NeurIPS Workshop, PMLR. 2019 Presented at: Machine Learning for Health Workshop; December 13, 2019; Vancouver, BC, Canada p. 246-259.
37. All about antidepressants. Medical News Today. URL: https://www.medicalnewstoday.com/articles/248320 [accessed 2023-03-09]
38. Haendel MA, Chute CG, Bennett TD, Eichmann DA, Guinney J, Kibbe WA, N3C Consortium. The National COVID Cohort Collaborative (N3C): Rationale, design, infrastructure, and deployment. J Am Med Inform Assoc 2021 Mar 01;28(3):427-443 [FREE Full text] [doi: 10.1093/jamia/ocaa196] [Medline: 32805036]
39. Li F, Morgan KL, Zaslavsky AM. Balancing Covariates via Propensity Score Weighting. Journal of the American Statistical Association 2017 Nov 13;113(521):390-400. [doi: 10.1080/01621459.2016.1260466]
40. Lee BK, Lessler J, Stuart EA. Improving propensity score weighting using machine learning. Stat Med 2010 Feb 10;29(3):337-346 [FREE Full text] [doi: 10.1002/sim.3782] [Medline: 19960510]
41. Pan W, Bai H. Propensity score methods for causal inference: an overview. Behaviormetrika 2018 Jul 24;45(2):317-334. [doi: 10.1007/s41237-018-0058-8]
42. Abramovich F, Grinshtein V. High-Dimensional Classification by Sparse Logistic Regression. IEEE Trans. Inform. Theory 2019 May;65(5):3068-3079. [doi: 10.1109/tit.2018.2884963]
43. Melamed R. Using indication embeddings to represent patient health for drug safety studies. JAMIA Open 2020 Oct;3(3):422-430 [FREE Full text] [doi: 10.1093/jamiaopen/ooaa040] [Medline: 33376961]
44. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. Am J Epidemiol 2016 Apr 15;183(8):758-764 [FREE Full text] [doi: 10.1093/aje/kwv254] [Medline: 26994063]
45. Ryan PB, Schuemie MJ, Gruber S, Zorych I, Madigan D. Empirical performance of a new user cohort method: lessons for developing a risk identification and analysis system. Drug Saf 2013 Oct 29;36 Suppl 1(S1):S59-S72. [doi: 10.1007/s40264-013-0099-6] [Medline: 24166224]
46. Schuemie MJ, Ryan PB, DuMouchel W, Suchard MA, Madigan D. Interpreting observational studies: why empirical calibration is needed to correct p-values. Stat Med 2014 Jan 30;33(2):209-218 [FREE Full text] [doi: 10.1002/sim.5925] [Medline: 23900808]
47. Lenze EJ, Mattar C, Zorumski CF, Stevens A, Schweiger J, Nicol GE, et al. Fluvoxamine vs Placebo and Clinical Deterioration in Outpatients With Symptomatic COVID-19: A Randomized Clinical Trial. JAMA 2020 Dec 08;324(22):2292-2300 [FREE Full text] [doi: 10.1001/jama.2020.22760] [Medline: 33180097]
48. Oskotsky T, Maric I, Tang A, Oskotsky B, Wong RJ, Aghaeepour N, et al. Mortality Risk Among Patients With COVID-19 Prescribed Selective Serotonin Reuptake Inhibitor Antidepressants. JAMA Netw Open 2021 Nov 01;4(11):e2133090 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.33090] [Medline: 34779847]
49. Hoertel N, Sánchez-Rico M, Cougoule C, Gulbins E, Kornhuber J, Carpinteiro A, et al. Repurposing antidepressants inhibiting the sphingomyelinase acid/ceramide system against COVID-19: current evidence and potential mechanisms. Mol Psychiatry 2021 Dec 12;26(12):7098-7099 [FREE Full text] [doi: 10.1038/s41380-021-01254-3] [Medline: 34385600]
50. Kornhuber J, Hoertel N, Gulbins E. The acid sphingomyelinase/ceramide system in COVID-19. Mol Psychiatry 2022 Jan 04;27(1):307-314 [FREE Full text] [doi: 10.1038/s41380-021-01309-5] [Medline: 34608263]
51. National COVID Cohort Collaborative. URL: https://covid.cd2h.org/ [accessed 2023-04-10]

Abbreviations

ATE: average treatment effect

N3C: National COVID Cohort Collaborative
SNOMED-CT: Systematized Nomenclature of Medicine-Clinical Terms
SNRI: serotonin and norepinephrine reuptake inhibitor
SSRI: selective serotonin reuptake inhibitor

Edited by T Leung, A Mavragani; submitted 10.05.22; peer-reviewed by K Adapa, V Chauhan; comments to author 23.11.22; revised version received 15.12.22; accepted 05.03.23; published 11.04.23.

Please cite as:
Rahman MM, Mahi AM, Melamed R, Alam MAU
Effects of Antidepressants on COVID-19 Outcomes: Retrospective Study on Large-Scale Electronic Health Record Data
Interact J Med Res 2023;12:e39455
URL: https://www.i-jmr.org/2023/1/e39455
doi:10.2196/39455
PMID:36881541

Original Paper

Face Mask Use and Associated Factors Among Students: Mixed Methods Study

Abreha Addis Gesese[1], MPH; Tut Duer Thot[2], MPH

[1]Department of Clinical Nursing, Gambella Teachers Education and Health Science College, Gambella, Ethiopia

[2]School of Public Health, College of Health and Medical Sciences, Haramaya University, Harar, Ethiopia

Corresponding Author:
Abreha Addis Gesese, MPH
Department of Clinical Nursing
Gambella Teachers Education and Health Science College
Gambella Town
Gambella, 113
Ethiopia
Phone: 251 949581901
Fax: 251 475510293
Email: abrhaddis09@gmail.com

Abstract

Background: COVID-19 has gravely affected the world, including students, due to the high level of contracting infections.

Objective: This study assessed the magnitude of mask use and associated factors among students.

Methods: A cross-sectional study using mixed methods was conducted among students at Gambella Teachers' Education and Health Science College, Gambella Region, Southwest Ethiopia, from March 5 to March 30, 2021. The stratified random sampling technique was used. Proportional allocation of samples was used to randomly select case teams, and a simple random sampling technique was used to recruit the students. The data were collected by trained and experienced enumerators. Data were entered into EpiData (version 3.1; EpiData Association) and exported to SPSS (version 22; IBM Corp) for analysis. Logistic regression was executed. The adjusted odds ratio (AOR) with the 95% CI was used to determine the association and strength with the outcome variable. The qualitative data were transcribed, translated, coded, and analyzed using thematic analysis. Then, the themes were used to triangulate the quantitative study.

Results: The study included a total of 379 participants and yielded a response rate of 95.5% (379/397). The majority of study participants were older than 25 years, with the mean age being 26.34 (SD 5.8) years. This study found that the magnitude of mask use among students was 87% (330/379). The odds of mask use were higher among students who were female (AOR 3.32, 95% CI 1.191-9.248), younger (AOR 2.55, 95% CI 1.155-5.627), agreed that not all persons with COVID-19 develop severe disease (AOR 3.38, 95% CI 1.36-8.41), agreed that there is currently no effective cure (AOR 6.28, 95% CI 1.36-28.99), performed proper washing with soap and water (AOR 0.027, 95% CI 0.004-0.182), had started to stay home (AOR 0.168, 95% CI 0.054-0.52), agreed that COVID-19 is fatal (AOR 0.236, 95% CI 0.084-0.666), agreed that a flu vaccine is sufficient for COVID-19 prevention (AOR 3.874, 95% CI 1.540-9.749), and disinfected equipment and working areas at least once a day (AOR 0.222, 95% CI 0.086-0.575).

Conclusions: This study found that the magnitude of mask use among students was relatively moderate in Ethiopia. Sex, age, agreeing that not all persons with COVID-19 develop severe disease, agreeing that there is currently no effective cure, performing proper washing with soap and water, starting to stay home, agreeing that COVID-19 is fatal, and agreeing that the flu vaccine is sufficient to prevent COVID-19 were independently associated with mask use among students. Therefore, colleges should aggressively encourage students to wear masks and monitor the implementation of COVID-19 prevention regulations along with the accessibility of masks.

(Interact J Med Res 2023;12:e41365) doi:10.2196/41365

KEYWORDS

face mask use; associated factors; COVID-19; Gambella; students; Ethiopia

Introduction

COVID-19 is a respiratory tract infection and a public health emergency of international concern [1]. It is epidemiologically associated with the Wuhan, Hubei Province, Seafood Wholesale Market, where birds, bats, snakes, and other wildlife animals are sold [2]. COVID-19 differs with respect to community spread and severity [3]. After exposure to the virus and an incubation period of 2 to 14 days, people with COVID-19 develop a wide range of symptoms, which have been reported to range from mild to severe illness and include fever or chills, cough, shortness of breath or difficulty breathing, fatigue, muscle or body aches, headache, loss of taste or smell, sore throat, congestion or runny nose, nausea or vomiting, and diarrhea. Patients with COVID-19 can experience only mild or uncomplicated illness, and approximately 14% develop a severe illness that requires hospitalization and oxygen support, with 5% requiring admission to an intensive care unit [1,4]. COVID-19 can be transmitted through droplets, direct and indirect contact, and aerosols in long-range transmission (ie, airborne transmission) [5].

Globally, as of June 28, 2021, there had been 182,037,151 confirmed cases, 3,942,149 deaths, and 166,525,346 recovered cases of COVID-19 reported to the Worldometer [6]. The United States, India, and Brazil were among the countries with the highest confirmed cases in 2021 [6-8]. In Africa, approximately 3,942,448 infected patients in 47 countries led to a cumulative 94,217 deaths [6]. In Ethiopia, after the first report of a COVID-19 case, more than 276,000 confirmed cases and 4300 deaths have occurred [8].

Rigorous trials have brought COVID-19 vaccines, although these have side effects and quality differences, and they are the subject of popular myths [9,10]. However, there is lack of effective and approved drugs to treat COVID-19 in certain countries. Instead, strategic efforts have focused on combatting the spread of the disease. Recommendations have been made to prevent the most contagious viral diseases; one of these is the use of face masks. This is the recommendation to the general public that they should wear nonmedical masks in indoor settings (eg, shops, shared workplaces, and schools) and outdoor settings where physical distancing is not possible to a minimum interval of 1 meter [11]. The universal use of face masks can contribute to the containment of the virus in the community if they are adequately available, properly used, and properly disposed of after use [12]. The absence of clear scientific evidence for aerosol transmission of SARS-CoV-2 provides the rationale for the current recommendations for the use of surgical masks among health care professionals. There are also other means of preventing COVID-19, such as frequent hand washing with soap and sanitizer [13]. A study from the United Kingdom suggested that making masks mandatory in secondary schools would be of benefit but would need to be combined with scaling up of test-trace-isolate coverage to prevent resurgence of COVID-19. We highlight that the adoption of masks in schools, in addition to community settings, can help reduce epidemic resurgence but that to do this effectively, access to masks has to be sufficiently high. Studies show that if access is lower, the estimated reduction in COVID-19 resurgence is smaller. Uncertainties concerning the effectiveness of masks remain, and these results add to the ongoing body of evidence on the impact of using face masks to combat epidemics [14]. Face mask use could result in a large reduction in the risk of infection, with stronger associations with N95 and similar respirators than with disposable surgical masks [15,16].

Despite the benefits of face mask use, studies have reported that face mask use varies across different settings. A study from Japan reported that most respondents showed a moderate or higher frequency of washing their hands or wearing masks (both at 96.4%) [17], while in Saudi Arabia, the majority (91%) reported that they were following the strategies recommended by the authorities to prevent the spread of the virus [18]. In Mizan-Tepi, Ethiopia, the majority (55%) did not wear face masks [19], while in the United Arab Emirates more than 90% did so [20]. Positive correlations were discovered between attitude and practice [21] among primary school students in Wuhan, China (51.6%) [22], in western Uganda (95.2%) [23], and in Dessie and Kombolcha towns in Ethiopia (74.1%) [24].

After the launch of safe school reopening programs in Ethiopia [25], there was high demand for prevention and control efforts to be strengthened. Students are among at-risk groups for contracting COVID-19, which in turn can affect a large number of families and communities. Knowing this will help to set priorities and design effective and consistent preventive measures at school. Thus, this study aimed to assess the level of mask use and associated factors among students at Gambella Teachers' Education and Health Science College, Gambella, Ethiopia, in 2021.

Methods

Study Area and Period

The study was conducted in Gambella Teachers' Education and Health Science College, Gambella Region, Southwest Ethiopia, from March 5 to 30, 2021. This college is located in Gambella Town, 766 km away from Addis Ababa, the capital city of Ethiopia. The region has a total population of 435,999 [26]. After the report of the first COVID-19 case in Ethiopia, the government, in collaboration with responsible bodies, implemented strict COVID-19 prevention and control measures. Despite that effort, the number of infected patients increased because of limited capacity and access to services. Gambella is one of the regional states that were among the settings of the COVID-19 program. As part of this program, a total of 5 hospitals and 148 other health institutions and isolation centers provided services all over the region. The college provides teaching and learning to regular, private, and extension-program students. A total of 1041 regular, 1262 summer, and 2563 extension students were enrolled from the teaching stream and 659 regular and 2256 private students were enrolled from the regular health stream. Of these students, 2763 were men and 1515 were women. The college also provides scholarship programs for students from South Sudan and refugees in partnership with various nongovernmental organizations [27].

Study Design

This was a cross-sectional study using mixed methods.

Source Population and Study Population

The source population included all students in the Gambella Teachers' Education and Health Science College during the study period. However, the study population included only students in selected streams and departments during the study period.

Inclusion and Exclusion Criteria

All active students registered in the second semester of the 2021 academic calendar year or later academic years were included in this study. Students who were seriously ill or absent during the data collection period were excluded from this study.

Sample Size Determination and Sampling

The sample size was determined using a single-population proportion formula with the following assumptions: 95% CI, 5% type I error, and a proportion of respondents whose attitude level was 56.4%. Finally, the researchers added 5% to compensate for the nonresponse of participants, making the final sample size 397 after considering other objectives.

A stratified random sampling technique was used. First, the college was stratified into teachers' education and health science streams based on academic setting. This college has 2 broad streams (health and teaching). Among academic departments, 5 are health departments and about 20 are teachers' streams or departments. Then, the sample was proportionally allocated to size. Finally, a simple random sampling technique was used to enroll 1 study participant from a health stream to every 4 participants from teaching streams. Purposely selected participants were interviewed for the qualitative study.

Data Collection Instrument and Procedure

A pretested and translated version of a self-administered questionnaire was used for data collection. This questionnaire measured sociodemographic characteristics, knowledge of COVID-19 infection and transmission, attitude-related variables, magnitude, and associated factors. Data collection tools were gathered and adapted from previous research [9,17,19,28]. The data were collected by trained and experienced enumerators. Two days' training was provided to the data collectors and supervisors on the study purpose and methodology and on how to conduct and administer the self-administered questionnaire, take consent, keep confidentiality, and respect the rights of the participants. A minimum of a 1-meter distance was kept between interviewers and interviewees. A semistructured interview guide was used for collecting qualitative interviews. The interviewer used taped records and took notes. Interviews lasting a minimum of 30 to 45 minutes were conducted with the study participants. The research was conducted after the lockdown.

Data Processing and Analysis

The collected data were first checked manually for completeness and consistency at the time of data collection and then rechecked at the office by the principal investigator before data entry. Then, the data were entered into EpiData (version 3.1; EpiData Association) and exported to SPSS (version 22; IBM Corp) for analysis.

Descriptive statistics are reported for sociodemographic characteristics and knowledge, attitude, and practice variables as the mean (SD) and range for numerical data and frequency and percentage for categorical data, including the magnitude of mask use.

Then, a logistic regression analysis was performed to determine the strength of associations with the outcome variable using the adjusted odds ratio (AOR) with the 95% CI. The final model fitness was checked using Hosmer-Lemeshow goodness of fit. The qualitative data were transcribed, translated, coded, and analyzed using thematic analysis. Then, the themes were used to triangulate the quantitative study.

Data Quality Management

Data collectors were trained on the data collection tool contents and how to collect them. A pretested questionnaire was used on 5% of respondents from another nearby college. The completeness, accuracy, and consistency of the collected data were checked by the principal investigator. Data were edited, coded, and entered into a computer. Then, computer data cleaning was performed to check for the consistency of data.

Operational Definition

Face masks were defined as disposable or reusable devices that create a physical barrier between the mouth and nose of the wearer and potential contaminants in the immediate environment.

Ethical Approval

Ethical approval was obtained from Gambella Teachers' Education and Health Science College Research Ethics Review Committee (registration number: GTE/939/355/2020). Verbal and written informed consent was obtained for the quantitative and qualitative data, respectively. The informants were assured that all written and recorded data would be kept confidential by using codes to identify participants instead of names or any other personal identifiers. Informants were clearly informed about their right to refuse to participate in the study or withdraw at any time during the interview session.

Results

Sociodemographic Characteristics

The study included total of 397 participants with a response rate of 95.5% (n=379). The majority of study participants were older than 25 years, and the mean age was 26.34 (SD 5.8) years. About 86% (327/379) of the study participants were men and 55.1% (209/379) were Protestant. A total of 261 (68.9%) of the participants earned a monthly income between US $9.29 and US $37.26 (Table 1).

Table 1. Sociodemographic characteristics of students (n=379) at Gambella Teachers' Education and Health Science College, Gambella Region, Southwest Ethiopia.

Variable	Values
Age (years), mean (SD)	26.34 (5.8)
Age group (years), n (%)	
<20	88 (23.2)
20-25	75 (19.8)
>25	216 (57)
Sex, n (%)	
Male	327 (86.3)
Female	52 (13.7)
Marital status, n (%)	
Single	151 (39.8)
Engaged/married	228 (60.2)
Religion, n (%)	
Muslim	27 (7.1)
Orthodox	48 (12.7)
Protestant	209 (55.1)
Catholic or other	72 (19)
Monthly income (US $), n (%)	
Up to 9.30	103 (27.2)
9.31-37.25	261 (68.9)
37.26 or more	15 (4)

Knowledge of COVID-19 Infection and Transmission Among Students

The following sections describe the students' knowledge on COVID-19 infection according to the quantitative data.

Symptoms

A high proportion (360/379, 95%) of the students knew that the main clinical symptoms of COVID-19 are fever, fatigue, dry cough, and myalgia, while 234/379 (61.7%) of the participants mentioned other symptoms, such as stuffy nose, runny nose, and sneezing, which distinguish COVID-19 from the common cold and flu.

Risk Factors and Prognosis

A total of 195 of 379 (51.5%) of the students knew that elderly people who have chronic illnesses and obesity are at higher risk of developing a severe form of COVID-19, whereas 83.1% (316/379) knew that COVID-19 has no effective cure, yet seeking early treatment increases the chance of recovery.

Mode of Transmission

About 83% (316/379) of the students knew that the COVID-19 virus spreads via respiratory droplets from infected people and 77.8% (295/379) of the students knew that asymptomatic transmission is possible.

Knowledge About Prevention

A higher proportion (n=362, 95.5%) of the 379 students knew that proper hand washing with soap and water and wearing of general medical masks by ordinary residents can prevent infection. Similar proportions of participants knew not to touch the eyes or nose with unwashed hands (n=295, 77.8%); that they should avoid going to crowded places, such as train stations and public transportation (n=333, 87.9%); that they should avoid contact with people with a known history of infection (n=317, 83.6%); and that isolation and treatment of people who are infected with COVID-19 are effective ways to prevent the virus (n=311, 82.5%). However, 125 (33%) agreed that children and young adults do not need to take measures to prevent infection with COVID-19. (Table 2)

Table 2. Knowledge of students about mode of transmissions and infectiousness at Gambella Teachers' Education and Health Science College, Gambella Region, Southwest Ethiopia.

Variables (n=379)	Respondents (n=379), n (%)	
	Correct	Incorrect
Knowledge of symptoms		
Main clinical symptoms of COVID-19 are fever, fatigue, dry cough, and myalgia	360 (95)	19 (5)
Unlike the common cold, stuffy nose, runny nose, and sneezing are less common in persons infected with COVID-19	234 (61.7)	145 (38.3)
Knowledge of high risk and prognosis		
Not all persons with COVID-19 will develop severe cases. Only those who are elderly, have chronic illnesses, and are obese are more likely to be severe cases	195 (51.5)	184 (48.5)
There currently is no effective cure for COVID-2019, but early symptomatic and supportive treatment can help most patients recover from the infection	316 (83.1)	61 (16.6)
Knowledge about mode of transmission and infectiousness		
COVID-19 spreads via respiratory droplets of infected individuals	316 (83.1)	61 (16.6)
Eating or contacting wild animals would result in the infection by COVID-19	255 (67.7)	124 (32.7)
Persons with COVID-19 cannot infect the virus to others when a fever is not present	122 (32.2)	257 (67.8)
Knowledge about ways of prevention		
Proper hand washing with soap and water is one method of preventing COVID-19	362 (95.5)	17 (4.5)
One way of preventing COVID 19 is not touching the eyes or nose with unwashed hands	295 (77.8)	84 (22.2)
To prevent infection with COVID-19, individuals should avoid going to crowded places, such as train stations, and avoid taking public transportation	333 (87.9)	46 (12.1)
Ordinary residents can wear general medical masks to prevent infection by COVID-19	362 (95.5)	17 (4.5)
People who have contact with someone infected with COVID-19 should be immediately isolated in a proper place	317 (83.6)	62 (16.4)
Isolation and treatment of people who are infected with COVID-19 are effective ways to reduce the spread of the virus	311 (82.5)	68 (17.9)
Children and young adults do not need to take measures to prevent COVID-19	125 (33)	254 (67)

Attitude-Related Variables of Students

Regarding the attitude of participants, 90.8% (344/379) of the students agreed that COVID-19 symptoms appear in 2 to 14 days, 88.7% (336/379) of the participants agreed COVID-19 is fatal, and 67.3% (255/379) of students knew that sick patients should share their recent travel history with health care providers. A total of 227 (59.9%) participants agreed that flu vaccination is sufficient, while 320 (84.4%) agreed that eating well-cooked food can sufficiently prevent COVID-19. About 70% (266/379, 70.2%) of students had a positive attitude toward disinfecting areas at least once a day and 98.4% (373/379) had a positive attitude toward washing hands to help prevent COVID-19 transmission (Table 3).

Table 3. Attitude-related variables of students and associated factors at Gambella Teachers' Education and Health Science College, Gambella Region, Southwest Ethiopia.

Attitude-related variables	Respondents (n=379), n (%)	
	Yes	No
COVID-19 symptoms appear in 2 to 14 days	344 (90.8)	35 (9.2)
COVID-19 is fatal	336 (88.7)	43 (11.3)
Flu vaccination is sufficient for COVID-19	227 (59.9)	152 (40.1)
Eating well cooked food and safely handled meat is safe	320 (84.4)	59 (15.6)
Sick patients should share their recent travel history with health care providers	255 (67.3)	124 (32.7)
Disinfect areas at least once a day	266 (70.2)	113 (29.8)
Washing hands helps prevent COVID-19 transmission	373 (98.4)	6 (1.6)

Magnitude of Mask Use Among Students

The magnitude of mask use among students was found to be 87% (330/379). The qualitative findings included the following observation: “about half of the students reported that mask use prevents COVID-19 infection; however, they failed to apply the regulation regularly. For this, students reported that lack of access to masks was a challenge during their stay in the college.”

Factors Associated With Mask Use Among Students

After selecting candidates in a bivariate analysis at *P*≤.25, the following variables were independently associated with mask use in the multivariate analysis using AOR and *P*≤.05: sex, age, agreement that not all persons with COVID-19 develop severe disease, agreement that there is currently no effective cure, adherence to proper washing with soap and water, other variables (rubbing with alcohol and avoiding contact with surfaces), having started to stay home, agreement that COVID-19 is fatal, and agreement that flu vaccination and eating well-cooked food are sufficient to prevent COVID-19.

The odds of mask use among female students were 3.3 times higher than among male students (AOR 3.32, 95% CI 1.191-9.248). Younger students was 2.55 times more likely to use face masks than older students (AOR 2.55, 95% CI 1.155-5.627) and students who agreed that not all persons with COVID-19 develop severe disease were 3.38 times more likely to wear a mask than those who did not (AOR 3.38, 95% CI 1.36-8.41). Students who responded that there was currently no effective cure were 6.28 times more likely to wear a mask than those who did not (AOR 6.28, 95% CI 1.36-28.99).

The qualitative interviews led to the following observation: “the majority of respondents reported that COVID-19 is a serious problem that affects all age groups regardless of their sex, race, level of income and educational status. No cure was affirmed globally; as a result, emphasis should be given to prevention and control.”

Students who reported that they performed proper washing with soap and water were 0.03 times as likely to wear a mask (AOR 0.027, 95% CI 0.004-0.182). Despite the cost-effectiveness of frequent hand washing, the qualitative report included the following observation: “40% of students reported that lack of access to filled water containers at every corner regularly has been a challenge at the school.”

In addition, a 21-year-old male participant reported in an interview that “along with mask use, students used to protect themselves using other alternative mechanisms too.”

Students who had started to stay home were 0.168 times as likely to wear a mask (AOR 0.168, 95% CI 0.054-0.52) and those who agreed that COVID-19 is fatal were 0.236 times as likely (AOR 0.236, 95% CI 0.084-0.666). Students who agreed that the flu vaccine is sufficient for preventing COVID-19 were 3.9 times more likely to wear a mask than those who did not (AOR 3.874, 95% CI 1.540-9.749), and those who disinfected equipment and working areas in market areas at least once a day were 0.22 times as likely compared to those who did not (AOR 0.222, 95% CI 0.086-0.575; Table 4).

Table 4. Factors associated with mask use among students and associated factors at Gambella Teachers' Education and Health Science College, Gambella Region, Southwest Ethiopia.

Variables and categories	COR[a] (95% CI)	AOR[b] (95% CI)	*P* value[c]
Sex			
Male	1	1	
Female	3.07 (1.52-6.23)[d]	3.31 (1.19-9.24)	.02
Age category (years)			
<20	2.01 (1.05-3.89)[d]	2.55 (1.15-5.62)	.009
20-25	0.41 (0.34-1.22)	0.27 (0.05-1.35)	.02
>25	1	1	
Agreement that not all persons with COVID-19 develop severe disease			
Yes	2.67 (1.38-5.12)[d]	3.37 (1.35-8.40)	.009
No	1	1	
Agreement that there is currently no effective cure			
Yes	3.42 (1.03-11.32)[d]	6.27 (1.35-28.98)	.02
No	1	1	
Perform proper washing with soap and water			
Yes	0.25 (0.09-0.70)[d]	0.02 (.004-0.18)	<.001
No	1	1	
Have started to stay home			
Yes	0.31 (0.14-0.69)[d]	0.16 (0.05-0.52)	.002
No	1	1	
Agreement that COVID-19 is fatal			
Yes	0.24 (0.12-0.49)[d]	0.23 (0.08-0.66)	.006
No	1	1	
Agreement that flu vaccine is sufficient for preventing COVID-19			
Yes	1	1	
No	2.69 (1.45-4.99)[d]	3.87 (1.54-9.74)	.004
Disinfect equipment and working areas at least once a day			
Yes	0.57 (0.31-1.06)	0.22 (0.08-0.57)	.002
No	1	1	

[a]COR: crude odds ratio.

[b]AOR: adjusted odds ratio.

[c]*P*≤.05 was considered statistically significant.

[d]Candidates selected at *P*<.25.

Discussion

Principal Findings

This study investigated face mask use and associated factors among students at Gambella Teachers' Education and Health Science College, Southwest Ethiopia, in 2021. The overall magnitude of face mask use was found to be 87% (330/379). Factors associated with mask use included being female and younger, agreeing that not all persons with COVID-19 develop severe disease, agreeing that there is currently no effective cure, performing proper washing with soap and water, having started to stay home, agreeing that COVID-19 is fatal, agreeing that flu vaccination is sufficient for preventing COVID-19, and disinfecting equipment and working areas at least once a day.

This study, conducted in Ethiopia, found that the magnitude of mask use among students was 87%, which is relatively lower than what was found by a study conducted in Japan; most respondents in that study showed a moderate or higher frequency of washing their hands or wearing a mask (both 96.4%) [17]; other studies found mask use magnitudes of 91% in Saudi

Arabia [18], 95.2% in western Uganda [23], 90% in the United Arab Emirates [20], and 87.94% in China (among undergraduate students) [21]. However, our finding of an 87% use rate is higher than some other studies: 45% in Mizan Tepi, Ethiopia [19], 74.1% in the towns of Dessie and Kombolcha in Ethiopia [24], and 51.6% in Wuhan, China [22]. The differences in the magnitude of mask use might be due to differences in the level of awareness of respondents, strictness of adherence to recommendations and regulations, access to services, and differences in services.

The odds of mask use among female students were 3.3 times higher than among male students. This finding is consistent with studies conducted among Japanese university students and in reopened secondary schools in Gonder, Northern Ethiopia [17,19]. Similarly, a study conducted in China revealed that women showed significantly higher levels of positive attitudes than men; moreover, there was a positive correlation between attitude and practice [19,21,29].

This study found that younger students were 2.55 times more likely to use face masks than older students. A study from the Amhara region of Ethiopia found that younger age was associated with better knowledge of COVID-19 [30], which, in turn, could be justified as being the result of those with good knowledge of COVID-19 being more likely to practice prevention measures [19].

Students who understood that not all persons with COVID-19 develop severe disease were about 3.4 times more likely to use masks than those who did not understand this. This seems to contradict previous literature showing that people with comorbid illness or other chronic diseases, such as diabetes mellitus and hypertension, and who are older were more likely to develop severe disease [31,32]. While it is true that not all persons will reach a severe stage of the disease, it has continued to be recommended that the public use face masks in specific situations with a high risk of infection, such as crowded places, irrespective of the local epidemiological situation.

This study found that students who agreed that there is currently no effective cure for COVID-19 were 6.3 times more likely to use masks than those who said they believed that there is an effective cure. Despite tremendous efforts to improve vaccine availability, access, and quality to control COVID-19, the World Health Organization (WHO) continues to recommend the use of masks by the public in specific situations and updated this advice to recommend their use irrespective of the local epidemiological situation, given the current spread of COVID-19 globally. Masks also reduce disease severity at schools, where there is a great deal of social gathering and interaction [5,11].

The literature overwhelmingly recommends proper washing with soap and water along with other behavior to help prevent COVID-19 infection [22]. Despite this, our study found that respondents who washed frequently were 0.03 times as likely to use a mask. It could be the case that it is an optional behavior for students to wash frequently, whereas using face masks is not optional. Meanwhile, studies have reported that some students do not use masks due to financial and affordability issues across different settings [12].

Students who had started to stay home were 87% less likely to use face masks than who did not stay home. The reason for this could be that masks are used in places where people move and interact, which can expose them to infection. On the other hand, those who stay home usually practice recommended prevention methods other than mask use.

Students who agreed that COVID-19 is fatal were about 74% less likely to use masks than those who did not agree. A study from Debreberhan University found that students perceived low risks related to the COVID-19 pandemic at the time of school resumption [33]. Despite the low perception of the deadliness of COVID-19, the students might have been exposed to other preventive measures, like hand washing and social distancing [15]. Meanwhile, a Polish study among health care workers suffering from face mask–induced itchiness found that they were 31.6% less likely to comply with mask use [34], while another study found that most students supported the required use of masks in schools and indoor public spaces [35].

The odds of mask use among respondents who agreed that flu vaccination was sufficient for preventing COVID-19 were 3.9 times greater than among those who did not agree. This contrasts with a study from Saudi Arabia that found that 90.4% of students avoided contact with other people if they had flu-like symptoms [20]. On the other hand, about 50% of students rightly stated that antibiotics and flu vaccines are not effective against COVID-19 infection [36]. A recent study conducted by Kinyili et al [37] showed that an increased extent of mask-wearing among vaccinated individuals increased with an increasing level of vaccination and that regular face mask use resulted in a sharp decrease in COVID-19 infections. However, wearing face masks alone also resulted in a reduction in the peak number of infections with an increasing level of face mask efficacy, although there can be a delay as the infections are cleared; that is, treatment of COVID-19 does not always lead to immediate recovery, as mild to moderate symptoms persist [38].

The WHO recommends cleaning and disinfecting bathroom and toilet surfaces at least once daily using regular household soap or detergent, and then, after rinsing, applying regular household disinfectant [39]. This agrees with our finding that respondents who disinfected their equipment and working areas at least once a day were 78% less likely to use masks. Students who were effective in maintaining washing, disinfecting, and applying social distancing might not have used masks. Similarly, about 80% (16/20) of students reported that lack of access to masks and affordability of masks were restraints on their regular use after schools were reopened.

Limitations

We made extensive efforts to reduce possible shortcomings of this study. Despite that, this study has certain limitations. First, the study was limited to students at a government college; therefore, there is a question of the representativeness of the findings to all students. In addition, the data presented in this study are self-reported and may thus be subject to recall bias. Furthermore, the study has a limited ability to determine cause-effect relationships because of its cross-sectional design.

Conclusion

The overall magnitude of mask use of 87% (330/379) shows that mask use is relatively moderate in Ethiopia. Several factors, such as sex, age, having started to stay home, agreeing that not all persons with COVID-19 develop severe disease, agreeing that there is currently no effective cure, agreeing that proper washing with soap and water is effective, agreeing that COVID-19 is fatal, and agreeing that the flu vaccine is sufficient to prevent COVID-19 were independently associated with mask use among students. Therefore, the college should aggressively encourage students to wear masks, monitor the implementation of COVID-19 prevention regulations, and improve the accessibility of masks.

Acknowledgments

The study participants are greatly acknowledged for the information they provided. Data collectors and the Gambella Teachers Education and Health Science College are also acknowledged for their kind response. This study was funded by the Gambella Teachers' Education and Health Science College. The views presented in the article are of the authors and do not necessarily express the views of the funding organization. All authors declared that they had insufficient or no funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this article.

Data Availability

The data sets used and/or analyzed during the current study are all presented within the manuscript.

Authors' Contributions

AAG carried out the statistical analysis. TDT wrote the manuscript. Both authors made contributions to the interpretation of results and revised the manuscript for important intellectual content. Both authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Clinical management of severe acute respiratory infection (SARI) when COVID-19 disease is suspected: interim guidance, 13 March 2020. World Health Organization. URL: https://apps.who.int/iris/handle/10665/331446 [accessed 2023-05-04]
2. Epidemiology Working Group for NCIP Epidemic Response, Chinese Center for Disease Control and Prevention. [The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China]. Zhonghua Liu Xing Bing Xue Za Zhi 2020 Feb 10;41(2):145-151. [doi: 10.3760/cma.j.issn.0254-6450.2020.02.003] [Medline: 32064853]
3. Critical preparedness, readiness and response actions for COVID-19: interim guidance. World Health Organization. URL: https://apps.who.int/iris/bitstream/handle/10665/331422/WHO-COVID-19-Community_Actions-2020.1-eng.pdf [accessed 2023-05-04]
4. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/index.html [accessed 2023-05-04]
5. World Health Organization, UNICEF, IFRC. Key Messages and Actions for COVID-19 Prevention and Control in Schools. 2020 Mar. URL: https://tinyurl.com/4357m657
6. WHO Coronavirus (COVID-19) Dashboard. World Health Organization. URL: https://COVID19.who.int/ [accessed 2023-05-04]
7. Africa CDC. African Union. URL: https://au.int/africacdc [accessed 2023-05-04]
8. Worldometer. Worldometer. URL: https://www.worldometers.info/ [accessed 2023-05-04]
9. Kebede Y, Yitayih Y, Birhanu Z, Mekonen S, Ambelu A. Knowledge, perceptions and preventive practices towards COVID-19 early in the outbreak among Jimma university medical center visitors, Southwest Ethiopia. PLoS One 2020;15(5):e0233744 [FREE Full text] [doi: 10.1371/journal.pone.0233744] [Medline: 32437432]
10. Awadasseid A, Wu Y, Tanaka Y, Zhang W. Initial success in the identification and management of the coronavirus disease 2019 (COVID-19) indicates human-to-human transmission in Wuhan, China. Int J Biol Sci 2020;16(11):1846-1860 [FREE Full text] [doi: 10.7150/ijbs.45018] [Medline: 32398954]
11. Responding to community spread of COVID-19: interim guidance. World Health Organization. URL: https://apps.who.int/iris/handle/10665/331421 [accessed 2023-05-04]
12. Tucho GT, Kumsa DM. Universal use of face masks and related challenges during COVID-19 in developing countries. Risk Manag Healthc Policy 2021;14:511-517 [FREE Full text] [doi: 10.2147/RMHP.S298687] [Medline: 33603517]

13. Sommerstein R, Fux CA, Vuichard-Gysin D, Abbas M, Marschall J, Balmelli C, Swissnoso. Risk of SARS-CoV-2 transmission by aerosols, the rational use of masks, and protection of healthcare workers from COVID-19. Antimicrob Resist Infect Control 2020 Jul 06;9(1):100 [FREE Full text] [doi: 10.1186/s13756-020-00763-0] [Medline: 32631450]
14. Panovska-Griffiths J, Kerr CC, Waites W, Stuart RM, Mistry D, Foster D, et al. Modelling the potential impact of mask use in schools and society on COVID-19 control in the UK. Sci Rep 2021 Apr 22;11(1):8747 [FREE Full text] [doi: 10.1038/s41598-021-88075-0] [Medline: 33888818]
15. Chu DK, Akl EA, Duda S, Solo K, Yaacoub S, Schünemann HJ, COVID-19 Systematic Urgent Review Group Effort (SURGE) study authors. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. Lancet 2020 Jun 27;395(10242):1973-1987 [FREE Full text] [doi: 10.1016/S0140-6736(20)31142-9] [Medline: 32497510]
16. Matuschek C, Moll F, Fangerau H, Fischer JC, Zänker K, van Griensven M, et al. Face masks: benefits and risks during the COVID-19 crisis. Eur J Med Res 2020 Aug 12;25(1):32 [FREE Full text] [doi: 10.1186/s40001-020-00430-5] [Medline: 32787926]
17. Hatabu A, Mao X, Zhou Y, Kawashita N, Wen Z, Ueda M, et al. Knowledge, attitudes, and practices toward COVID-19 among university students in Japan and associated factors: An online cross-sectional survey. PLoS One 2020;15(12):e0244350 [FREE Full text] [doi: 10.1371/journal.pone.0244350] [Medline: 33347488]
18. Alrasheedy AA, Abdulsalim S, Farooqui M, Alsahali S, Godman B. Knowledge, attitude and practice about coronavirus disease (COVID-19) pandemic and its psychological impact on students and their studies: a cross-sectional study among pharmacy students in Saudi Arabia. Risk Manag Healthc Policy 2021;14:729-741 [FREE Full text] [doi: 10.2147/RMHP.S292354] [Medline: 33654441]
19. Angelo AT, Alemayehu DS, Dacho AM. Knowledge, attitudes, and practices toward COVID-19 and associated factors among university students in Mizan Tepi University, 2020. Infect Drug Resist 2021;14:349-360 [FREE Full text] [doi: 10.2147/IDR.S299576] [Medline: 33564243]
20. Hasan H, Raigangar V, Osaili T, Neinavaei NE, Olaimat AN, Aolymat I. A cross-sectional study on university students' knowledge, attitudes, and practices toward COVID-19 in the United Arab Emirates. Am J Trop Med Hyg 2021 Jan;104(1):75-84 [FREE Full text] [doi: 10.4269/ajtmh.20-0857] [Medline: 33236710]
21. Peng Y, Pei C, Zheng Y, Wang J, Zhang K, Zheng Z, et al. A cross-sectional survey of knowledge, attitude and practice associated with COVID-19 among undergraduate students in China. BMC Public Health 2020 Aug 26;20(1):1292 [FREE Full text] [doi: 10.1186/s12889-020-09392-z] [Medline: 32847554]
22. Chen X, Ran L, Liu Q, Hu Q, Du X, Tan X. Hand hygiene, mask-wearing behaviors and its associated factors during the COVID-19 epidemic: a cross-sectional study among primary school students in Wuhan, China. Int J Environ Res Public Health 2020 Apr 22;17(8):2893 [FREE Full text] [doi: 10.3390/ijerph17082893] [Medline: 32331344]
23. Sikakulya FK, Ssebuufu R, Mambo SB, Pius T, Kabanyoro A, Kamahoro E, et al. Use of face masks to limit the spread of the COVID-19 among western Ugandans: Knowledge, attitude and practices. PLoS One 2021;16(3):e0248706 [FREE Full text] [doi: 10.1371/journal.pone.0248706] [Medline: 33760882]
24. Natnael T, Alemnew Y, Berihun G, Abebe M, Andualem A, Ademe S, et al. Facemask wearing to prevent COVID-19 transmission and associated factors among taxi drivers in Dessie City and Kombolcha Town, Ethiopia. PLoS One 2021;16(3):e0247954 [FREE Full text] [doi: 10.1371/journal.pone.0247954] [Medline: 33711038]
25. Honein MA, Barrios LC, Brooks JT. Data and policy to guide opening schools safely to limit the spread of SARS-CoV-2 infection. JAMA 2021 Mar 02;325(9):823-824 [FREE Full text] [doi: 10.1001/jama.2021.0374] [Medline: 33497433]
26. Federal Democratic Republic of Ethiopia CSA. Population Projection of Ethiopia for All Regions At Wereda Level from 2014-2017. Archived from the original on 6 June. Retrieved 4 June 2018:1-118 [FREE Full text]
27. Gambell Teachers Education and Health Science College A. Ten year strategic plan 2020-2030. June 2021. Gambella Southwest Ethiopia 2021 Jun 15;1:1-76.
28. Aynalem YA, Akalu TY, Gebresellassie Gebregiorgis B, Sharew NT, Assefa HK, Shiferaw WS. Assessment of undergraduate student knowledge, attitude, and practices towards COVID-19 in Debre Berhan University, Ethiopia. PLoS One 2021;16(5):e0250444 [FREE Full text] [doi: 10.1371/journal.pone.0250444] [Medline: 34003825]
29. Handebo S, Adugna A, Kassie A, Shitu K. Determinants of COVID-19-related knowledge and preventive behaviours among students in reopened secondary schools: cross-sectional study. BMJ Open 2021 Apr 24;11(4):e050189 [FREE Full text] [doi: 10.1136/bmjopen-2021-050189] [Medline: 33895723]
30. Tadesse AW, Abebe NM, Tadesse SE, Wube MC, Abate AA. Preventive practice and associated factors towards COVID-19 among college students in Amhara Region, Ethiopia: A cross-sectional study. Ethiop J Health Sci 2021 Jan;31(1):3-14 [FREE Full text] [doi: 10.4314/ejhs.v31i1.2] [Medline: 34158747]
31. Hu L, Chen S, Fu Y, Gao Z, Long H, Ren H, et al. Risk factors associated with clinical outcomes in 323 coronavirus disease 2019 (COVID-19) hospitalized patients in Wuhan, China. Clin Infect Dis 2020 Nov 19;71(16):2089-2098 [FREE Full text] [doi: 10.1093/cid/ciaa539] [Medline: 32361738]
32. Petrilli CM, Jones SA, Yang J, Rajagopalan H, O'Donnell L, Chernyak Y, et al. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. BMJ 2020 May 22;369:m1966 [FREE Full text] [doi: 10.1136/bmj.m1966] [Medline: 32444366]

33. World Health Organization. Home care for patients with suspected novel coronavirus (COVID-19) infection presenting with mild symptoms, and management of their contacts. Interim guidance 04 February 2020 Aug 12:1-9 [FREE Full text]
34. Tadese M, Haile AB, Moltot T, Silesh M. Perceived risk of COVID-19 and related factors among university students in Ethiopia during school reopening. Infect Drug Resist 2021;14:953-961 [FREE Full text] [doi: 10.2147/IDR.S302126] [Medline: 33737817]
35. Patte KA, Wade TJ, MacNeil AJ, Bélanger RE, Duncan MJ, Riazi N, et al. Support for mask use as a COVID-19 public health measure among a large sample of Canadian secondary school students. BMC Public Health 2022 Aug 22;22(1):1598 [FREE Full text] [doi: 10.1186/s12889-022-14011-0] [Medline: 35996138]
36. Reszke R, Matusiak, Krajewski PK, Szepietowska M, Białynicki-Birula R, Szepietowski JC. The utilization of protective face masks among Polish healthcare workers during COVID-19 pandemic: do we pass the exam? Int J Environ Res Public Health 2021 Jan 19;18(2):841 [FREE Full text] [doi: 10.3390/ijerph18020841] [Medline: 33478170]
37. Kinyili M, Munyakazi JB, Mukhtar AYA. To use face masks or not after COVID-19 vaccination? An impact analysis using mathematical modeling. Front Appl Math Stat 2022 Apr 27;8:872284 [FREE Full text] [doi: 10.3389/fams.2022.872284]
38. Gohel KH, Patel PB, Shah PM, Patel JR, Pandit N, Raut A. Knowledge and perceptions about COVID-19 among the medical and allied health science students in India: An online cross-sectional survey. Clin Epidemiol Glob Health 2021;9:104-109 [FREE Full text] [doi: 10.1016/j.cegh.2020.07.008] [Medline: 32838066]
39. Cho FN, Ngah YE, Tassang AN, Fru CN, Kuku Elad PC, Jokwi PK, et al. Face mask ownership/utilisation and COVID-19 vaccine hesitancy amongst patients recovering from COVID-19 in Cameroon: A cross-sectional study. PLoS One 2023;18(1):e0280269 [FREE Full text] [doi: 10.1371/journal.pone.0280269] [Medline: 36662746]

Abbreviations

AOR: adjusted odds ratio
WHO: World Health Organization

Edited by A Mavragani; submitted 23.07.22; peer-reviewed by M Mulaudzi, S Yoon; comments to author 27.03.23; revised version received 15.04.23; accepted 17.04.23; published 22.05.23.

Please cite as:
Addis Gesese A, Duer Thot T
Face Mask Use and Associated Factors Among Students: Mixed Methods Study
Interact J Med Res 2023;12:e41365
URL: https://www.i-jmr.org/2023/1/e41365
doi:10.2196/41365
PMID:37130014

Original Paper

The Relationship Between Face Mask Use and Face-Touching Frequency in Public Areas: Naturalistic Study

Sydney Niesen[1], BSc; Daniel Ramon[2], BSc, MPH; Rhonda Spencer-Hwang[3], MPH, DrPH; Ryan Sinclair[3], MPH, PhD

[1]San Diego State University, San Diego, CA, United States

[2]Loma Linda University, Loma Linda, CA, United States

[3]Loma Linda University School of Public Health, Loma Linda, CA, United States

Corresponding Author:
Sydney Niesen, BSc
San Diego State University
5500 Campanile Dr
San Diego, CA, 92182
United States
Phone: 1 6195945200
Email: sydney.niesen@gmail.com

Abstract

Background: Throughout the COVID-19 pandemic in the United States, a major public health goal has been reducing the spread of the virus, with particular emphasis on reducing transmission from person to person. Frequent face touching can transmit viral particles from one infected person and subsequently infect others in a public area. This raises an important concern about the use of face masks and their relationship with face-touching behaviors. One concern discussed during the pandemic is that wearing a mask, and different types of masks, could increase face touching because there is a need to remove the mask to smoke, drink, eat, etc. To date, there have been few studies that have assessed this relationship between mask wearing and the frequency of face touching relative to face-touching behaviors.

Objective: This study aimed to compare the frequency of face touching in people wearing a mask versus not wearing a mask in high–foot traffic urban outdoor areas. The purpose of this study was to assess if mask wearing was associated with increased face touching.

Methods: Public webcam videos from 4 different cities in New York, New Jersey, Louisiana, and Florida were used to collect data. Face touches were recorded as pedestrians passed under the webcam. Adult pedestrians wearing masks were compared to those not wearing masks. Quantitative measures of frequency, duration, site of touch, and oral activities were recorded. Linear regression analysis was used to assess the association between mask use and face touching.

Results: Of the 490 observed subjects, 241 (49.2%) were wearing a mask properly and 249 (50.8%) were not. In the unmasked group, 33.7% (84/249) were wearing it improperly, covering the mouth only. Face touching occurred in 11.4% (56/490) of the masked group and 17.6% (88/490) in the unmasked group. Of those who touched their face, 61.1% (88/144) of people were not wearing a mask. The most common site of face touching was the perioral region in both groups. Both the masked and unmasked group had a frequency of face touching for 0.03 touches/s. Oral activities such as eating or smoking increased face touching in the unmasked group.

Conclusions: Contrary to expectations, non–mask-wearing subjects touched their face more frequently than those who were wearing a mask. This finding is substantial because wearing a face mask had a negative association with face touching. When wearing a mask, individuals are less likely to be spreading and ingesting viral particles. Therefore, wearing a mask is more effective in preventing the spread of viral particles.

(Interact J Med Res 2023;12:e43308) doi:10.2196/43308

KEYWORDS

COVID-19; mask wearing; face-touching; self-inoculation; public health; digital surveillance; webcam video; prevention; health risk; health; risk; mask; surveillance; transmission; behavior

Introduction

Over the past 2 and a half years, the world has been throttled with a massive pandemic. In August 2020, the COVID-19 disease had caused more than 29,880 deaths in the United States [1]. At some point in time during the pandemic, many county jurisdictions had mandates in place that required face masks to be worn in public outdoor areas to reduce the transmission of respiratory viruses. This recommendation was based on evidence from the outbreak of severe acute respiratory syndrome (SARS) in 2003 followed by the hemagglutinin type 5 neuraminidases type 1 (H5N1) and hemagglutinin type 1 neuraminidases type 1 (H1N1) influenza outbreaks [2]. During these previous outbreaks, it was found that people who are infected with respiratory viruses have the potential to transmit viruses through respiratory secretions that become airborne or adhere to public surfaces [3]. This ultimately justified the mandated use of face masks and the recommendation of other infection prevention practices (eg, frequent handwashing to reduce the spread of respiratory secretions). Along with these recommendations, hand to face contact is another important behavioral factor to control the spread of infectious disease. An exposure assessment is an ideal tool to measure the use of masks, handwashing, and the interaction where face masks impact the frequency of face touching.

Some hypotheses proposed by popular media early in the pandemic claimed that wearing a face mask can heighten facial awareness and sensitivity, prompting an increase in face touching [4]. Such activity can work against the barrier concept of mask wearing. Other concerns addressed a fear of breathing difficulties, constitutional rights being taken away, and hygiene concerns [5]. The idea that an increase in face touching occurs when wearing a mask was introduced for a short time by the US Surgeon General at the beginning of the pandemic [6].

Microbial transport from hand to face has been evaluated in several microbial risk assessments and used to advocate for better handwashing practices in clinical, public, and private environments [7]. Recent studies have investigated face-touching behaviors because they are a known risk for disease transmission. One study suggests that wearing a face mask is associated with decreased face touching, thereby enhancing the protection barrier for which the masks were originally designed [8]. Another study comparing face touching before and during the pandemic found that the frequency of face touching decreased as mask mandates were being implemented [9].

Despite the current published studies, there needs to be an investigation of face-touching behavior at highly frequented outdoor public sites where different activities of human behavior can be naturally observed. For this reason, our study is needed to evaluate how daily human activities affect human behavior and the frequency of face touching. We tested the hypothesis that wearing a face mask will increase face touching while engaged in different activities such as eating, drinking, or smoking.

Methods

Study Overview

This study used a video-based, naturalistic, and observational approach to assess the relationship between face mask use and face-touching behaviors of people in public spaces. Public webcams from EarthCam [10] were used to conduct real-time observations in New York, Louisiana, Florida, and New Jersey. Data were recorded on different face-touching behaviors in high-traffic public locations. This study was designed to be empirically focused and methodologically quantitative.

Recruitment

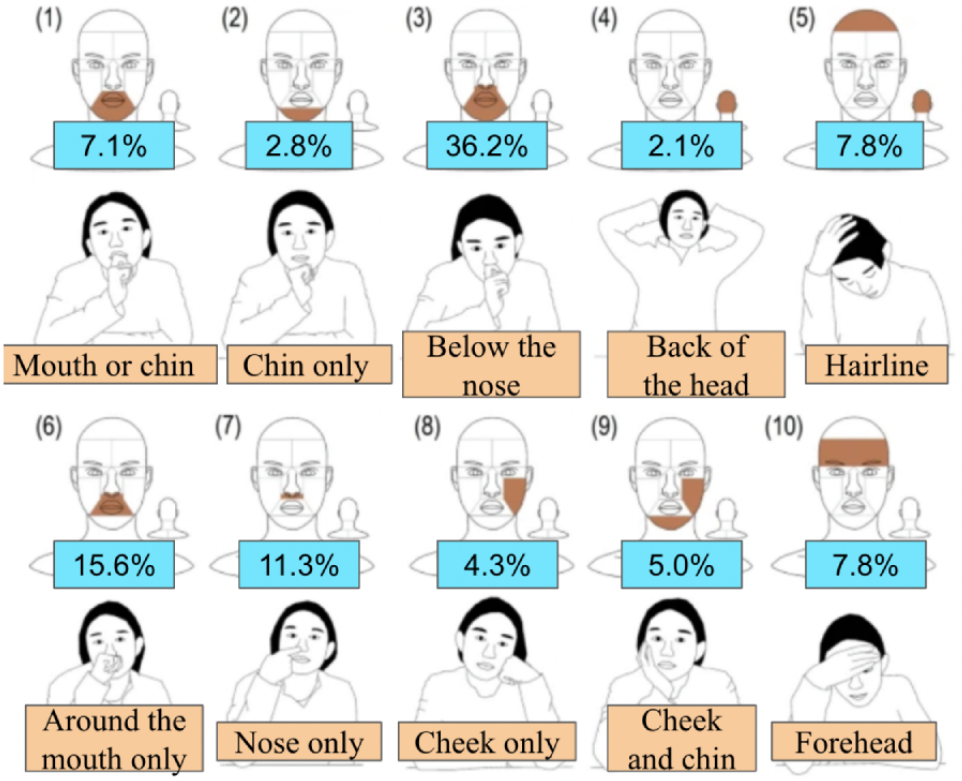
The target population included people from different ages but were categorized as either adults or children. Those viewed as 16 years of age or older were considered adults, and those younger were classified as children. Individuals were also categorized in the observation as being with or without a mask and whether or not they were observed eating, drinking, or smoking. For individuals to be included, the lighting had to be clear enough to differentiate their hands and face.

EarthCam was used to remotely view and record real-time footage of popular locations in the United States. The locations chosen for this study were selected based on several important factors that included a high density of people, video resolution, and the proximity of the camera to the people to be able to view face masks and behaviors of interest to this study. Five locations were selected based on this criteria: North and South Seaside Heights, New Jersey; Bourbon Street in New Orleans, Louisiana; Times Square in New York City, New York; and Key West, Florida.

Observation Instrument

A survey form using Google Forms was used to structure data entry on subject observations from recorded videos available on EarthCam. These videos were recorded using the screen capture function of a multimedia file player (QuickTime Player, Apple) and then uploaded to a file folder in Google Drive for collaboration. Each file was limited to 3 minutes to conserve file size and allow for systematic review. Recordings were made from October 1, 2020, to November 21, 2020, consistently between 3 and 5 PM (PST) for Tuesdays and Thursdays and between 8 and 10 AM (PST) for Saturdays and Sundays. The structured survey form consisted of observed demographics that included gender appearance, categorized as heteronormative male or female, and age, distinguished as child or adult. The time and date of the recorded observation, the location of the observation, and the duration of the walk was based on when the subject entered and exited the frame. The form also included the number of face touches and the duration of each touch for up to 3 consecutive touches. Each touch was then classified based on where the person touched their face (Figure 1) and if any of those touches were done while eating, drinking, vaping, or smoking. Lastly, the form recorded several types of masks worn and the overall mask-wearing style, classified as correct when it was covering the nose and mouth or incorrect when it covered the nose or mouth only or when worn on the chin or other areas.

Figure 1. Face touch locations for 10 regions showing the percentage of study participants observed to contact that region. Image adapted from Zhang et al [11], which is published under Creative Commons Attribution 4.0 International License [12].



Video Data Review and Record

Observations were made on the web via live public webcams posted on the EarthCam website at various locations. Individuals were selected for inclusion in the study if their hands and face were clearly visible in the frame for at least 5 seconds and not more than 300 seconds. Subjects were included in the observation starting from when they entered the frame to when they exited the frame or when their hands and face became unclear. The video files were then renamed with a predefined format and transferred to a Google Drive account for storage.

Observations were recorded using Google Forms to include time and date, location, age, the duration of the walk, the number of face touches, the duration of each touch up to 3 touches, the part of face touched, oral activity, mask-wearing style, the type of mask, and whether or not they were by themselves or with others. The longest observations were often made in Time Square, New York, because people were observed with a wide field of view sitting in a public courtyard. The shortest observations were in Key West, Florida, due to the placement of the camera by a narrow sidewalk with no benches. Data are available in Multimedia Appendix 1.

Ethical Considerations

The Loma Linda University institutional review board (IRB) determined that this research does not meet the definitions of human subject research and does not need or require IRB review or approval. The IRB listed 3 reasons. First, it does not obtain or receive private individually identifiable information. Second, there are no data or specimens collected specifically for use in this study. Third, the study does not have direct intervention or interaction with study subjects. The notice of determination from the Loma Linda University IRB was given the number 5210315.

Statistical Analysis

Data Analysis

We used SPSS (version 27; IBM Corp) statistical software to organize data and report frequencies. Within SPSS, we assessed our hypothesis questions by evaluating the asymptotic significance from Pearson chi-square test and by evaluating the unstandardized β with its asymptotic significance using multiple linear regression, seen in Table 1.

To determine the ideal sample size, we used R statistical software with the *pwr* package (R Foundation for Statistical Computing) [13] to analyze the minimum observations that would detect mask wearing with a medium effect size of 0.5, a significance level of .01, and a power of 90%. To observe the difference in mask-wearing practices, we needed 350 participants for this power and sample size. We aimed to collect beyond that number to allow a buffer for anticipated stratification in bivariate and regression analyses.

Table 1. Multiple linear regression for the frequency of face touching in relation to the type of face mask among mask wearers (n=434)[a].

Linear regression model	Unstandardized β	Coefficients SE	Standardized coefficients β	*t* test (*df=1*)	*P* value
Frequency of touching face (constant)	0.38	0.266	N/A[b]	1.428	.15
Washable mask	0.132	0.236	0.029	0.558	.58
Surgical mask	–0.192	0.273	–0.036	–0.702	.48
N95 mask	–0.374	0.74	–0.023	–0.506	.61
Neck gaiter	–0.3	0.646	–0.021	–0.464	.64
Other type of mask	–0.516	1.535	–0.015	–0.336	.74
Oral activity	1.251	0.376	0.157	3.325	.001
With others or by themselves (no or yes)	0.273	0.257	0.05	1.06	.29
Gender	–0.207	0.207	–0.047	–0.999	.32

[a]Dependent variable: the frequency of face touches with “zero face touching” as the reference group.

[b]N/A: not applicable.

Data Exclusion

Individuals where the sun or a streetlight washed out their face or hands were excluded. Additionally, if the individual was in the frame for less than 5 seconds, they were excluded because it was considered an insufficient amount of time for this study.

Results

Selected Population

We selected 490 individuals from August to November 2020 who met the study criteria. Over 65 hours of video were reviewed involving subject observations at 4 different United States locations including New York, New York (n=283, 57.8%); Seaside Heights, New Jersey (n=14, 2.9%); Key West, Florida (n=18, 3.7%); and New Orleans, Louisiana (n=175, 35.7%). Table 2 represents the demographics of all the subjects involved in the study with a total of 274 (55.9%) male and 216 (44.1%) female subject observations. Of these, we observed 20 (4.1%) individuals that were identified as children and appeared to be younger than 16 years old. For sample size, we needed a minimum of 241 observations for the detection of mask wearing with a medium effect size of 0.5, a significance level of .01, and a power of 90%. We collected 490 observations to allow a buffer for stratification in bivariate and regression analyses.

Most observations were made at Bourbon Street and Times Square due to the high foot traffic in those cities. The 3 other cameras (2 in Seaside Heights and 1 in Key West) had frequent visual obstructions from direct sunlight or fog, were too far away from individuals, or had a small field of view with several individuals that passed too quickly.

Table 2. Demographic characteristics of people observed in New Orleans, New York, Florida, and New Jersey.

Characteristic	Mask (n=241)	No mask (n=249)	Total (n=490)	Chi-square test		
				Chi-square, (*df*)	Value, n	*P* value
Gender appearance, n (%)				0.0026 (1)	488	.96
Male	120 (49.8)	154 (61.8)	274 (55.9)			
Female	121 (50.2)	95 (38.2)	216 (44.1)			
Age, n (%)				0.2725 (1)	489	.60
Child	11 (4.6)	9 (3.6)	20 (4.1)			
Adult	230 (95.4)	240 (96.4)	470 (95.9)			
Face touch, n (%)				8.64 (1)	490	<.05
Yes	56 (23.2)	88 (35.3)	144 (29.4)			
No	185 (76.8)	161 (64.7)	346 (70.6)			
Multiple face touch, n (%)				9.672 (1)	490	<.05
Yes	17 (7.1)	40 (16.1)	57 (11.6)			
No	224 (92.9)	209 (83.9)	433 (88.4)			
Long face touch, n (%)				5.864 (1)	490	<.05
Yes	29 (12)	50 (20)	79 (16.1)			
No	212 (88)	199 (80)	411 (83.9)			
Site of face touch, n				15.68 (9)	144	<.05
1	5	10	15			
2	1	3	4			
3	21	28	49			
4	1	2	3			
5	4	5	9			
6	2	18	20			
7	9	7	16			
8	5	2	7			
9	1	6	7			
10	7	7	14			
Oral activity, n (%)				36.84 (1)	490	<.05
Yes	1 (0.4)	38 (15.3)	39 (8)			
No	240 (99.6)	211 (84.7)	451 (92)			
Oral detail, n (%)				19.11 (3)	484	<.05
Smoking	0 (0)	11 (4.4)	11 (2.2)			
Vaping	0 (0)	2 (0.8)	2 (0.4)			
Drinking	1 (0.4)	19 (7.6)	20 (4.1)			
Eating	0 (0)	6 (2.4)	6 (1.2)			
Frequency of face touch (touch/s)	0.03	0.03	0.03			
Location, n (%)				19.11 (3)	484	<.05
New Orleans	70 (29)	105 (42.2)	175 (35.7)			
New York	161 (66.8)	122 (49)	283 (57.8)			
New Jersey	6 (2.5)	8 (3.2)	14 (2.9)			
Florida	4 (1.7)	14 (5.6)	18 (3.7)			

Face-Touching Observations

From our observations, the majority of the population touched their face in the same area. As seen in Figure 1, area 3, below the nose to the bottom of their chin, was the most common place that subjects touched their face. Most other areas fell within this larger area and were coded as areas 1, 2, 3, 6, or 7 (mask region as "mskreg1"). Area 3 was the most frequently observed, because in many cases, the video quality was insufficient to determine the exact location.

We observed a total of 144 people touching their face at least once, with many touching different regions of their face. From the population that touched their face, 88 (61.1%) people were not wearing a mask and 56 (38.9%) people were wearing a mask. We counted a total of 273 discrete face touches in all 490 observed subjects. Of everyone who touched their face for more than one second, 37% (29/79) were wearing a mask and 63% (50/79) were not wearing a mask. Face touches longer than 6 seconds accounted for 4.2% (6/144) of all face touch observations, with a 15-second touch being the lengthiest touch (n=1).

Of those who were wearing a mask, only 7.1% (17/241) touched the face more than once, seen on Table 3. Of those who touched their face for longer than 1 second, 37% (29/79) touched their face more than once.

Table 3. Frequency of single touch versus multiple touches shown across subject's oral activity, touch duration and mask wearing.

Variable	Multitouch	No multitouch	Chi-square test		
			Chi-square (*df*)	Value, n	*P* value
Oral activity (n=39)	16 (41)	23 (59)	35.6 (1)	490	<.001
Long touch duration (>1 second; n=79)	29 (36.7)	50 (63.3)	57.6 (1)	490	<.001
Wearing a mask (n=241)	17 (7.1)	224 (92.9)	9.67 (1)	490	.002

Mask-Wearing Observations

Mask-wearing style was recorded to observe if subjects were wearing their mask properly, covering their nose and mouth. Subjects who were only covering their mouth with their mask, wearing it as a chin strap, or taking their mask on and off were considered to be not wearing a mask for this study. In Table 4, the frequency of mask style is presented.

Washable masks (homemade or manufactured) accounted for the most frequently observed type of mask at 53.1% (128/241). At 43.2% (104/241), disposable surgical masks were the second most frequently observed type of mask. Subsequent observations of each additional type of mask drastically fell to less than 8%: N95 masks only accounted for 3.7% (9/241) of all observed face masks. Mask types were only recorded if the subject was wearing a mask properly.

Table 4. Frequency of mask style.

Mask styles	Frequency observed (N=490), n (%)
Covering nose and mouth	241 (49.2)
Not wearing one	158 (32.2)
Covering mouth only	24 (4.9)
Chin strap	49 (10)
Partially wearing mask[a]	18 (3.7)

[a]Partially wearing mask consists of taking the mask on and off and dangling it from the ear.

Discussion

Principal Findings

This exposure assessment used a naturalistic observation method and found that the average face touching frequency was 0.03 touches per second or 1.8 touches per minute for over 400 individuals. This rate was comparable to other studies, which found a frequency of 0.8 touches per minute in an indoor environment [7] for 10 subjects. Our original hypothesis was that wearing a face mask will increase the frequency of face touching. Contrary to our hypothesis, the regression shows a negative association between mask wearing and the frequency of face touching.

The results indicate that almost half of the observations made (241/490, 49.2%) were of people wearing a mask properly, covering the nose and mouth. In both the mask-wearing and non–mask-wearing groups, the most frequent area of face touch was the space between the nostrils and the chin, as shown in Figure 1. From our observations of individuals that touched their face, 61.1% (88/144) were not wearing a mask. Of all the individuals that touched their face more than once, 70.2% (40/57) of them were not wearing a mask. These observations show that those wearing a mask had a lower face-touching frequency compared to those who were not wearing a mask.

Face masks covering the nose and mouth have been proven to limit the spread of the COVID-19 disease [2,14]. In public spaces, people infected with COVID-19 can contaminate their environment, which will later on contaminate the hands of the general public. Hand to face transmission is a critical transmission route to study in public areas where there are significantly more objects or materials that are likely to carry infection [7]. Face-touching behaviors are important to study as it relates to exposure assessment science. It is crucial to understand how these behaviors impact the spread of disease or viral particles.

Limitations

The use of public webcam footage only allowed the subjects to be observed for a short walking distance in the few seconds they were in frame of the shot. This is due to the constraints of the focal length in the camera lens. Therefore, subjects were only observed for anywhere between our minimum inclusion criteria of 5 seconds up through the longest recorded duration of 5 minutes. A wider camera lens would have been useful for monitoring subjects at a greater distance and for a longer period of time. However, the constraint of public webcams created a standardized focal length that allowed for a consistent review of the footage. In other similar studies, the monitoring distance and observation time are not clear [15]. Additionally, this study took place exclusively in outdoor public spaces and not in enclosed spaces such as offices, markets, restaurants, etc. Thus, the findings in this study can only be applicable to face-touching behaviors in public spaces and not in enclosed spaces.

Comparison With Prior Work

The findings of this study concluded that mask wearing is not associated with an increased frequency of face touching. Another study investigated this hypothesis by comparing face-touching behaviors before and during the COVID-19 pandemic [9]. Their study took place in China, Japan, South Korea, and Western Europe and found that the frequency of face touching decreased as mask mandates were being implemented. This is important because it demonstrates that mask wearers have been shown to reduce face-touching behaviors. Therefore, face masks offer a double advantage in decreasing viral transmission through the protection of the oropharyngeal area and decreasing the potential for face-touching frequency.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Video data review and record.
[XLSX File (Microsoft Excel File), 82 KB - ijmr_v12i1e43308_app1.xlsx]

References

1. COVIDView, key updates for week 33. Centers for Disease Control and Prevention. 2020 Aug 21. URL: https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/past-reports/08212020.html [accessed 2022-08-11]
2. Sim SW, Moey KSP, Tan NC. The use of facemasks to prevent respiratory infection: a literature review in the context of the Health Belief Model. Singapore Med J 2014 Mar;55(3):160-167 [FREE Full text] [doi: 10.11622/smedj.2014037] [Medline: 24664384]
3. Tao Z, Dong J, Culleton R. The use of facemasks may not lead to an increase in hand-face contact. Transbound Emerg Dis 2020 Nov;67(6):3038-3040 [FREE Full text] [doi: 10.1111/tbed.13698] [Medline: 32594652]
4. Taylor S, Asmundson GJG. Negative attitudes about facemasks during the COVID-19 pandemic: the dual importance of perceived ineffectiveness and psychological reactance. PLoS One 2021 Feb 17;16(2):e0246317 [FREE Full text] [doi: 10.1371/journal.pone.0246317] [Medline: 33596207]
5. Al-Ramahi M, Elnoshokaty A, El-Gayar O, Nasralah T, Wahbeh A. Public discourse against masks in the COVID-19 era: infodemiology study of Twitter data. JMIR Public Health Surveill 2021 Apr 05;7(4):e26780 [FREE Full text] [doi: 10.2196/26780] [Medline: 33720841]
6. Masks may actually increase your coronavirus risk if worn improperly, surgeon general warns. CNN. 2020 Mar 2. URL: https://www.cnn.com/2020/03/02/health/surgeon-general-coronavirus-masks-risk-trnd/index.html [accessed 2021-10-12]
7. Nicas M, Best D. A study quantifying the hand-to-face contact rate and its potential application to predicting respiratory tract infection. J Occup Environ Hyg 2008 Jun;5(6):347-352 [FREE Full text] [doi: 10.1080/15459620802003896] [Medline: 18357546]
8. Liebst LS, Ejbye-Ernst P, de Bruin M, Thomas J, Lindegaard MR. Face-touching behaviour as a possible correlate of mask-wearing: a video observational study of public place incidents during the COVID-19 pandemic. Transbound Emerg Dis 2022 May 18;69(3):1319-1325 [FREE Full text] [doi: 10.1111/tbed.14094] [Medline: 33817991]
9. Chen YJ, Qin G, Chen J, Xu JL, Feng DY, Wu XY, et al. Comparison of face-touching behaviors before and during the coronavirus disease 2019 pandemic. JAMA Netw Open 2020 Jul 01;3(7):e2016924 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.16924] [Medline: 32725247]
10. EarthCam - webcam network. EarthCam. URL: https://www.earthcam.com [accessed 2021-08-18]
11. Zhang N, Jia W, Wang P, King M, Chan P, Li Y. Most self-touches are with the nondominant hand. Sci Rep 2020 Jun 26;10(1):10457 [FREE Full text] [doi: 10.1038/s41598-020-67521-5] [Medline: 32591572]
12. Creative Commons License - Attribution 4.0 International - CC BY 4.0. Creative Commons. URL: https://creativecommons.org/licenses/by/4.0/ [accessed 2023-05-26]

13. Champely S, Ekstrom C, Dalgaard P, Gill J, Weibelzahl S, Anandkumar A, et al. pwr: basic functions for power analysis. R Project. 2020 Mar 17. URL: https://cran.r-project.org/web/packages/pwr/ [accessed 2023-04-26]
14. Brooks JT, Butler JC. Effectiveness of mask wearing to control community spread of SARS-CoV-2. JAMA 2021 Mar 09;325(10):998-999 [FREE Full text] [doi: 10.1001/jama.2021.1505] [Medline: 33566056]
15. Arp NL, Nguyen TH, Linck EJG, Feeney AK, Schrope JH, Ruedinger KL, et al. Use of face coverings in public during the COVID-19 pandemic: an observational study. medRxiv reprint posted on online on June 23, 2020. [FREE Full text] [doi: 10.1101/2020.06.09.20126946] [Medline: 32587989]

Abbreviations

IRB: institutional review board
H1N1: hemagglutinin type 1 neuraminidases type 1
H5N1: hemagglutinin type 5 neuraminidases type 1
SARS: severe acute respiratory syndrome

Edited by A Mavragani; submitted 07.10.22; peer-reviewed by M Kapsetaki, M Bin Mohd Hanim, E Baca-García, B Lukhele; comments to author 11.01.23; revised version received 17.02.23; accepted 03.04.23; published 29.05.23.

Please cite as:
Niesen S, Ramon D, Spencer-Hwang R, Sinclair R
The Relationship Between Face Mask Use and Face-Touching Frequency in Public Areas: Naturalistic Study
Interact J Med Res 2023;12:e43308
URL: https://www.i-jmr.org/2023/1/e43308
doi:10.2196/43308
PMID:37094229

Original Paper

Mobile Health–Supported Active Syndrome Surveillance for COVID-19 Early Case Finding in Addis Ababa, Ethiopia: Comparative Study

Haileleul Bisrat[1*], MSc; Tsegahun Manyazewal[1*], PhD; Abebaw Fekadu[1*], PhD

Center for Innovative Drug Development and Therapeutic Trials for Africa (CDT-Africa), College of Health Sciences, Addis Ababa University, Addis Abeba, Ethiopia

[*]all authors contributed equally

Corresponding Author:
Haileleul Bisrat, MSc
Center for Innovative Drug Development and Therapeutic Trials for Africa (CDT-Africa)
College of Health Sciences
Addis Ababa University
BL_03_420 St.
Addis Abeba, 9086
Ethiopia
Phone: 251 911940459
Email: haylishb2@gmail.com

Abstract

Background: Since most people in low-income countries do not have access to reliable laboratory services, early diagnosis of life-threatening diseases like COVID-19 remains challenging. Facilitating real-time assessment of the health status in a given population, mobile health (mHealth)–supported syndrome surveillance might help identify disease conditions earlier and save lives cost-effectively.

Objective: This study aimed to evaluate the potential use of mHealth-supported active syndrome surveillance for COVID-19 early case finding in Addis Ababa, Ethiopia.

Methods: A comparative cross-sectional study was conducted among adults randomly selected from the Ethio telecom list of mobile phone numbers. Participants underwent a comprehensive phone interview for COVID-19 syndromic assessments, and their symptoms were scored and interpreted based on national guidelines. Participants who exhibited COVID-19 syndromes were advised to have COVID-19 diagnostic testing at nearby health care facilities and seek treatment accordingly. Participants were asked about their test results, and these were cross-checked against the actual facility-based data. Estimates of COVID-19 detection by mHealth-supported syndromic assessments and facility-based tests were compared using Cohen Kappa (κ), the receiver operating characteristic curve, sensitivity, and specificity analysis.

Results: A total of 2741 adults (n=1476, 53.8% men and n=1265, 46.2% women) were interviewed through the mHealth platform during the period from December 2021 to February 2022. Among them, 1371 (50%) had COVID-19 symptoms at least once and underwent facility-based COVID-19 diagnostic testing as self-reported, with 884 (64.5%) confirmed cases recorded in facility-based registries. The syndrome assessment model had an optimal likelihood cut-off point sensitivity of 46% (95% CI 38.4-54.6) and specificity of 98% (95% CI 96.7-98.9). The area under the receiver operating characteristic curve was 0.87 (95% CI 0.83-0.91). The level of agreement between the mHealth-supported syndrome assessment and the COVID-19 test results was moderate (κ=0.54, 95% CI 0.46-0.60).

Conclusions: In this study, the level of agreement between the mHealth-supported syndromic assessment and the actual laboratory-confirmed results for COVID-19 was found to be reasonable, at 89%. The mHealth-supported syndromic assessment of COVID-19 represents a potential alternative method to the standard laboratory-based confirmatory diagnosis, enabling the early detection of COVID-19 cases in hard-to-reach communities, and informing patients about self-care and disease management in a cost-effective manner. These findings can guide future research efforts in developing and integrating digital health into continuous active surveillance of emerging infectious diseases.

(Interact J Med Res 2023;12:e43492) doi:10.2196/43492

KEYWORDS

mobile health; mHealth; digital health; COVID-19; syndrome assessment; surveillance; Ethiopia; public health; syndrome surveillance; self-care; telemedicine; telecom, SARS-CoV-2

Introduction

The global health community learned a lesson from the COVID-19 pandemic: each country needs to strengthen its epidemic preparedness and response capacity to mitigate emerging and reemerging infectious diseases at an earlier stage. It was noted that trajectories of the COVID-19 pandemic would have been curtailed through active and adequate disease surveillance, laboratory infrastructure, health workforce, pandemic planning, and management systems [1,2]. With limited disease preparedness and response capacity and a largely susceptible economy, Africa was endangered by the pandemic and its interrelated economic consequences [3-5]. Although the pandemic progressed more slowly in Africa compared to the rest of the globe, almost all African countries have been affected by the pandemic in one way or another [5-7]. More than 9.5 million COVID-19 confirmed cases were recorded across the African continent as of May 08, 2023 [8].

Ethiopia, a country in sub-Saharan Africa, confirmed its first case of COVID-19 on March 13, 2020. Two days later, the World Health Organization declared a pandemic of the disease [9]. As of May 10, 2023, there have been 500,853 confirmed cases of COVID-19, with 7574 deaths, reported to the World Health Organization [8]. COVID-19 placed a significant burden on patients with chronic diseases in Ethiopia, affecting their ability to access their routine clinical care and treatment [10,11]. Health care workers [12,13] and the community at large [14,15] had uncertainties in implementing preventive measures against the disease. Working with the global health community, the Ethiopian government introduced vaccines against COVID-19, and as of May 6, 2023, the country administered 54,041,862 vaccine doses [9]. Ethiopia is one of the resource-constrained countries, overwhelmed by a double burden of infectious and noninfectious diseases and with limited capacity to find and treat cases at an early stage [15-18]. The country has limited health care infrastructure and workforce [19-21].

The systematic digitalization of the health care industry may improve access to quality care and lower health care costs. Through the use of smartphones, health information technology, wearable devices, telemedicine, and personalized medicine, digital health technologies have the potential to facilitate health care and attain intended health outcomes [22-24]. Mobile health (mHealth) apps are emerging as a strategy to improve health care delivery and outcomes [25-27]. As mobile phones are more accessible to many people in low- and middle-income nations like Ethiopia, the technology is likely a means to offer impactful and affordable solutions to address diseases of significant public health importance.

In the context of COVID-19, the potential use of mHealth technology to prevent and control the pandemic in Africa has been studied [28-32]; however, evidence is limited, especially on the potential impact of mHealth technology on active syndrome surveillance. As the number of mobile phone owners continues to rise in many African countries, including Ethiopia, mHealth technologies may aid in the early identification of COVID-19 cases within the community and link them to diagnostic testing centers.

Africa showed limited capacity and flexibility to scale up COVID-19 testing, effectively tracing contacts of confirmed cases, and promptly training and deploying community health workers to help in the early identification of cases and their connection to appropriate care [33]. Existing conventional surveillance systems were not sufficient enough to respond to the COVID-19 pandemic. Moreover, it was difficult for the majority of the African population to fully comply with the preventive measures due to socioeconomic consequences. Hence, the problem of mitigating COVID-19 spanned across infrastructure, human resources for health, diagnosis, logistics, population literacy, and economy. There was a strong need to develop a simple alternative method to help detect COVID-19 cases in the community quickly and cost-effectively. Digital health interventions offered such potential.

Therefore, we aimed to evaluate the potential use of mHealth-supported active syndrome surveillance for the early identification of COVID-19 cases in Addis Ababa, Ethiopia.

Methods

Study Design and Participants

This study is part of an ongoing national, population-based cohort mHealth-supported study—"mHealth-supported continuous national surveillance of COVID-19 for early case finding and population-level impact and control in Ethiopia (EPIC)." This specific study is a population-based cross-sectional comparative study, using mobile phone call surveys and a cross-sectional comparison of mHealth-supported COVID-19 syndrome diagnosis versus confirmatory laboratory tests. In this study, a mobile call survey was deemed an appropriate data collection method, given the nature of the pandemic, the wide geographic area of the country, and economic feasibility.

Eligible participants were adults aged 18 years and older living in Ethiopia, speaking one or more of the 3 Ethiopian working languages (Amharic, Afan Oromo, and Tigrigna), and having no hearing or cognitive impairment or serious mental illness that impedes participation in the study. Potential participants were selected from the population of individuals with mobile phones registered centrally with either the federal or Addis Ababa authorities. Hence, participants were randomly selected from the list of mobile phone numbers available in the country using computer-generated random numbers. Initially, 11 million numbers were generated, from which 30,000 phone numbers were randomly selected. Our study uses the first 4180 phone numbers from the 30,000 randomly generated numbers.

Data Collection

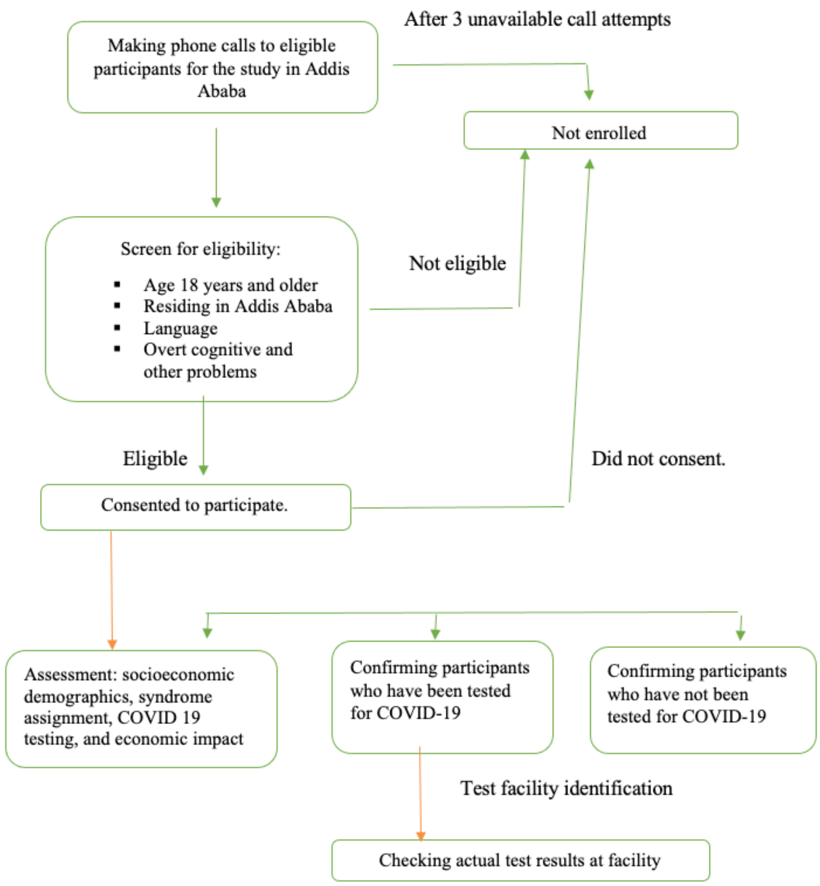
Participants underwent a comprehensive phone interview for COVID-19 syndromic assessments, and their symptoms were scored and interpreted based on national guidelines. COVID-19–like symptoms were measured using a syndromic assessment for acute respiratory illnesses (fever and at least one sign or symptom of respiratory diseases, such as cough, sore throat, runny nose, shortness of breath, loss of smell, and loss of taste). Participants who had COVID-19 were advised to have COVID-19 diagnostic testing at nearby health care facilities and seek treatment accordingly. Participants were asked about their test results, and these were crosschecked against the actual facility-based data.

Questionnaires were implemented on an electronic data capture platform, Open Data Kit, the free app for Android devices that is used to collect and compile data. Whenever a phone number was not responsive or unanswered during the first attempt, repeated calls were made up to 3 times before excluding it from the study.

Data on COVID-19 laboratory test results were collected by reviewing log charts in hospitals and health centers where the participants had undergone laboratory testing. During the phone interview, those who had been tested and had received the COVID-19 test results, along with other indicators, were cross-checked. From the data gathered from government referrals and specialized hospitals, government health centers, private hospitals, and private clinics, it was possible to determine whether the information provided to the participants who indicated they had undergone a COVID-19 test, along with other indicators, matched the registration logs at health care facilities.

Different quality assurance strategies were followed to ensure data integrity, quality, and reliability. In the data collection process, 13 data collectors with a minimum of BSc degrees in the health profession, along with 1 supervisor, undertook the data collection responsibilities after all contractual and training aspects were finalized. The survey procedures and tools were pretested for utility, feasibility, and acceptability. There was a face-to-face meeting with the supervisor every morning for reviewing and planning purposes. The data were stored daily on a central server at Center for Innovative Drug Development and Therapeutic Trials (CDT)-Africa in Addis Ababa University, Ethiopia. Figure 1 presents the overall procedures followed for data collection.

Figure 1. Data collection procedure.



Ethics Approval

This study has been reviewed and approved by the Institutional Review Board of the College of Health Sciences at Addis Ababa University (086/20/CDT). The study participants were briefed about the study over the phone and asked about their willingness to participate in the study. They were enrolled only after giving verbal consent. There was no compensation for participation in the study. The data were kept confidential and used for study purposes only.

Statistical Analysis

Interpretation of syndrome assessment followed the guidelines of the Ethiopian Ministry of Health. COVID-19 and non–COVID-19 signs and symptoms were categorized by assigning a score of 1 to COVID-19 and a score of 0 to Non–COVID-19 suspects. The formula used to determine the distribution of participants based on syndrome assessment and test results is as follows (Table 1):

$$\kappa = P_o - P_e / 1 - P_e = 1 - 1 - P_o / 1 - P_e$$

Where P_o = relative observed agreement among raters, P_e = hypothetical probability of chance agreement, and κ = kappa status.

$$Po = (a + d) / N$$

Where *a* = the total number of instances that both raters said were correct (ie, the raters are in agreement), *b* = the total number of instances where what rater 2 said was incorrect, but what rater 1 said was correct (ie, rater disagreement), *c* = the total number of instances where what rater 1 said was incorrect, but what rater 2 said was correct (ie, rater disagreement), and *d* = the total number of instances where what both raters said was incorrect (ie, raters are in agreement).

For Cohen kappa, the chance agreement is defined as the sum of the products of marginal distributions, which is as follows:

$$P_e (\kappa) = (P.1\ P1.) + (P.2\ P2.)$$

Table 1. Determination of the distribution of participants based on syndrome assessment and test results.

Test result	Syndrome assessment			
	Category 1 (yes)	Category 2 (no)	Total	
Category 1 (yes)	a	b	a + b	P1. = (a + b) / N
Category 2 (no)	c	d	c + d	P2. = (c + d) / N
Total	a + c	b + d	N	
	P.1 = (a + c) / N	P.2 = (b + d) / N		

For both the syndrome assessment and test results, the area under the receiver operating characteristic (ROC) curve was used to assess the overall diagnostic performance, indicating the possibility of accurately diagnosing all symptoms. The area under the curve of a procedure should be close to 1 for it to be highly sensitive and specific. The approach is considered more accurate if the curve closely aligns with the left-hand border and the top border of the ROC space. If the area under the ROC curve exceeded 0.75, we deemed our methods to be appropriate.

For validity measures, sensitivity and specificity along with their corresponding 95% CIs were compared between syndrome assessment and laboratory test results. Sensitivity refers to the ability of the COVID-19 syndrome assessment to correctly identify COVID-19 test results as designated by the test outcome, while specificity refers to the proportion of cases caused by other factors, correctly identified as non–COVID-19. These 2 measures are closely related to type 1 and type 2 errors. Hence, both sensitivity and specificity were calculated. The formula for the calculation was defined as follows: sensitivity = *TP / (TP + FN)* and specificity = *TN / (FP + TN)*. In these formulas, TP is true positive, FP is false positive, TN is true negative, and FN is false negative.

Data analysis was conducted using STATA 17 (StataCorp) software. Individual participants had a unique ID, and all data sets were merged using this ID before analysis. Two separate variables, COVID-19 syndrome assessment and COVID-19 laboratory test results, were generated for comparison purposes and for documenting the trend of COVID-19 results. Coding, recording, labeling, and analysis were also done.

Estimates of COVID-19 detection through syndrome assessment and test results were compared using Cohen kappa (κ) and ROC curve analysis. A κ value less than 0 indicates no agreement, and values between 0 and 0.20 indicate slight agreement; values from 0.21 to 0.40 denote fair agreement; values from 0.41 to 0.60 suggest moderate agreement; those from 0.61 to 0.80 imply substantial agreement; and finally, values from 0.81 to 1 indicate an almost perfect agreement [34].

Results

Background Characteristics

In this study, a total of 35,646 national calls were made, of which 2741 participants were from Addis Ababa. Of the 6818 calls made, 4077 were characterized as unavailable, unanswered, switched off, disconnected, or hung up, resulting in a response rate of 67.2% among the participants from Addis Ababa (Figure 2).

A total of 2741 data points were collected through telephone (mobile phone) interviews for the period from December 2021 to February 2022. Of these, 1476 (53.85%) were male and 1265 (46.15%) were female individuals; 1213 (44.25%) were from the age group of 18-29 years, 1151 (41.99%) had a diploma or higher academic qualification, and 1443 (52.65%) were married (Table 2).

Figure 2. Background characteristics.

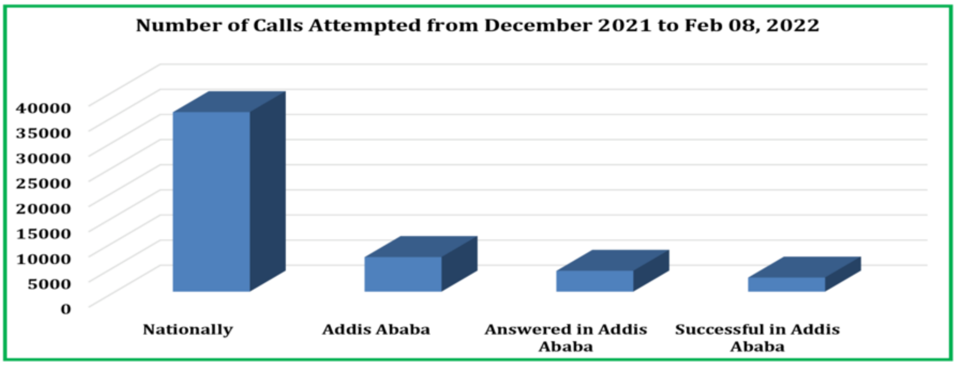


Table 2. Background characteristics of the study participants (N=2741).

Background characteristics	Values, n (%)
Sex	
Male	1476 (53.85)
Female	1265 (46.15)
Age category (years)	
18-29	1213 (44.25)
30-39	879 (32.07)
40-49	379 (13.83)
50-59	159 (5.80)
60-69	78 (2.85)
70-79	31 (1.13)
≥80	2 (0.07)
Level of education	
Cannot read or write	75 (2.74)
Primary school	450 (16.42)
Secondary school	993 (36.23)
Certificate	72 (2.63)
Diploma or above	1151 (41.99)
Marital status	
Single	1162 (42.39)
Married	1443 (52.65)
Divorced	86 (3.14)
Widowed	50 (1.82)

Comparison of mHealth-Supported Syndrome Assessment Versus Laboratory Test Results

Of the total 2741 data points collected through mobile phones, 1371 participants had undergone a COVID-19 test. Among the tested participants, 884 (64.5%) received confirmation of their results from the respective hospitals and health centers; however, the remaining 487 participants did not have their results recorded in the health care facilities and were consequently excluded from this particular analysis.

There were some variations in laboratory results collected in different months. In December 2021, confirmed COVID-19 cases accounted for 37% (32/86), while results from syndrome assessment constituted only 29% (45/155) of the participants with a positive COVID-19 outcome. In January, confirmed results accounted for 63% (54/86), while syndrome assessment indicated that 67% (489/729) were non–COVID-19 cases. Table 3 summarizes the COVID-19 mHealth-supported syndrome assessment and laboratory test results categorized by months and COVID-19 results.

Table 3. COVID-19 mHealth-supported syndrome assessment and laboratory test results by months and test results (N=884).

Month and year	COVID-19 laboratory test result		COVID-19 mHealth-supported syndrome assessment	
	Non–COVID-19 (n=798), n (%)	COVID-19 (n=86), n (%)	Non–COVID-19 (n=729), n (%)	COVID-19 (n=155), n (%)
December 2021	252 (32)	32 (37)	239 (33)	45 (29)
January 2022	544 (68)	54 (63)	489 (67)	109 (70)
February 2022	2 (0.25)	0 (0)	1 (0.14)	1 (1)

Agreement Between mHealth-Supported Syndrome Assessment and Laboratory Test Results

The observed agreement between the mHealth-supported syndrome assessment and laboratory results was 0.89 and Cohen kappa for the hypothetical probability of chance agreement was 0.54. The syndrome assessment model had an optimal likelihood cut-off point sensitivity of 46% (95% CI 38.4-54.6) and specificity of 98% (95% CI 96.7-98.9; Table 4).

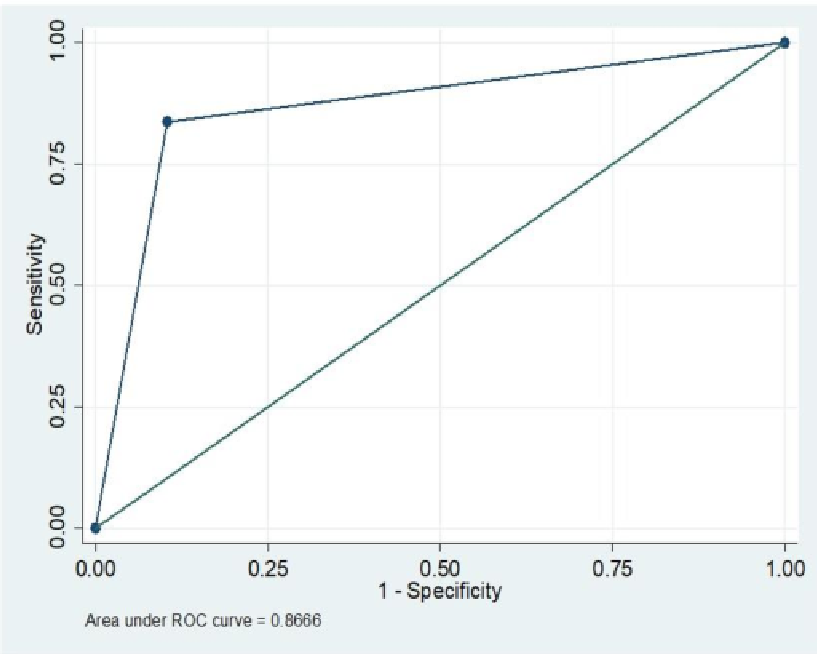
Table 4. Distribution of participants by mHealth-supported syndrome assessment and test results for COVID-19 using mHealth categories.

Laboratory test results	mHealth-supported syndrome assessment			
	COVID-19	Non–COVID-19	Total	
COVID-19	72	14	86	P1. = 0.097
Non–COVID-19	83	715	798	P2. = 0.902
Total	155	729	884	
	P.1 = 0.175	P.2 = 0.824		

ROC analysis

Figure 3 presents the ROC analysis in a one-to-one square. The area under the curve captures the relationship between the sensitivity and specificity of the COVID-19 test results and reflects the performance of the method used. The curve follows the left-hand border and the top border of the ROC space, indicating an acceptable level of accuracy. The ROC curve shows that the predicted area under the ROC curve for COVID-9 is 0.87 for the mHealth-supported syndrome assessment results compared to the laboratory test results. This indicates the good diagnostics performance of the syndrome assessment method used.

Figure 3. Receiver operating characteristic (ROC) curve.



Discussion

Principal Findings

In this study, we evaluated the potential use of mHealth-supported active syndrome surveillance for COVID-19 early case finding. Interviewing a total of 2741 adults through the mHealth platform, we found 50% (1371/2741) had COVID-19 symptoms at least once, and 64.5% (884/2741) had laboratory test results recorded in facility-based registries. The syndrome assessment model had an optimal likelihood cut-off point sensitivity and specificity of 46% and 98%, respectively, with an area under the ROC curve of 0.87 and a moderate level of agreement between the mHealth-supported syndrome assessment and the laboratory-confirmed COVID-19 test results (κ=0.54).

One of the efficient methods for addressing epidemic or pandemic outbreaks is the use of health information technologies, such as mHealth, which facilitate remote communication. To manage and regulate the recently emerged COVID-19 pandemic, studies are underway to assess the functionalities of different digital health technologies. This study described the use of mHealth-supported active surveillance for COVID-19 early case findings by evaluating the agreement between the mHealth-supported syndrome assessment and laboratory-confirmed COVID-19 test results. The outcomes revealed a moderate level of agreement between the results of the COVID-19 syndrome assessment supported by mHealth and the laboratory test results, at 89%, with a kappa value of 0.54. This suggests that mHealth could be a potential alternative to the standard laboratory-based confirmatory diagnosis to find COVID-19 cases. Previous studies have shown that mHealth technologies could have significant contributions to self-care for patients with COVID-19 [35] and the dissemination of COVID-19–related information [36,37]. As mHealth and other digital health technologies are yet at an early stage of development in Africa, the capacity and readiness of each country to effectively adopt, implement, and scale up digital health interventions require due diligence [38-42].

The findings of this study show that the mHealth-supported COVID-19 syndrome assessment model has an optimal likelihood cut-off point sensitivity and specificity when compared with the laboratory-based tests. The ROC curve indicates the good diagnostics performance of the mHealth syndrome assessment model. The study used a holistic syndrome assessment approach, based on the national guidelines, to examine and interpret the COVID-19 status of individuals participating in the study. Previous studies show that syndromic diagnosis of COVID-19 based on a single symptom cannot accurately identify individuals who might have the virus, and hence, investigating cases through combinations of syndromes along with additional information, such as recent contacts, travel history, or vaccination status, is necessary [43-45]. Studies show that although digital health–enabled communication may not be as effective as in-person communication, it represents a safe and efficient alternative to collecting evidence-based medical history, especially during the COVID-19 period, when in-person care cannot be provided [46].

Investigating the potential use of mHealth-supported active syndrome surveillance for COVID-19 early case finding in Ethiopia, this study provides important insights. Future studies can further explore how digital health technologies can be used for the early identification of emerging infectious diseases, reducing their transmissions, as well as monitoring and mitigating their undue impact. Our findings inform the scientific community about how mHealth can be adapted for health system responses and the implementation challenges and opportunities within a resource-constrained country context. Scientific evidence regarding the potential use of such digital health technologies and their interrelated challenges is important for guiding policy and practice, especially in countries that have not yet fully embraced digital health interventions.

Study Limitations

This research has certain limitations. The response rate of the participants was not as anticipated. Because the survey was conducted over the phone, only those who had cell phones at the time of data collection were included in the survey, which limits the generalizability of our findings. Interrupted call connectivity, inconsistency between the participants' responses and the data in facility registries, and participants undergoing COVID-19 testing for reasons other than syndromes (eg, travel health certificates) were some of the shortcomings in the data collection process. To mitigate these issues, only clear and consistent data were included and used in the analysis and interpretation of the study.

Conclusions

In this study, the level of agreement between the mHealth-supported syndrome assessment results and the actual laboratory-confirmed COVID-19 results was reasonable at 89%. mHealth-supported syndromic assessment of COVID-19 is a potential alternative method to the standard laboratory-based confirmatory diagnosis to detect COVID-19 cases at an earlier stage in hard-to-reach communities and to advise patients on self-care and disease management in a cost-effective way. The findings show that the mHealth platform is valid for COVID-19 surveillance in Ethiopia, where health infrastructures are limited. Given the growing digital health landscape, the findings of our study offer valuable insight for guiding future research efforts in developing and integrating digital health into continuous active surveillance of emerging infectious diseases.

Acknowledgments

This research received funding from Addis Ababa University through the thematic grant scheme. Additional support was obtained from the National Institute for Health Research (NIHR) Global Health Research Unit on Health System Strengthening in Sub-Saharan Africa, King's College London (GHRU 16/136/54) using UK aid from the UK Government to support global health research, and the World Bank grant (P151847-IDA57940) supporting the Center for Innovative Drug Development and Therapeutic

Trials (CDT)-Africa. The authors are thankful to CDT-Africa in Addis Ababa University for the successful coordination of this work. The authors are also grateful to the research team of “the early case finding and population-level impact and control in Ethiopia (EPIC)” study.

All authors declared that they had insufficient or no funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this article.

Data Availability

The data set supporting the conclusions of this study is included within the paper. Any additional material can be obtained upon reasonable request.

Authors' Contributions

HB implemented the study, analyzed the data, and developed the first draft of the manuscript. AF and TM implemented and supervised the study, supported the data analysis, and revised the manuscript. All authors approved the final version for publication.

Conflicts of Interest

None declared.

References

1. Global Burden of Disease 2021 Health Financing Collaborator Network. Global investments in pandemic preparedness and COVID-19: development assistance and domestic spending on health between 1990 and 2026. Lancet Glob Health 2023 Mar;11(3):e385-e413. [doi: 10.1016/S2214-109X(23)00007-4] [Medline: 36706770]
2. Fauci A, Folkers G. Pandemic preparedness and response: lessons from COVID-19. J Infect Dis 2023 Apr 10:jiad095. [doi: 10.1093/infdis/jiad095] [Medline: 37035891]
3. Morsy H, Balma L, Mukasa AN. 'Not a good time': assessing the economic impact of COVID-19 in Africa using a macro-micro simulation approach. Afr Dev Rev 2021 Apr 18;33(Suppl 1):S17-S30 [FREE Full text] [doi: 10.1111/1467-8268.12526] [Medline: 34149238]
4. Bitanihirwe BKY, Ssewanyana D. The health and economic burden of the coronavirus in sub-Saharan Africa. Glob Health Promot 2021 Mar 10;28(1):70-74. [doi: 10.1177/1757975920977874] [Medline: 33302803]
5. Atim MG, Kajogoo VD, Amare D, Said B, Geleta M, Muchie Y, et al. COVID-19 and health sector development plans in Africa: the impact on maternal and child health outcomes in Uganda. RMHP 2021 Oct;Volume 14:4353-4360. [doi: 10.2147/rmhp.s328004]
6. Salyer SJ, Maeda J, Sembuche S, Kebede Y, Tshangela A, Moussif M, et al. The first and second waves of the COVID-19 pandemic in Africa: a cross-sectional study. The Lancet 2021 Apr;397(10281):1265-1275. [doi: 10.1016/s0140-6736(21)00632-2]
7. Kawuki J, Chan PS, Fang Y, Chen S, Mo PKH, Wang Z. Knowledge and practice of personal protective measures against COVID-19 in Africa: systematic review. JMIR Public Health Surveill 2023 May 16;9:e44051 [FREE Full text] [doi: 10.2196/44051] [Medline: 37058578]
8. WHO coronavirus (COVID-19) dashboard. World Health Organization. URL: https://covid19.who.int/ [accessed 2023-08-09]
9. Mohammed H, Oljira L, Roba KT, Yimer G, Fekadu A, Manyazewal T. Containment of COVID-19 in Ethiopia and implications for tuberculosis care and research. Infect Dis Poverty 2020 Sep 16;9(1):131 [FREE Full text] [doi: 10.1186/s40249-020-00753-9] [Medline: 32938497]
10. Chilot D, Woldeamanuel Y, Manyazewal T. COVID-19 burden on HIV patients attending antiretroviral therapy in Addis Ababa, Ethiopia: a multicenter cross-sectional study. Front Med (Lausanne) 2022 Mar 2;9:741862 [FREE Full text] [doi: 10.3389/fmed.2022.741862] [Medline: 35308528]
11. Chilot D, Woldeamanuel Y, Manyazewal T. Real-time impact of COVID-19 on clinical care and treatment of patients with tuberculosis: a multicenter cross-sectional study in Addis Ababa, Ethiopia. Ann Glob Health 2021;87(1):109 [FREE Full text] [doi: 10.5334/aogh.3481] [Medline: 34824990]
12. Hasen AA, Seid AA, Mohammed AA. Anxiety and stress among healthcare professionals during COVID-19 in Ethiopia: systematic review and meta-analysis. BMJ Open 2023 Feb 07;13(2):e070367 [FREE Full text] [doi: 10.1136/bmjopen-2022-070367] [Medline: 36750289]
13. Deressa W, Worku A, Abebe W, Gizaw M, Amogne W. Risk perceptions and preventive practices of COVID-19 among healthcare professionals in public hospitals in Addis Ababa, Ethiopia. PLoS One 2021 Jun 25;16(6):e0242471 [FREE Full text] [doi: 10.1371/journal.pone.0242471] [Medline: 34170910]
14. Umer A, Abdella K, Tekle Y, Debebe A, Manyazewal T, Yuya M, et al. Community engagement in the fight against COVID-19: knowledge, attitude, and prevention practices among Dire Dawa residents, Eastern Ethiopia. Front Public Health 2021 Nov 30;9:9. [doi: 10.3389/fpubh.2021.753867]

15. Mose A, Zewdie A, Sahle T. Pregnant women's knowledge, attitude, and practice towards COVID-19 infection prevention in Ethiopia: a systematic review and meta-analysis. PLoS One 2022 Oct 26;17(10):e0276692 [FREE Full text] [doi: 10.1371/journal.pone.0276692] [Medline: 36288349]
16. Mengesha D, Manyazewal T, Woldeamanuel Y. Five-year trend analysis of tuberculosis in Bahir Dar, Northwest Ethiopia, 2015-2019. Int J Mycobacteriol 2021;10(4):437-441 [FREE Full text] [doi: 10.4103/ijmy.ijmy_181_21] [Medline: 34916465]
17. Gebrie D, Manyazewal T, Ejigu DA, Makonnen E. Metformin-insulin versus metformin-sulfonylurea combination therapies in type 2 diabetes: a comparative study of glycemic control and risk of cardiovascular diseases in Addis Ababa, Ethiopia. DMSO 2021 Jul;Volume 14:3345-3359. [doi: 10.2147/dmso.s312997]
18. Manyazewal T, Sisay Z, Biadgilign S, Abegaz W. Hepatitis B and hepatitis C virus infections among antiretroviral-naive and -experienced HIV co-infected adults. J Med Microbiol 2014 May;63(Pt 5):742-747. [doi: 10.1099/jmm.0.063321-0] [Medline: 24757219]
19. Gile PP, van de Klundert J, Buljac-Samardzic M. Strategic human resource management and performance in public hospitals in Ethiopia. Front Public Health 2022 Oct 20;10:915317 [FREE Full text] [doi: 10.3389/fpubh.2022.915317] [Medline: 36339178]
20. Manyazewal T, Matlakala M. Implementing health care reform: implications for performance of public hospitals in central Ethiopia. J Glob Health 2018 Jun;8(1):010403 [FREE Full text] [doi: 10.7189/jogh.08.010403] [Medline: 29497501]
21. Manyazewal T, Oosthuizen MJ, Matlakala MC. Proposing evidence-based strategies to strengthen implementation of healthcare reform in resource-limited settings: a summative analysis. BMJ Open 2016 Sep 20;6(9):e012582 [FREE Full text] [doi: 10.1136/bmjopen-2016-012582] [Medline: 27650769]
22. Manyazewal T, Woldeamanuel Y, Blumberg HM, Fekadu A, Marconi VC. The potential use of digital health technologies in the African context: a systematic review of evidence from Ethiopia. NPJ Digit Med 2021 Aug 17;4(1):125 [FREE Full text] [doi: 10.1038/s41746-021-00487-4] [Medline: 34404895]
23. Manyazewal T, Woldeamanuel Y, Holland DP, Fekadu A, Marconi VC. Effectiveness of a digital medication event reminder and monitor device for patients with tuberculosis (SELFTB): a multicenter randomized controlled trial. BMC Med 2022 Sep 28;20(1):310 [FREE Full text] [doi: 10.1186/s12916-022-02521-y] [Medline: 36167528]
24. Manyazewal T, Woldeamanuel Y, Getinet T, Hoover A, Bobosha K, Fuad O, et al. Patient-reported usability and satisfaction with electronic medication event reminder and monitor device for tuberculosis: a multicentre, randomised controlled trial. EClinicalMedicine 2023 Feb;56:101820 [FREE Full text] [doi: 10.1016/j.eclinm.2022.101820] [Medline: 36684395]
25. Cucciniello M, Petracca F, Ciani O, Tarricone R. Development features and study characteristics of mobile health apps in the management of chronic conditions: a systematic review of randomised trials. NPJ Digit Med 2021 Oct 05;4(1):144 [FREE Full text] [doi: 10.1038/s41746-021-00517-1] [Medline: 34611287]
26. Katapally TR, Bhawra J, Leatherdale ST, Ferguson L, Longo J, Rainham D, et al. The SMART study, a mobile health and citizen science methodological platform for active living surveillance, integrated knowledge translation, and policy interventions: longitudinal study. JMIR Public Health Surveill 2018 Mar 27;4(1):e31 [FREE Full text] [doi: 10.2196/publichealth.8953] [Medline: 29588267]
27. Morjaria P, Bastawrous A, Murthy GVS, Evans J, Sagar MJ, Pallepogula DR, et al. Effectiveness of a novel mobile health (Peek) and education intervention on spectacle wear amongst children in India: Results from a randomized superiority trial in India. EClinicalMedicine 2020 Nov;28:100594 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100594] [Medline: 33294811]
28. Rinke de Wit TF, Janssens W, Antwi M, Milimo E, Mutegi N, Marwa H, et al. Digital health systems strengthening in Africa for rapid response to COVID-19. Front Health Serv 2022 Nov 28;2:987828 [FREE Full text] [doi: 10.3389/frhs.2022.987828] [Medline: 36925782]
29. Getachew E, Adebeta T, Muzazu SGY, Charlie L, Said B, Tesfahunei HA, et al. Digital health in the era of COVID-19: reshaping the next generation of healthcare. Front Public Health 2023 Feb 15;11:942703 [FREE Full text] [doi: 10.3389/fpubh.2023.942703] [Medline: 36875401]
30. Nachega JB, Leisegang R, Kallay O, Mills EJ, Zumla A, Lester RT. Mobile health technology for enhancing the COVID-19 response in Africa: a potential game changer? Am J Trop Med Hyg 2020 Jul;103(1):3-5 [FREE Full text] [doi: 10.4269/ajtmh.20-0506] [Medline: 32476643]
31. Kamulegeya LH, Bwanika JM, Musinguzi D, Bakibinga P. Continuity of health service delivery during the COVID-19 pandemic: the role of digital health technologies in Uganda. Pan Afr Med J 2020 May 20;35(Suppl 2):43 [FREE Full text] [doi: 10.11604/pamj.supp.2020.35.2.23115] [Medline: 33623568]
32. Chereka AA, Demsash AW, Ngusie HS, Kassie SY. Digital health literacy to share COVID-19 related information and associated factors among healthcare providers worked at COVID-19 treatment centers in Amhara region, Ethiopia: A cross-sectional survey. Inform Med Unlocked 2022;30:100934 [FREE Full text] [doi: 10.1016/j.imu.2022.100934] [Medline: 35441087]
33. Nkengasong JN, Mankoula W. Looming threat of COVID-19 infection in Africa: act collectively, and fast. The Lancet 2020 Mar;395(10227):841-842. [doi: 10.1016/s0140-6736(20)30464-5]
34. Kovács A, Palásti P, Veréb D, Bozsik B, Palkó A, Kincses ZT. The sensitivity and specificity of chest CT in the diagnosis of COVID-19. Eur Radiol 2021 May;31(5):2819-2824 [FREE Full text] [doi: 10.1007/s00330-020-07347-x] [Medline: 33051732]

35. Montazeri M, Galavi Z, Ahmadian L. The role of mobile health in prevention, diagnosis, treatment and self-care of COVID-19 from the healthcare professionals' perspectives. Digit Health 2023 May 02;9:20552076231171969 [FREE Full text] [doi: 10.1177/20552076231171969] [Medline: 37152239]
36. Zamberg I, Manzano S, Posfay-Barbe K, Windisch O, Agoritsas T, Schiffer E. A mobile health platform to disseminate validated institutional measurements during the COVID-19 outbreak: utilization-focused evaluation study. JMIR Public Health Surveill 2020 Apr 14;6(2):e18668 [FREE Full text] [doi: 10.2196/18668] [Medline: 32250958]
37. Hajizadeh A, Monaghesh E. Telehealth services support community during the COVID-19 outbreak in Iran: activities of Ministry of Health and Medical Education. Inform Med Unlocked 2021;24:100567 [FREE Full text] [doi: 10.1016/j.imu.2021.100567] [Medline: 33842687]
38. Manyazewal T, Ali MK, Kebede T, Magee MJ, Getinet T, Patel SA, et al. Mapping digital health ecosystems in Africa in the context of endemic infectious and non-communicable diseases. NPJ Digit Med 2023 May 26;6(1):97 [FREE Full text] [doi: 10.1038/s41746-023-00839-2] [Medline: 37237022]
39. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
40. Manyazewal T, Woldeamanuel Y, Fekadu A, Holland DP, Marconi VC. Effect of digital medication event reminder and monitor-observed therapy vs standard directly observed therapy on health-related quality of life and catastrophic costs in patients with tuberculosis: a secondary analysis of a randomized clinical trial. JAMA Netw Open 2022 Sep 15;5(9):e2230509 [FREE Full text] [doi: 10.1001/jamanetworkopen.2022.30509] [Medline: 36107429]
41. Getachew E, Woldeamanuel Y, Manyazewal T. Digital health interventions in the clinical care and treatment of tuberculosis and HIV in central Ethiopia: an initial provider perceptions and acceptability study using the unified theory of acceptance and use of technology model. Int J Mycobacteriol 2022;11(1):1-9 [FREE Full text] [doi: 10.4103/ijmy.ijmy_235_21] [Medline: 35295017]
42. Manyazewal T, Woldeamanuel Y, Blumberg HM, Fekadu A, Marconi VC. The fight to end tuberculosis must not be forgotten in the COVID-19 outbreak. Nat Med 2020 Jun 12;26(6):811-812 [FREE Full text] [doi: 10.1038/s41591-020-0917-1] [Medline: 32398877]
43. Struyf T, Deeks J, Dinnes J, Takwoingi Y, Davenport C, Leeflang M, Cochrane COVID-19 Diagnostic Test Accuracy Group. Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19. Cochrane Database Syst Rev 2022 May 20;5(5):CD013665 [FREE Full text] [doi: 10.1002/14651858.CD013665.pub3] [Medline: 35593186]
44. Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. JAMA 2020 Feb 24:1239-1242. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
45. Gebrie D, Getnet D, Manyazewal T. Efficacy of remdesivir in patients with COVID-19: a protocol for systematic review and meta-analysis of randomised controlled trials. BMJ Open 2020 Jun 04;10(6):e039159. [doi: 10.1136/bmjopen-2020-039159]
46. Verma N, Buch B, Taralekar R, Acharya S. Diagnostic concordance of telemedicine as compared with face-to-face care in primary health care clinics in rural India: randomized crossover trial. JMIR Form Res 2023 Jun 23;7:e42775 [FREE Full text] [doi: 10.2196/42775] [Medline: 37130015]

Abbreviations

CDT: Center for Innovative Drug Development and Therapeutic Trials
EPIC: early case finding and population-level impact and control in Ethiopia
mHealth: mobile health
ROC: receiver operating characteristic

Edited by T de Azevedo Cardoso, A Mavragani; submitted 13.10.22; peer-reviewed by A Anteneh, Y Jeem, S Pesälä; comments to author 17.04.23; revised version received 17.05.23; accepted 30.06.23; published 28.08.23.

Please cite as:
Bisrat H, Manyazewal T, Fekadu A
Mobile Health–Supported Active Syndrome Surveillance for COVID-19 Early Case Finding in Addis Ababa, Ethiopia: Comparative Study
Interact J Med Res 2023;12:e43492
URL: https://www.i-jmr.org/2023/1/e43492
doi:10.2196/43492
PMID:37556182

Original Paper

Consumer Willingness to Pay for Food Defense and Food Hygiene in Japan: Cross-Sectional Study

Shinya Matsumoto[1], PhD; Yoshiyuki Kanagawa[2], MD, PhD; Kiwamu Nagoshi[1], MD, PhD; Takemi Akahane[3], MD, PhD; Tomoaki Imamura[2], MD, PhD; Manabu Akahane[4], MD, PhD

[1]Department of Environmental Medicine and Public Health, Faculty of Medicine, Shimane University, Izumo, Japan

[2]Department of Public Health, Health Management and Policy, Nara Medical University, Kashihara, Japan

[3]Department of Gastroenterology, Nara Medical University, Kashihara, Japan

[4]Department of Health and Welfare Services, National Institute of Public Health, Wako, Japan

Corresponding Author:
Shinya Matsumoto, PhD
Department of Environmental Medicine and Public Health
Faculty of Medicine
Shimane University
Enya-cho 89-1
Izumo, 693-8501
Japan
Phone: 81 853 20 2162
Fax: 81 853 20 2160
Email: smatsumo@med.shimane-u.ac.jp

Abstract

Background: In Japan, incidents of falsified expiration dates on popular cookie brands and health hazards associated with frozen Chinese dumplings have raised food safety awareness. To prevent the intentional contamination of food by foreign substances, large food manufacturing companies have adopted the concept of food defense.

Objective: The aim of this study was to assess people's willingness to pay for food protection measures. In addition, the impact of participants' personalities and considerations regarding their purchase choices on how much they were willing to pay when shopping for food and other products were measured.

Methods: A questionnaire on willingness to pay for food hygiene and food defense was administered via a web survey and 1414 responses were included in the analysis. Univariate logistic regression analyses were performed with individuals willing and unwilling to pay additional costs as the objective variable and other questionnaire items as explanatory variables. A principal component analysis was performed on 12 questions regarding how much additional money people were willing to pay, and the principal component scores and other questions were examined for implications and other information.

Results: Approximately one-third of the respondents stated that they were unwilling to pay additional costs and reported a willingness to consume delivery food even if it contained items that were not part of the original order. The first principal component reflected the extent to which people were willing to pay additional money, and if so, how much. This tendency existed even if the individual foods and amounts varied. The third principal component reflected the amount of extra money that people were willing to pay, which was determined by the amount people had to pay toward food safety measures. Those who answered "zero" were more likely to believe that consumers should not have to pay to ensure food safety. The second principal component reflected an axis separating food defense and food hygiene. Some items not directly related to food were correlated with this axis.

Conclusions: In Japan, the concept of food hygiene is well-established and is generally taken for granted. In contrast, the concept of food defense is relatively new and has not yet fully penetrated the Japanese market. Our research shows that people who think that clothing brands provided added value to clothing products may have similar feelings about food defense. In addition, food hygiene efforts to prevent outbreaks of food poisoning are common in Japan and have been established as the basis of food safety. While food defense efforts are spreading, mainly in companies, it is presumed that they are valuable for the general public as supplementary measures to routine (or basic) food hygiene.

(Interact J Med Res 2023;12:e43936) doi:10.2196/43936

KEYWORDS

food defense; food hygiene; contingent valuation method; willingness; food; cost; awareness; food safety; questionnaire; Japan; prevention; food poisoning; safety

Introduction

Terrorism can be carried out using various means, including food terrorism. Food terrorism is the act of intentionally contaminating food using poisonous substances to achieve political goals, causing harm to people and stirring up social unrest. Food defense represents an effort to prevent the intentional contamination of food [1-6]. Since the attacks of September 11, 2001, widespread attention has been directed to the danger of so-called food terrorism worldwide and building food defense [7]. Fortunately, no large-scale terrorist attacks using food have been reported to date; however, other issues plague food safety, such as cases of food counterfeiting or the falsification of expiration dates and a food's place of origin. These issues have been reported as threats to food safety and security in Japan and elsewhere [8,9].

In addition to food hygiene measures, food defense measures are becoming more common to ensure food safety [10-13]. Food hygiene measures involve the monitoring of food and beverages to prevent foodborne health hazards. These measures include preventing spoilage by pathogenic microorganisms as well as contamination by harmful chemicals. To prevent spoilage and contamination, each stage of the food supply chain, such as production, transportation, packaging, preparation, and the sale of food and beverages, is undertaken with utmost care. In other words, the prevention of food poisoning enables maintaining a safe food supply chain while avoiding food poisoning. By contrast, food defense is a means to prevent health hazards caused by the intentional contamination of food and beverages with foreign substances [10,12,13]. Contamination in the food defense context is not accidental, such as that caused by human error, but rather involves incidents in which food or beverages are intentionally contaminated by foreign substances with malicious intent. Food defense measures are required to ensure food safety in all food aspects, both at the manufacturing level and at the retail and consumer levels.

Although not directly considered an act of food terrorism, intentional contamination or adulteration of food products with foreign substances has been reported in Japan and elsewhere [9]. In Japan, the health hazards caused by Chinese-made frozen dumplings in 2008 and the pesticide contamination of frozen foods manufactured in Japan in 2013 raised awareness for ensuring food safety during the food manufacturing process [10]. It has been approximately 10 years since efforts to prevent the intentional contamination of food were widely recognized as a food defense measure, especially by major food manufacturing companies in Japan [1].

To prevent health hazards caused by the intentional contamination of food, measures must be taken throughout the food supply chain. In other words, in addition to food defense measures adopted by food companies in food manufacturing and distribution and by retail stores and restaurants, citizens, as consumers, should also be aware of food safety. Therefore, in this study, a questionnaire survey was conducted to determine if and to what degree people are willing to pay for food safety and to identify the characteristics of people who perceive the risk of incidents that will threaten food in the future. In addition, the degree of willingness to pay to ensure food safety was surveyed. This study is the first to clarify whether and how much consumers are willing to pay for food safety.

Methods

Data Acquisition

This study was conducted as a cross-sectional survey using an internet panel survey company (Macromill Inc) and all participants were registered with the company. First, to recruit study participants, the survey company randomly sampled a list of registered participants between the ages of 15 and 79 years. Second, an email was sent to everyone on the list to inquire whether they were interested in participating in this study. Enrollment closed when the number of participants in each group by gender and age in 10-year increments reached the target sample size (103 people per group). Participants completed the survey via a web survey screen and received a small monetary reward upon survey completion. There were 1442 participants in the study; after data cleaning, 1414 participants were included in the analysis with 706 men and 708 women.

Although the Japanese population has a higher number of people in their 70s than in their 20s, all age/sex groups had an equal number of respondents in this study owing to the method used to enroll participants. We did not employ data augmentation methods because we used traditional statistical methods for the analysis. Regarding regions, there was a slightly higher number of respondents in the city prefectures of Kanto and Kansai than in other areas. The respondents' incomes were categorized in ranges: the median household income of the respondents was 6-8 million yen (100 yen=US $0.9596), which is higher than the national median household income of 4.4 million yen.

Questionnaire Items

Age, gender, weight, height, and region of residence were recorded when the participants first registered with the panel survey company.

In addition to items related to food safety, food hygiene, and food defense, the survey included questions about the participants' personalities (self-evaluation) and lifestyles (what they considered important when purchasing electrical appliances and clothes).

The value of food hygiene and food defense was evaluated using the contingent valuation method (CVM). There were six questions each pertaining to the specific additional costs to be paid for food hygiene measures (Q27) and for food defense measures (Q28) for a total of 12 questions. The unit prices of several food products were provided as examples, and the respondents were asked how much extra they would be willing to pay for food hygiene or food defense measures to be incorporated. The relevant questions are presented in Textbox 1. The unit price is expressed in Japanese yen, and the exchange rate at the time of the survey, January 2021, was 100 yen=US $0.9596.

Textbox 1. Survey questions about additional costs (100 yen=US $0.9596).

Q27. Food manufacturing companies in Japan have implemented food hygiene measures (ie, measures to prevent food poisoning), but these incur additional costs. How much extra would you be willing to pay for each of the following items owing to these measures?

(Please write “0” if you would not pay anything extra)

- 1000-yen lunch box
- 1000-yen juice
- 500-yen frozen food
- 500-yen side dish
- 500-yen snack
- 200-yen soft drink

Q28. Not all food manufacturing companies in Japan have implemented food defense measures (measures to prevent intentional contamination of food with foreign substances/drugs). Nevertheless, as these require additional expenditure, how much extra would you be willing to pay for each of the following items due to the measures?

(Please write “0” if you would not pay anything extra)

- 1000-yen lunch box
- 1000-yen juice
- 500-yen frozen food
- 500-yen side dish
- 500-yen snack
- 200-yen soft drink

Data Analysis

We performed a distribution check on the payment amount for additional costs for food defense. As many people answered “0 yen” (not willing to pay), we performed comprehensive univariate logistic regression analyses for all combinations of the 12 objective variables and other questionnaire items as explanatory variables to analyze the characteristics of this subgroup. The univariate logistic regression analysis variable was evaluated using z-scores. By averaging the z-scores for the 12 questions (Textbox 1), we considered the strength of the relationship between the other questions and willingness to pay. The z-scores of the 12 logistic regression analyses, which changed the objective variable, were averaged and arranged in order of absolute values of the average of z-scores. The z-score is a statistic affected by sampling and therefore has small variation owing to sampling, whereas large fluctuations are unlikely. In this study, the responses to the 12 similar questions were converted to binary values and logistic regression was performed with the converted values as the objective variables. Given the slight variation in z-scores among the 12 logistic regression analyses, evaluating the 12 z-scores individually was not required and comparisons were made using mean values.

Since our questions concerned amounts of money, the data distribution was often skewed to the right (positive skewness). In the case of such a distribution, a transformation is often implemented to ensure that the distribution is closer to a normal distribution. As the Box-Cox transformation cannot be defined for 0, we added 1 to the transformation so that the distribution shape would follow a normal distribution. Since λ was estimated to be approximately 0 for each variable, we set $\lambda=0$ for all 12 variables. In other words, a logarithmic transformation was performed. Next, principal component analysis was performed on the 12 variables after the Box-Cox transformation and a principal component score was calculated for each respondent.

Univariate linear regression analyses were performed for all combinations with the top three principal component scores as the objective variables and all questionnaire items not used in the principal component analysis as the explanatory variables. Consequently, items that had strong relationships were identified. The relationships between each principal component score and the questionnaire items were examined.

Ethical Considerations

This study was conducted with approval of the Ethics Committee of the National Institute of Public Health, Japan (approval number NIPH-IBRA#12302). All participants provided informed consent online for data collection and storage. Informed consent for participation in the study was obtained at the time of registration. In cases where participants were under the age of 18 years, a consent screen was displayed to obtain permission from their parents or guardians. The web-based questionnaire survey was conducted by an authorized research company compliant with personal information protection and we obtained anonymized data from the company.

Results

The survey included questions regarding how much additional cost participants would pay for certain products. However, several respondents indicated they would pay many times more than the original unit price (double or more). The respondents who answered that they would pay more than 1.5 times the product's unit price were considered outliers and excluded from the analysis. Therefore, responses for 1414 participants were included in the final analysis.

Figure 1 shows how much more money a respondent was willing to pay for each product according to the product type and unit price. Approximately 30% of respondents answered that they were not willing to pay any additional amount (0 yen) for each product. The most frequent value range, excluding 0 yen, was the same for food defense (Q28) and food hygiene (Q27): 51-100 yen for a 1000-yen lunch; 21-50 yen for a 1000-yen fruit juice drink, 500-yen frozen food, or 500-yen side dish; and 6-10 yen for a 200-yen snack and 200-yen soft drink.

Figure 1. Distribution of the amount of additional costs (100 yen=US $0.9596). JPY: Japanese yen.

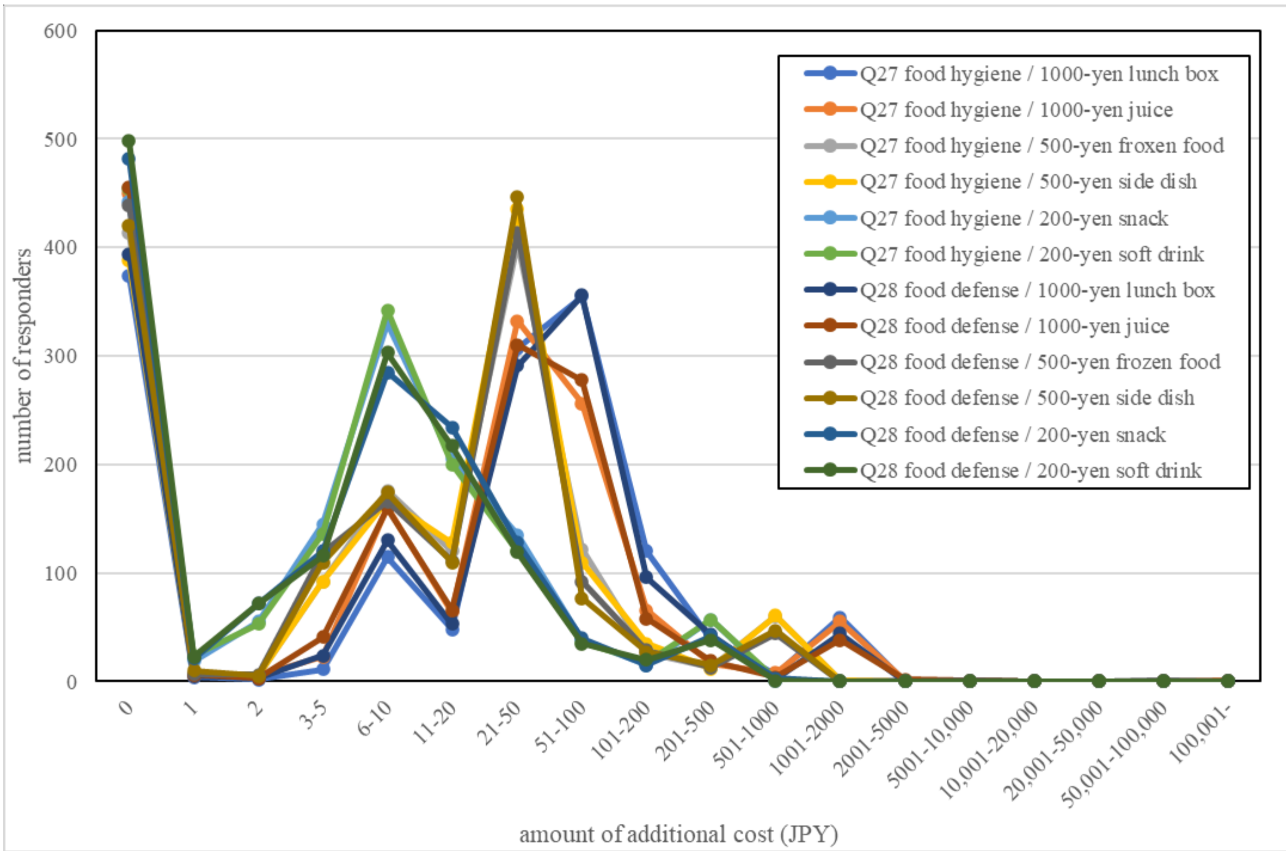


Table 1 shows the median and mean additional costs for all respondents and for the respondents after data cleaning. Focusing on the average value after data cleaning, which is often used for CVM, the values were approximately 7% to 10% of the unit price. Comparing the same products, the amount of additional cost for food hygiene (Q27) was slightly higher than that for food defense (Q28).

Table 1. Summary of additional costs for the contingent valuation analysis (100 yen=US $0.9596).

Question	All responses		After data cleaning	
	Median additional cost (yen)	Mean additional cost (yen)	Median additional cost (yen)	Mean additional cost (yen)
Q27: food hygiene				
1000-yen lunch box	50	121.80	50	101.60
1000-yen juice	30	94.77	27.5	79.38
500-yen frozen food	20	98.18	20	48.10
500-yen side dish	20	64.58	20	49.28
200-yen snack	10	24.47	10	20.71
200-yen soft drink	10	24.12	10	20.21
Q28: food defense				
1000-yen lunch box	50	103.8	50	87.85
1000-yen juice	20	7715	20	69.91
500-yen frozen food	10	108.6	10	41.57
500-yen side dish	15	47.59	10	41.33
200-yen snack	10	22.62	10	18.92
200-yen soft drink	10	20.34	8	18.13

Table 2 shows the top 30 items in the analysis results. The first item was a direct question regarding paying additional costs; however, many questionnaire items that showed positive attitudes toward food defense and food hygiene were also present. Although they were not included in the top 30 items, people who were indifferent to food safety, such as those who would eat frozen food even if it had a strange smell, were also associated with paying 0 yen.

A principal component analysis was conducted on the 12 variables related to additional cost in the two major questions Q27 and Q28. The SD of the first principal component was greater than 5 and that of the second and third principal components exceeded 1.2 and were considered to be valid. Table 3 lists the factor loadings up to the third component. All the signs were the same for the first component. For the second component, Q27 and Q28 had differing positive and negative signs, respectively. In the third component, the loadings of the questions for expensive products were positive and those for cheap food products were negative.

Table 2. Top 30 items correlated with “0 yen” responses (unwilling to pay additional cost).

Question/statement	Average z-score
Please tell us about your thoughts on safety: I will pay additional fees to ensure safety	–7.423
When making a purchase, you take into consideration product labels (marks, etc) indicating that the product has been manufactured in a factory that adopts the following measures: food hygiene measures	–5.818
The following are important for you when buying food (here, the rating would indicate the degree of importance): safety	–5.373
Please tell us about yourself: I am cooperative	–5.202
You consider the following when purchasing food (here, the rating would indicate the degree of consideration given): expiry date/best-before date	–5.053
The following are important to prevent food poisoning when eating at home: cleanliness of the place where you are cooking	–5.023
The following are important for you when buying food (here, the rating would indicate the degree of importance): manufactured domestically	–4.841
When making a purchase, you take into consideration product labels (marks, etc) indicating that the product has been manufactured in a factory that adopts the following measures: food defense measures	–4.731
The following are important to prevent food poisoning when eating at home: expiry date	–4.622
Please tell us about yourself: I have a strong sense of morality	–4.561
The following are important to prevent food poisoning when eating at home: heat cooking	–4.505
You are worried the following will increase the risk of infection when eating at a restaurant amid the COVID-19 pandemic: condiments put on the table (containers with a lid)	–4.414
You practice the following often to maintain hygiene and cleanliness: wearing a mask	–4.400
The following are important for you when buying electrical appliances (here, the rating would indicate the degree of importance): safety	–4.384
You try to obtain new information on the following topics: domestic situation	–4.375
You are worried the following will increase the risk of infection when eating at a restaurant amid the COVID-19 pandemic: buffet-style layout	–4.326
When making a purchase, you take into consideration product labels (marks, etc) indicating that the product has been manufactured in a factory that adopts the following measures: allergen labeling	–4.274
You are worried the following will increase the risk of infection when eating at a restaurant amid the COVID-19 pandemic: drinks put on the table (pitchers with a lid)	–4.160
You try to obtain new information on the following topics: international situation	–4.150
When making a purchase, you take into consideration product labels (marks, etc) indicating that the product has been manufactured in a factory that adopts the following measures: labeling of foods for specified health use (FOSHU)	–4.145
Have you heard of the following terms? (Yes/No): food security and safety	4.107
You consider the following when purchasing food (here, the rating would indicate the degree of consideration given): ingredient labeling	–4.085
You practice the following often to maintain hygiene and cleanliness: hand washing	–4.020
Please tell us about yourself: I am diligent	–3.966
Considering that there have been several incidents of intentional contamination of food with foreign substances (pesticides, etc) in Japan, such incidents will recur in the future at the following places: food-related logistics facilities	–3.819
Please tell us about yourself: I have a strong sense of responsibility	–3.756
You are anxious about eating in the following style/places amid the COVID-19 pandemic (here, the rating would indicate the degree of anxiousness): buffet	–3.743
You worry about the following at international sports events held in the summer in Japan: heat stroke	–3.722
Personal income	3.719
You try to obtain new information on the following topics: environment	–3.677

Table 3. Factor loadings from principal component analysis.

Question	1st component	2nd component	3rd component
Q27: food hygiene			
1000-yen[a] lunch box	–0.33213	0.40084	0.29172
1000-yen juice	–0.32377	0.41225	0.15839
500-yen frozen food	–0.29709	0.24143	–0.18636
500-yen side dish	–0.29346	0.25852	–0.12869
200-yen snack	–0.23692	0.13195	–0.42668
200-yen soft drink	–0.23624	0.13591	–0.41529
Q28: food defense			
1000-yen lunch box	–0.33156	–0.19234	0.45951
1000-yen juice	–0.32478	–0.20582	0.35734
500-yen frozen food	–0.29400	–0.33029	–0.01778
500-yen side dish	–0.28771	–0.32699	0.03124
200-yen snack	–0.24059	–0.33121	–0.27116
200-yen soft drink	–0.23670	–0.32140	–0.26522

[a]100 yen=US $0.9596.

Univariate linear regressions were conducted with the first to third principal components as objective variables and the remaining questionnaire items as explanatory variables. For the first principal component, there were 78 significant items (*P*≤.05). The top 30 items are listed in Table 4. For the first principal component, the strongest questionnaire item was "Please tell us about your thoughts on safety: I will pay additional fees to ensure safety." The second strongest questionnaire item was "Please tell us about yourself: I am cooperative." Other items also indicated that people who were interested in food hygiene and food defense were willing to pay additional costs. The items strongly correlated with the first principal component were considered to indicate the difference in attitudes toward food safety.

Table 4. Top 30 significant ($P≤.05$) questionnaire items in univariate linear regression for the first principal component.

Question/statement	*t* value (*df*=1412)
Please tell us about your thoughts on safety: I will pay additional fees to ensure safety	–7.4080
Please tell us about yourself: I am cooperative	–5.6810
When making a purchase, you take into consideration product labels (marks, etc) indicating that the product has been manufactured in a factory that adopts the following measures: food hygiene measures	–5.2480
The following are important for you when buying food: manufactured domestically	–5.0140
You are worried the following will increase the risk of infection when eating at a restaurant amid the COVID-19 pandemic: buffet-style layout	–4.9980
The following are important for you when buying food: safety	–4.9220
You are worried the following will increase the risk of infection when eating at a restaurant amid the COVID-19 pandemic: drinks put on the table (pitchers with a lid)	–4.8350
The following are important to prevent food poisoning when eating at home: expiry date	–4.8280
You practice the following often to maintain hygiene and cleanliness: wearing a mask	–4.7360
You are worried the following will increase the risk of infection when eating at a restaurant amid the COVID-19 pandemic: condiments put on the table (containers with a lid)	–4.7010
The following are important to prevent food poisoning when eating at home: heat cooking	–4.6800
You consider the following when purchasing food: expiry date/best-before date	–4.5550
When making a purchase, you take into consideration product labels (marks, etc) indicating that the product has been manufactured in a factory that adopts the following measures: allergen labeling	–4.4420
When making a purchase, you take into consideration product labels (marks, etc) indicating that the product has been manufactured in a factory that adopts the following measures: food defense measures	–4.3700
When making a purchase, you take into consideration product labels (marks, etc) indicating that the product has been manufactured in a factory that adopts the following measures: labeling of foods for specified health use (FOSHU)	–4.3230
You are anxious about eating in the following style/places amid the COVID-19 pandemic: buffet	–4.1600
You are worried the following will increase the risk of infection when eating at a restaurant amid the COVID-19 pandemic: disposable chopsticks (kept in a chopstick stand without packaging)	–4.1520
Have you heard of the following terms? (Yes/No): food security and safety	4.1430
The following are important to prevent food poisoning when eating at home: cleanliness of the place where you are cooking	–4.0640
Please tell us about yourself: I have a strong sense of morality	–3.9950
Have you heard of the following terms? (Yes/No): food terrorism	3.9630
The following are important for you when buying electrical appliances: safety	–3.8220
The following are important to prevent food poisoning when eating at home: best-before date	–3.7600
Age group	3.6820
If the frozen food you purchased is contaminated with foreign substances (metal, hair, etc), you…: dispose of it	–3.6720
Please tell us about yourself: I am diligent	–3.6160
The following are important for you when buying food: reputation (word of mouth)	–3.6090
The following are important for you when buying electrical appliances: reputation (word of mouth)	–3.5770
Personal income	3.5100
Age	3.4710

For the second principal component, there were 14 significant questionnaire items ($P≤.05$), which is considerably fewer than the number of significant items along the first principal component (Table 5). The strongest questionnaire item was "When you buy food, how important are the following: reputation (word of mouth)," followed by "When you buy clothes, how important are the following: price." These are suggested to be related not only to food reputation but also to perspectives when purchasing items other than food.

Table 5. Significant ($P \leq .05$) questionnaire items in univariate linear regression for the second principal component.

Question/statement	*t* value (*df*=1412)
The following are important for you when buying food: reputation (word of mouth)	–3.2050
The following are important for you when buying clothes: price	–3.0380
The following are important for you when buying food: price	–2.9940
Have you heard of the following terms? (Yes/No): food hygiene	2.6100
The following are important for you when buying food: brand	–2.4850
The frozen food you purchased has a bad (rotten, chemical, etc) smell. You…: dispose of it	–2.4530
You consider the following as “unhygienic” when eating at a restaurant: disposable chopsticks (kept in a chopstick stand without packaging)	2.4110
When making a purchase, you take into consideration product labels (marks, etc) indicating that the product has been manufactured in a factory that adopts the following measures: labeling of foods for specified health use (FOSHU)	–2.3840
When making a purchase, you take into consideration product labels (marks, etc) indicating that the product has been manufactured in a factory that adopts the following measures: food defense measures	–2.3780
The following are important for you when buying food: customer service/troubleshooting	–2.3310
The following are important for you when buying electrical appliances: brand	–2.0830
The following are important to prevent food poisoning when eating at home: order of cooking	–2.0710
When making a purchase, you take into consideration product labels (marks, etc) indicating that the product has been manufactured in a factory that adopts the following measures: food hygiene measures	–1.9830
The following are important to prevent food poisoning when eating at home: cold storage of ingredients	–1.9650

For the third principal component, there were 100 significant ($P \leq .05$) items in univariate regression; the top 30 items are listed in Table 6. The strongest item was “age.” This analysis indicated that the items constituting respondents’ basic information were more relevant than the questionnaire items regarding food safety.

Table 6. Top 30 significant items (*P*≤.05) for the third principal component in univariate linear regression.

Question statement	*t* value (*df*=1412)
Age	6.2660
Age group	5.9030
Have child (Yes/No)	4.6800
Marital status	4.1900
You find extra item(s), which you had not selected, in the food parcel delivered to you. You…: eat it without worrying	–3.5670
You consider the following when purchasing food: place (ie, country) of production	3.4980
Please tell us about yourself: I am diligent	3.4700
You find extra item(s), which you had not selected, in the food parcel delivered to you. You…: contact the shop from where you had ordered it	3.4660
You consider the following when purchasing food: expiry date/best-before date	3.4190
Please tell us about your thoughts on safety: terror acts do not occur in Japan	3.3410
Please tell us about yourself: I am honest	3.1820
The following are important for you when buying food: safety	3.0710
There have been several instances of intentional contamination of food with foreign substances/drugs in the past, and implementing measures to prevent this is called “food defense.” Considering this, such measures should be adopted for food manufactured/provided at the following places: international political event venues (summits, etc)	3.0220
You are anxious about eating in the following style/places amid the COVID-19 pandemic: bars	2.9770
There have been several instances of intentional contamination of food with foreign substances/drugs in the past, and implementing measures to prevent this is called “food defense.” Considering this, such measures should be adopted for food manufactured/provided at the following places: food service chain stores (restaurants, etc)	2.9610
The frozen food you purchased has a bad (rotten, chemical, etc) smell. You…: contact the manufacturer	2.9070
There have been several instances of intentional contamination of food with foreign substances/drugs in the past, and implementing measures to prevent this is called “food defense.” Considering this, such measures should be adopted for food manufactured/provided at the following places: international sports event venues (Olympics, etc)	2.8990
You worry about the following at international sport events held in the summer in Japan: terror attack(s)	–2.8910
There have been several instances of intentional contamination of food with foreign substances/drugs in the past, and implementing measures to prevent this is called “food defense.” Considering this, such measures should be adopted for food manufactured/provided at the following places: food factories	2.7030
You are anxious about eating in the following style/places amid the COVID-19 pandemic: event venues (stores)	2.7010
Please tell us about yourself: I have a strong sense of responsibility	2.6900
The frozen food you purchased has a bad (rotten, chemical, etc) smell. You…: eat it without worrying	–2.5700
The frozen food you purchased has a bad (rotten, chemical, etc) smell. You…: contact the shop where you purchased it	2.5530
Please tell us about yourself: I have a strong sense of morality	2.4920
The following are important for you when buying food: manufactured domestically	2.4530
You practice the following often to maintain hygiene and cleanliness: wearing a mask	2.4410
You are anxious about eating in the following style/places amid the COVID-19 pandemic: home	–2.4120
When making a purchase, you take into consideration product labels (marks, etc) indicating that the product has been manufactured in a factory that adopts the following measures: food defense measures	2.3950
You practice the following often to maintain hygiene and cleanliness: cleaning	2.3640
The following are important for you when buying food: customer service/troubleshooting	2.1760

Discussion

Principal Results

In this study, a web-based survey was conducted in Japan to examine people’s willingness to pay for food safety and food defense. Approximately one-third of the respondents stated that they were unwilling to incur any additional costs. This subgroup also reported their willingness to consume food items included in their delivery even if those items were not part of their original order.

To analyze the additional amount paid, principal component analyses were performed. The first principal component reflected whether one is willing to pay extra and if so, how much. This propensity was maintained even if individual foods and amounts of money differed and therefore does not need to be discussed further. However, the second principal component reflected an axis indicating a difference in the importance of food defense and food hygiene among respondents.

Food Hygiene and Food Defense in Japan

Individuals who expressed unwillingness to pay additional costs for food hygiene and food defense may perceive bearing the costs associated with ensuring food safety as the responsibility of businesses rather than that of consumers. The first principal component represented the amount paid, and many items strongly associated with this component were related to perceptions of safety. Some people may consider food safety as something that should be provided free of charge, whereas others are willing to pay for it.

Respondents who were unwilling to pay for additional costs were often those who would consume food included in their delivery even if it posed a risk of intentional harm. Therefore, it can be concluded that this type of respondent is less concerned about food safety. The third principal component reflected the influence of the original price. Few people indicated that they would pay a fixed amount higher than 0 yen, regardless of the product's price. Those who claimed that the price would not affect their decision may also be considered in the 0 yen category; in other words, they have no intention of spending additional money. Consequently, it can be inferred that Japanese food is perceived as safe and trustworthy [2], and respondents with this perception do not intend to increase their expenditure.

Some individuals in Japan believe that investing in safety is unnecessary. For example, Japan's tap water is inexpensive and suitable for drinking, and it is considered a safe country overall [14]. In 1970, Yamamoto (writing under the pseudonym Isaiah BenDasan) [15] astutely pointed out that "Japanese people think water and safety are free." Despite this perspective, citizens do pay for water and contribute through taxes for police services. Owing to the relatively low cost and its integration with other needs, there is a high possibility that some citizens are unaware of the necessity of investing in safety.

In Japan, both food defense and food hygiene are crucial for ensuring safe and high-quality food. Food defense involves protecting food from intentional contamination, whereas food hygiene focuses on preventing unintentional contamination. Although government regulations strictly establish standards for food hygiene, guidelines for food defense are developed and published by a research team funded by the Ministry of Health, Labour, and Welfare's scientific research grant [1,3]. In recent years, concerns have arisen regarding the intentional contamination of food products with foreign substances in Japan. Consequently, stronger food defense measures are required, and the industry is working toward addressing these issues and promoting greater transparency and traceability in the food supply chain.

Thorough measures to improve food safety in the food supply chain have become imperative recently [16,17]. As all factors at each stage of the food chain, from primary production to consumers, can affect food safety, necessary actions must be taken [18,19]. To mitigate the risk of food-related health hazards, both businesses involved in the food supply chain as well as consumers must adopt appropriate measures.

Overall, Japan has fostered a strong food safety culture and remains committed to ensuring citizen health and well-being through safe and nutritious food. However, the results of this survey indicate that some consumers may be unaware that maintaining Japan's robust food safety culture, built over many years, may entail a financial burden.

The second principal component revealed an axis that distinguishes between food hygiene and food defense. The most important factor indicating this distinction was "When buying food, the following are important to you: reputation (word of mouth)." The second most important factor was "When buying clothes, the following are important to you: price." The fifth most important factor was "When buying food, the following are important to you: brand." Based on these results, the second most important factor, "clothing," is not directly related to food, whereas the fifth factor, "brand," emphasizes the value of intangible qualities. The value of a brand has also been demonstrated by Aaker [20] and Winters [21].

In the case of agricultural products, Koike et al [22] conducted an internet survey and found that 64.5% of consumers preferred domestically produced agricultural products, 35.2% had no specific preference, and only 0.3% preferred foreign-produced products. This indicates the existence of the "domestic myth" (an emotional belief that domestic production is superior) or the perception that food produced in Japan is inherently safe, particularly regarding agricultural products. In our study's questionnaire, most respondents stated that domestic production is important when purchasing food because "domestic products are trusted by many people." Additionally, individuals who value brands in contexts other than food, such as clothing, also support food defense. We asked respondents whether they considered "manufactured domestically" to be important when buying food, with 1.3% responding "strongly disagree," 2.9% responding "disagree," 10.9% responding "somewhat disagree," 33.5% responding "somewhat agree," 31.2% responding "agree," and 20.2% responding "strongly agree." This indicates that the "domestic myth" of domestic production is prevalent in the food industry in Japan as 84.9% of respondents answered "somewhat agree" or higher. Our findings suggest that food hygiene is well-managed and recognized in Japan, while food defense is considered an added value, similar to brands.

Future Work

We intend to investigate causal relationships and other factors using analysis methods such as structural equation modeling to identify the type of consciousness the current result is based on.

Limitations

Several limitations to this study must be acknowledged. First, the data were collected only from respondents registered with an internet panel survey company, which may have biased the sample. Older people who use computers and the internet may be healthier and more active than those who do not. However, panel surveys are increasingly widely used in questionnaires

[5,23-26]. Second, respondents received a brief description of the survey via email and then chose whether to participate in the survey. Thus, there is a possibility of participant bias owing to their preferences. Third, we included an explanation of food defense in our survey because we thought that the concept of food defense was not well understood by the general public. In the future, we must clarify the difference between food defense and food hygiene and then ask the participants to respond. This may have affected some participants' responses regarding food defense. Fourth, this was not a population-based study; therefore, the participants may not be representative of the general population in Japan. Fifth, participants received a small cash reward for their participation, which may have affected the randomness of the sample. However, the research firm minimized bias by using a large enrollment population to create the survey panel (the total number of people registered with the research firm was more than 1,300,000, accounting for approximately 1% of the Japanese population). Despite these limitations, internet questionnaires can be considered representative of the general public; therefore, we believe that the study results could represent the entire Japanese population to some extent.

Conclusions

Approximately one-third of respondents said they were unwilling to incur additional costs for ensuring food safety. This subgroup also described that they would eat foods included in their delivery that they had not ordered. We performed a principal component analysis on the amount of additional money that people were willing to pay for food defense and food hygiene measures, and found an axis that indicated the divergent importance of food defense and food hygiene for consumers.

In Japan, food hygiene is taken for granted and food defense is considered of little significance by many people; it is only considered important by those who seek to gain peace of mind. Many companies are applying food defense efforts, but it is necessary to also educate consumers so that they understand the need for these measures.

Acknowledgments

This study was partially supported by a Grant-in-Aid for Scientific Research from the Ministry of Health, Labour and Welfare, Japan (21KA1009).

Data Availability

The data used in this study are available from the corresponding author upon reasonable request.

Authors' Contributions

SM designed the project, executed the analysis, and drafted the initial manuscript. YK, TI, TA, and MA developed the questionnaires. SM, KN, TI, and MA interpreted the results. MA directed and coordinated the project. All authors approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Kanagawa Y, Akahane M, Hasegawa A, Yamaguchi K, Onitake K, Takaya S, et al. Developing a national food defense guideline based on a vulnerability assessment of intentional food contamination in Japanese food factories using the CARVER+Shock vulnerability assessment tool. Foodborne Pathog Dis 2014 Dec;11(12):953-959. [doi: 10.1089/fpd.2014.1767] [Medline: 25496071]
2. Matsumoto S, Kanagawa Y, Nagoshi K, Imamura T, Akahane M. Characteristics of people in Japan based on their perceptions about food-related risks. Shimane J Med 2022;39:15-25. [doi: 10.51010/sjms.39.1_15]
3. Kanagawa Y, Akahane M, Imamura T, Hasegawa A, Yamaguchi K, Onitake K, et al. Tentative food defense guidelines for food producers and processors in Japan. Nihon Koshu Eisei Zasshi 2014;61(2):100-109 [FREE Full text] [doi: 10.11236/jph.61.2_100] [Medline: 24739880]
4. Sugiura H, Ohkusa Y, Akahane M, Sugahara T, Okabe N, Imamura T. Construction of syndromic surveillance using a web-based daily questionnaire for health and its application at the G8 Hokkaido Toyako Summit meeting. Epidemiol Infect 2010 Oct;138(10):1493-1502. [doi: 10.1017/S095026880999149X] [Medline: 20067657]
5. Sugiura H, Ohkusa Y, Akahane M, Sano T, Okabe N, Imamura T. Development of a web-based survey for monitoring daily health and its application in an epidemiological survey. J Med Internet Res 2011 Sep 23;13(3):e66 [FREE Full text] [doi: 10.2196/jmir.1872] [Medline: 21946004]
6. Maeyashiki A, Akahane M, Sugiura H, Ohkusa Y, Okabe N, Imamura T. Development and application of an alert system to detect cases of food poisoning in Japan. PLoS One 2016;11(5):e0156395 [FREE Full text] [doi: 10.1371/journal.pone.0156395] [Medline: 27231884]
7. Hennessey M, Kennedy S, Busta F. Demeter's Resilience: an International Food Defense exercise. J Food Prot 2010 Jul;73(7):1353-1356 [FREE Full text] [doi: 10.4315/0362-028x-73.7.1353] [Medline: 20615353]

8. Everstine K, Spink J, Kennedy S. Economically motivated adulteration (EMA) of food: common characteristics of EMA incidents. J Food Prot 2013 Apr;76(4):723-735 [FREE Full text] [doi: 10.4315/0362-028X.JFP-12-399] [Medline: 23575142]
9. Spink J, Moyer DC. Defining the public health threat of food fraud. J Food Sci 2011;76(9):R157-R163. [doi: 10.1111/j.1750-3841.2011.02417.x] [Medline: 22416717]
10. Jurica K, Vrdoljak J, Karačonji IB. Food defence systems as an answer to food terrorism. Arh Hig Rada Toksikol 2019 Dec 01;70(4):232-255 [FREE Full text] [doi: 10.2478/aiht-2019-70-3357] [Medline: 32623862]
11. Manning L, Soon JM. Food safety, food fraud, and food defense: a fast evolving literature. J Food Sci 2016 Apr;81(4):R823-R834 [FREE Full text] [doi: 10.1111/1750-3841.13256] [Medline: 26934423]
12. Xirasagar S, Kanwat CP, Qu H, Smith LU, Patterson NJ, Shewchuk RM. Preventing intentional food contamination: a survey to assess restaurant preparedness. J Public Health Manag Pract 2010;16(4):E7-E17. [doi: 10.1097/PHH.0b013e3181c4d950] [Medline: 20520363]
13. Xirasagar S, Kanwat CP, Smith LU, Li Y, Sros L, Shewchuk RM. Restaurant industry preparedness against intentional food contamination: results of a South Carolina survey. J Public Health Manag Pract 2010;16(4):E18-E30. [doi: 10.1097/PHH.0b013e3181c6b676] [Medline: 20520362]
14. Hino K, Schneider RH. Planning for crime prevention in Japan. Built Eviron 2013 Mar 01;39(1):114-139. [doi: 10.2148/benv.39.1.114]
15. Yamamoto S. The Japanese and the Jews. Tokyo: Yamamoto-Shoten; 1970.
16. Sarno E, Pezzutto D, Rossi M, Liebana E, Rizzi V. A review of significant European foodborne outbreaks in the last decade. J Food Prot 2021 Dec 01;84(12):2059-2070 [FREE Full text] [doi: 10.4315/JFP-21-096] [Medline: 34197583]
17. Cortés-Sánchez ADJ, Espinosa-Chaurand LD, Díaz-Ramirez M, Torres-Ochoa E. Plesiomonas: a review on food safety, fish-borne diseases, and tilapia. ScientificWorldJournal 2021;2021:3119958. [doi: 10.1155/2021/3119958] [Medline: 34594160]
18. Insfran-Rivarola A, Tlapa D, Limon-Romero J, Baez-Lopez Y, Miranda-Ackerman M, Arredondo-Soto K, et al. A systematic review and meta-analysis of the effects of food safety and hygiene training on food handlers. Foods 2020 Aug 25;9(9):1169 [FREE Full text] [doi: 10.3390/foods9091169] [Medline: 32854221]
19. Lee JH, Seo KH. An integrative review of hygiene practice studies in the food service sector. J Food Prot 2020 Dec 01;83(12):2147-2157 [FREE Full text] [doi: 10.4315/JFP-19-488] [Medline: 32692821]
20. Aaker D. The value of brand equity. J Bus Strategy 1992 Apr;13(4):27-32. [doi: 10.1108/eb039503]
21. Winters L. Brand equity measures: some recent advances. Mark Res 1991;3:70.
22. Koike N, Yamamoto Y, Demura K. Measuring Japanese consumer's evaluation on regional brand power of agricultural products: an approach using Internet research methodology. Rev Agric Econ Hokkaido Univ 2006;62:129-139.
23. Akahane M, Maeyashiki A, Tanaka Y, Imamura T. The impact of musculoskeletal diseases on the presence of locomotive syndrome. Mod Rheumatol 2019 Jan;29(1):151-156. [doi: 10.1080/14397595.2018.1452173] [Medline: 29529893]
24. Min YH, Lee JW, Shin Y, Jo M, Sohn G, Lee J, et al. Daily collection of self-reporting sleep disturbance data via a smartphone app in breast cancer patients receiving chemotherapy: a feasibility study. J Med Internet Res 2014 May 23;16(5):e135 [FREE Full text] [doi: 10.2196/jmir.3421] [Medline: 24860070]
25. Akahane M, Maeyashiki A, Yoshihara S, Tanaka Y, Imamura T. Relationship between difficulties in daily activities and falling: loco-check as a self-assessment of fall risk. Interact J Med Res 2016 Jun 20;5(2):e20 [FREE Full text] [doi: 10.2196/ijmr.5590] [Medline: 27323871]
26. Akahane T, Nakanishi Y, Yoshiji H, Akahane M. Esophagogastroduodenoscopy screening intentions during the COVID-19 pandemic in Japan: web-based survey. JMIR Cancer 2022 Nov 11;8(4):e40600 [FREE Full text] [doi: 10.2196/40600] [Medline: 36343187]

Abbreviations

CVM: contingent valuation method

Edited by G Eysenbach, T de Azevedo Cardoso; submitted 07.11.22; peer-reviewed by J Nielsen; comments to author 31.12.22; revised version received 21.02.23; accepted 29.09.23; published 23.10.23.

Please cite as:
Matsumoto S, Kanagawa Y, Nagoshi K, Akahane T, Imamura T, Akahane M
Consumer Willingness to Pay for Food Defense and Food Hygiene in Japan: Cross-Sectional Study
Interact J Med Res 2023;12:e43936
URL: https://www.i-jmr.org/2023/1/e43936
doi:10.2196/43936
PMID:37870896

Original Paper

Evaluating the Effectiveness of School Closure in COVID-19–Related Syndromes From Community-Based Syndromic Surveillance: Longitudinal Observational Study

Ping-Chen Chung[1], MSc, DDS; Kevin J Chen[2], EdD; Hui-Mei Chang[2], MEd; Ta-Chien Chan[3,4], PhD

[1]Department of Dentistry, Puzi Hospital, Ministry of Health and Welfare, Chiayi County, Taiwan

[2]Department of Health, Taipei City Government, Taipei City, Taiwan

[3]Research Center for Humanities and Social Sciences, Academia Sinica, Taipei City, Taiwan

[4]Institute of Public Health, School of Medicine, National Yang Ming Chiao Tung University, Taipei City, Taiwan

Corresponding Author:
Ta-Chien Chan, PhD
Research Center for Humanities and Social Sciences
Academia Sinica
128 Academia Road, Section 2
Taipei City, 115
Taiwan
Phone: 886 227898160
Email: tachien@sinica.edu.tw

Abstract

Background: During the COVID-19 pandemic, a school closure policy was adopted to prevent cluster transmission in schools and subsequent household transmission. However, the effectiveness of school closure is not consistent in studies conducted in different countries.

Objective: This study aimed to explore the association between school closure and the daily standardized incidence of COVID-19–related syndromes in an outpatient syndromic surveillance system.

Methods: We calculated the incidence of COVID-19–related syndromes derived from a community-based syndromic surveillance system between the first week of January and the second or fourth weeks after school closure in 2021 and 2022 in Taipei City, Taiwan. The effect of school closure on the standardized incidence of COVID-19–related syndromes was evaluated by interrupted time series analysis using an autoregressive integrated moving average with a distributed lag function. The exogenous variables were changes in human mobility measured by Google COVID-19 community mobility reports. Furthermore, the models quantified the influence of different age groups and the hierarchy of medical facilities, such as clinics or community hospitals.

Results: School closure was only negatively and significantly associated with the overall standardized incidence of COVID-19–related syndromes in 2021 for 2 weeks after the intervention (coefficient −1.24, 95% CI −2.40 to −0.08). However, in different age groups, school closure had a significantly negative association with the standardized incidence among people aged 13-18 years and ≥65 years for 2 weeks after the intervention in clinics in 2021. In community hospitals, school closure was significantly positively associated with the standardized incidence among people aged 19-24 years in 2021. In 2022, 2 weeks after the intervention, school closure had a significantly negative association with the standardized incidence among people aged 0-6, 7-12, and 19-24 years in community hospitals and aged >45 years in clinics. Furthermore, the standardized incidence was positively associated with movement change toward grocery and pharmacy stores in all age groups in 2022. In addition, movement changes toward residences were significantly positively associated with the standardized incidence among all age groups.

Conclusions: Overall, school closure effectively suppresses COVID-19–related syndromes in students owing to the reduction of physical contact. In addition, school closure has a spillover effect on elderly people who stay at home.

(Interact J Med Res 2023;12:e44606) doi:10.2196/44606

KEYWORDS

school closure; COVID-19; syndromic surveillance; outpatient; mobility

Introduction

Before the massive COVID-19 vaccination campaign for school children, school closure was the most adopted nonpharmaceutical policy to prevent cluster transmission in schools and subsequently household transmission. However, the effectiveness of school closure is not consistent among studies conducted in different countries [1]. With the adaptation of SARS-CoV-2 to humans, the highly attacked population has shifted from elderly people and adults to school children [2]. As epidemiological characteristics change, the evaluation of the school closure policy may be different. One study conducted during the early phase of the pandemic in 2020 [3] examined the effect of policies, including school closure, on reducing local transmission of COVID-19. At that time, the majority of the infected population was adults. Therefore, the effectiveness of school closure was not significant. However, as the pandemic progressed, the low vaccination coverage of school children and the reopening of schools [4] caused the infection rate among school children to surge. In the United States, children aged <18 years accounted for 17.3% of reported cases of COVID-19 until August 30, 2022 [5]. In Taiwan, up to calendar week 31 in 2022, the proportion of confirmed cases of COVID-19 among those aged under 20 years was 13.9% [6]. The attack rate and susceptibility to COVID-19 infection are lower in children than in adults according to current evidence [7]. Current research suggests that COVID-19 angiotensin-converting enzyme 2 receptors are scarce in the respiratory tract of children, leading to fewer receptors for viruses to bind to [8]. Although there appear to be few confirmed cases among children, some data suggest that children are more likely to be asymptomatic or mild and that they are not initially tested, which leads to underreported cases [9]. Children may still be a source of infection transmission [9].

The prevention of infection and severe complications, especially in children, remains a public health issue. Nonpharmaceutical interventions (NPIs) are effective strategies to mitigate the spread of epidemics in the community. Strategies include maintaining personal hygiene, wearing masks, social distancing, daily temperature measurements, and symptom monitoring. School closure or class suspension as an NPI is often considered an option for suppressing the spread of influenza and enterovirus epidemics [10-12]. School closure can reduce transmission first among children and then in the community [13]. A published systematic review noted that the mean reduction in the peak of the influenza epidemic was 29.65% (SD 23.63) after implementing school closure [10]. The earlier implementation of school closure is more likely to reduce and delay the peak of the influenza epidemic [10]. School closure can decrease the number of daily contacts between schoolmates in close proximity. A simulation study showed that reproduction number and age-specific susceptibility to infection influenced the policy effect of school closure [14]. School closure is less effective under the condition of a larger reproduction number (>1.8) and a population with less susceptibility to infection [14]. Greater transmissibility (reproduction number) and lower susceptibility to COVID-19 in children are opposed to influenza. The effects of school closure to combat COVID-19 at different times and in different countries are still disputed [1].

A recent systematic review revealed that the impact of school closure on COVID-19 is smaller than that of other social distancing interventions, and school closure alone only prevents 2%-4% of deaths [1]. In the fall of 2020 in Croatia, an association was found between school closure and COVID-19 morbidity and mortality, while in the winter of 2021, the association was insignificant [15]. According to a United Nations International Children's Emergency Fund (UNICEF) report, from March 2020 to February 2021, schools for more than 168 million children worldwide were closed for nearly a year [16]. During the COVID-19 pandemic, school closure policies have been widely implemented worldwide to reduce exposure rates and transmission risks ahead of mass vaccination. According to a systematic review of school closure during the pandemic, school closure and in-school mitigation measures were associated with reduced COVID-19 transmission in the community [17]. However, they also noted that assessing the impact of school closure is challenging because many nonpharmacological interventions are being implemented simultaneously, and the potential negative impacts on students' mental and physical health were also widely discussed [17,18]. School closure can negatively influence children's physical and mental health, and education, and can have an economic impact on working parents. Therefore, before deciding on the implementation of school closure, we must consider its effects and costs. The timing of the intervention to mitigate emerging infectious diseases was referenced by prompt surveillance. Sentinel-based surveillance for monitoring disease activity has been widely used in different countries. Sentinel surveillance can detect early aberrations in the daily incidence of a disease or the outpatient consultation rate [19]. In this study, we built community-based syndromic surveillance (ie, Sentinel Plus) in sentinel clinics and community hospitals since June 2018 in Taipei City, Taiwan. The advantages of Sentinel Plus can be used to monitor a variety of syndromes and detect changes across 7 age groups and health care facility levels, including community hospitals and clinics [19]. Sentinel Plus is designed for daily automated symptom monitoring of 34 current symptom groups in an outpatient setting to complement emergency room surveillance. In our previous study [19], Sentinel Plus performed better than other surveillance systems in early aberration detection in influenza-like illnesses and enterovirus-like syndromes. Owing to the lack of empirical evidence on the influence of the COVID-19 epidemic during the implementation of school closure, this study aimed to explore the association between school closure and the standardized incidence of COVID-19–related syndromes derived from daily outpatient syndromic surveillance data. Furthermore, we elucidated the association between 7 different age stratifications and the hierarchy of medical facilities. Moreover, the effects of school closure were explored in different extents of mobility changes between 2021 and 2022.

Methods

Data Source

We collected specific syndromic groups daily from a community-based enhanced sentinel surveillance system named "Sentinel Plus," which has been designed for the early detection of aberrations of epidemics in clinics and community hospitals since June 2018 [19]. After November 2020, Sentinel Plus monitors expanded the syndromic groups from 23 to 34 syndromes owing to the COVID-19 pandemic. The specific International Classification of Diseases, 10th Revision (ICD-10) diagnoses from the hospital information systems of participating clinics and community hospitals were computed on-site and aggregated into 34 syndromic groups and 7 age groups without any patient identifiers. The aggregated data were then sent to Sentinel Plus through a secure channel.

In this study, we used COVID-19–related syndromes from Sentinel Plus. ICD-10 diagnoses of daily visits from December 2020 to June 2022 were obtained from 130 participating clinics, 12 health centers, and 8 community hospitals in Taipei City. The definition of COVID-19–related syndromes was discussed with the family physician and infectious disease physician, and was therefore identified by ICD-10 codes, including R05, R06.02, R50.9, R43.0, R43.1, R43.2, R43.8, R43.9, R19.7, J06.9, J12.89, J34.89, R07.0, R51, and R19.7. Although the confirmed diagnosis of COVID-19 was implemented in the system in November 2020, reimbursement for the diagnosis of COVID-19 through the routine national health insurance system began at the end of May 2021. From June 2021 to late May 2022, COVID-19 testing by polymerase chain reaction (PCR) was only implemented in designated hospitals and community screening stations. Clinics did not make a diagnosis of COVID-19 but transferred possible patients to hospitals and screening stations during that time. Therefore, we did not use a COVID-19–confirmed diagnosis as the primary outcome.

In addition to the surveillance data, we also incorporated the policy of school closure in Taipei City and the Google COVID-19 Community Mobility Report for analysis. School closure information was obtained from a press release from the Department of Education of Taipei City Government. Daily percentage changes in people's visits to and staying time in the 6 categories of places were obtained from the COVID-19 Community Mobility Report [20]. The data charted movement trends over time compared to baseline days, which was the median value for the corresponding day of the week during the 5-week period from January 3 to February 6, 2020. Categories for grouping places with similar characteristics for social distancing guidance included grocery stores and pharmacies, parks, transit stations, retail and recreation, residences, and workplaces. Daily alert levels for COVID-19 from the Central Epidemic Command Center were downloaded from the Taiwan Centers for Disease Control (CDC) [21] as a reference. In Sentinel Plus, the age group was classified into 7 categories as follows: 0-6, 7-12, 13-18, 19-24, 25-44, 45-64, and ≥65 years.

Ethics Approval

The study was approved by the Institutional Review Board of the Biomedical Science Research, Academia Sinica (AS-IRB-BM-18017).

Data Analysis

In Taipei City, school closure and distance learning at home were implemented from May 18, 2021, to July 2, 2021. The time intervals from January 1, 2021, to June 1, 2021 (T1-1) and June 15, 2021 (T1-2) were analyzed. In 2022, Taipei's school closure was from May 23 to June 5. The time intervals from January 1, 2022, to June 6, 2022 (T2-1) and June 20, 2022 (T2-2) were analyzed. We assumed that the school closure on COVID-19–related syndromes lasted 1-14 days, 2-15 days, or 3-16 days after the intervention. We explored the effect of school closure on the standardized incidence of COVID-19–related syndromes, which were derived from Sentinel Plus by autoregressive integrated moving average interrupted time series-distributed lag (ARIMAITS-DL) [22], and quantified the influence of different age groups and the hierarchy of medical facilities, such as clinics and community hospitals. Moreover, the effects of school closure under movement changes between 2021 and 2022 were explored. Incidence was defined as the daily counts for the diagnosis of COVID-19–related syndromes divided by the total daily outpatient visits multiplied by 1000. Patients may be diagnosed with multiple ICD codes in the same syndromic group, and thus, the incidence may be >1000. Owing to the low number of outpatient visits on weekends and holidays, we removed daily visits below 200 from our analysis and in our plot. The ARIMAITS-DL model with exogenous variables simultaneously models both the unclear intervention time and the distributed effect of the intervention over time. It is assumed that the intervention effect is uniformly distributed over time. We checked the collinearity between the 6 categories of grouping places from the Google Community Mobility Report and deleted categories, including retail, recreation, and transit stations, owing to a variance inflation factor >10. Parks were also excluded because parks usually refer to official national parks and not the general outdoors. Workplaces were excluded because children and retirees rarely go there. Exogenous variables were grocery stores, pharmacies, and residences. The regression parameters were estimated using the maximum likelihood method. The appropriate autoregressive integrated moving average (ARIMA) parameters (p,d,q) were selected using the Akaike information criterion (AIC) and model fitting with different age groups and hierarchies of medical facilities. We checked the sensitivity of the results by changing the duration of the school closure effect with model-fitting statistics using the AIC. Finally, the residuals of the selected model and autocorrelation were tested using a residual plot and the Ljung-Box test, respectively. If autocorrelation still existed in the residuals, different autoregressive or moving average orders were chosen. All analyses were performed using R software (version 4.2.1) [23], including the forecast [24] and ggplot2 [25] packages.

Results

In the first observation period of 2021 (T1-1 and T1-2), the alert level of COVID-19 was elevated to Level 3 on May 19, 2021. In the second observation period in the first half of 2022 (T2-1 and T2-2), the alert level of COVID-19 was Level 2 (Figure 1). The standardized incidence of the diagnosis of COVID-19–related syndromes separately from clinics and community hospitals in the observed period is illustrated in Figure 1. The epi-curve varied regularly on workdays and weekends, and the trend has been decreasing since school closure, particularly in clinics, in 2021.

The upper panel represents data from clinics, and the lower panel represents data from community hospitals. The color of the first row represents 3 different alert levels in Taipei City issued by the Central Epidemic Command Center (CECC) of Taiwan. Blue represents Level 1, green represents Level 2, and red represents Level 3.

The epi-curves marked between the 2 red dashed lines are the periods of school closure. Daily visits below 200 (weekends and spring festivals) were removed from the plot. Removed dates in clinics: February 12, 2021; February 01, 2022; and February 02, 2022. Removed dates in community hospitals: December 20, 2020; February 14, 2021; July 25, 2021; September 26, 2021; October 11, 2021; November 21, 2021; December 05, 2021; December 12, 2021; January 09, 2022; and February 13, 2022.

The median incidence of COVID-19–related syndromes was higher in clinics than in community hospitals in 2021 and 2022 (Tables 1 and 2, respectively). Before the implementation of school closure, the median incidence of COVID-19–related syndromes was higher in 2021 than in 2022 in both clinics and community hospitals. In different age groups, the incidence of patients aged <18 years was relatively high in clinics. In contrast, in community hospitals, the incidence of patients aged over 25 years was relatively high in 2021 and 2022.

In the study time interval of 2021 (T1-1), the overall standardized incidence of COVID-19–related syndromes was not significantly associated with the intervention of school closure in clinics (coefficient −0.84, 95% CI −1.69 to 0.01), but was negatively associated in community hospitals (coefficient −1.24, 95% CI −2.40 to −0.08) (Table 3).

In different age groups, school closure had a significant negative association with the standardized incidence among people aged 13-18 years and ≥65 years in clinics. School closure had a significant positive association with the standardized incidence among people aged 19-24 years in community hospitals. Furthermore, the standardized incidence had significant positive associations with the change in movement toward grocery and pharmacy stores among people aged 7-12 years and 25-44 years in clinics. Furthermore, changes in movement toward residences were significantly positively associated with standardized incidence rates in all age groups, in the 13-18 and >25 age groups in clinics, and in all age groups in community hospitals. However, changes in movement toward residences were significantly negatively associated with the standardized incidence among people aged 7-18 years in community hospitals.

In the study time interval of 2021 (T1-2), the overall standardized incidence of COVID-19–related syndromes in both clinics and community hospitals was not significantly associated with the school closure intervention (clinic: coefficient −0.42, 95% CI −1.34 to 0.50; community hospital: coefficient −0.45, 95% CI −1.65 to 0.76) (Table 4). In different age groups, school closure was significantly negatively associated with the standardized incidence among people aged 0-6, 13-18, and ≥65 years in clinics. However, school closure had a significant positive association with the standardized incidence among people aged 19-24 years in community hospitals.

In the first half of 2022 (T2-1), school closure had no relationship with the overall standardized incidence in both clinics and community hospitals (clinic: coefficient −0.40, 95% CI −0.90 to 0.10; community hospital: coefficient −0.29, 95% CI −0.60 to 0.03) (Table 5). In the 45-64 and ≥65 age groups in clinics, school closure had a significant negative association with the standardized incidence. In the 0-6, 7-12, and 19-24 age groups in community hospitals, school closure had a significant negative association with the standardized incidence. Among people of all ages and those aged >25 years in clinics and people of all ages and those aged >45 years in community hospitals, changes in movement toward grocery and pharmacy stores were positively associated with the standardized incidence. In clinics, among people of all ages and those aged >19 years, movement changes toward residences had a significantly positive association with the standardized incidence. In community hospitals, among people of all ages and those aged 0-6, 7-12, 19-24, and >45 years, movement changes toward residences had a significantly positive association with the standardized incidence.

In the first half of 2022 (T2-2), school closure had no relation with the overall standardized incidence in both clinics and community hospitals (clinic: coefficient −0.29, 95% CI −0.77 to 0.19; community hospital: coefficient −0.19, 95% CI −0.48 to 0.10) (Table 6). In the 45-64 and ≥65 age groups in clinics, school closure had a significant negative association with the standardized incidence.

In 2021, during a small-scale epidemic with a soft lockdown policy in effect from May to July, school closure effectively reduced the COVID-19 incidence among preschool students, junior and senior high school students, and elders. During both time intervals, there was an increase in mobility to residences, grocery stores, and pharmacies based on the Google Community Report [20].

Figure 1. Time series plot of COVID-19–related syndromes in Taipei City.

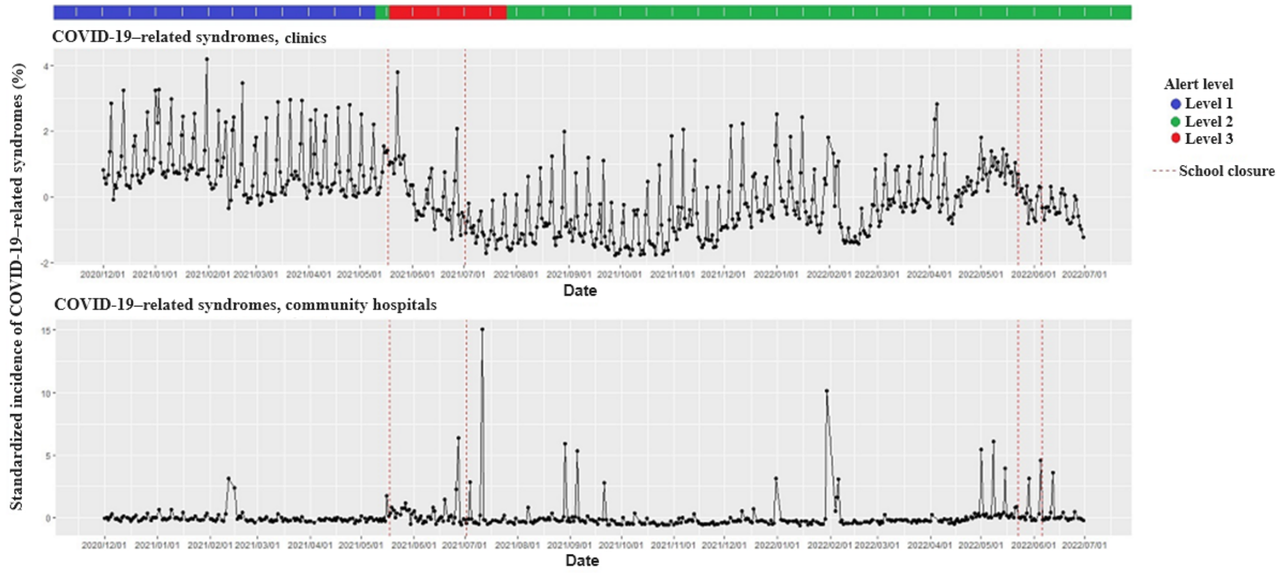


Table 1. Comparison of the daily incidence of the diagnosis of COVID-19–related syndromes before and after school closure between clinics and community hospitals in 2021.

Age	Before school closure in 2021		Two weeks after school closure in 2021		Four weeks after school closure in 2021	
	Clinic, median (IQR)	Hospital, median (IQR)	Clinic, median (IQR)	Hospital, median (IQR)	Clinic, median (IQR)	Hospital, median (IQR)
Overall	59.88 (19.22)	1.78 (0.47)	64.93 (12.35)	3.46 (0.97)	54.10 (22.89)	2.53 (1.71)
0-6 years	108.67 (18.29)	0.49 (0.29)	108.36 (21.32)	0.23 (0.45)	98.40 (26.99)	0.00 (0.24)
7-12 years	126.15 (25.89)	0.34 (0.31)	127.85 (29.46)	0.00 (0.31)	127.85 (22.71)	0.00 (0.06)
13-18 years	101.23 (32.37)	0.25 (0.55)	108.70 (48.43)	0.00 (0.47)	94.59 (39.02)	0.00 (0.06)
19-24 years	85.14 (19.58)	0.59 (0.49)	99.34 (42.19)	1.61 (2.06)	77.27 (59.28)	1.01 (1.69)
25-44 years	71.04 (14.90)	2.83 (1.09)	74.96 (21.83)	7.05 (2.17)	67.29 (38.91)	6.14 (3.20)
45-64 years	42.01 (16.98)	3.54 (1.47)	50.84 (7.48)	6.55 (1.11)	48.01 (17.37)	5.86 (3.84)
≥65 years	36.90 (20.91)	4.16 (1.79)	48.07 (11.50)	6.57 (2.93)	46.62 (13.66)	5.05 (3.89)

Table 2. Comparison of the daily incidence of the diagnosis of COVID-19–related syndromes before and after school closure between clinics and community hospitals in 2022.

Age	Before school closure in 2022		Two weeks after school closure in 2022		Four weeks after school closure in 2022	
	Clinic, median (IQR)	Hospital, median (IQR)	Clinic, median (IQR)	Hospital, median (IQR)	Clinic, median (IQR)	Hospital, median (IQR)
Overall	46.82 (18.10)	2.24 (1.91)	46.20 (10.97)	2.10 (0.55)	43.16 (7.01)	1.99 (0.34)
0-6 years	117.46 (17.09)	0.42 (0.27)	90.98 (45.42)	0.25 (0.38)	85.54 (17.40)	0.37 (0.26)
7-12 years	117.05 (30.00)	0.22 (0.29)	78.59 (52.81)	0.66 (0.25)	83.12 (48.34)	0.42 (0.30)
13-18 years	85.41 (40.21)	0.27 (0.58)	30.56 (68.81)	0.75 (0.44)	38.22 (42.48)	0.49 (0.50)
19-24 years	51.65 (32.97)	0.39 (0.66)	50.36 (24.44)	0.42 (0.32)	48.75 (8.29)	0.42 (0.33)
25-44 years	50.68 (16.11)	3.07 (1.36)	59.42 (15.25)	4.16 (1.71)	54.02 (11.25)	4.16 (1.72)
45-64 years	31.49 (20.66)	3.17 (2.19)	49.00 (10.89)	3.75 (1.55)	39.27 (11.46)	3.92 (1.25)
≥65 years	32.10 (21.02)	3.48 (1.61)	40.48 (11.62)	4.37 (0.64)	35.34 (7.12)	4.37 (0.89)

Table 3. Effect of school closure on the standardized incidence of COVID-19–related syndromes in clinics and community hospitals from January 1, 2021, to June 1, 2021.

Age and variable	Clinic (estimation), coefficient (95% CI)	Hospital (estimation), coefficient (95% CI)
Overall		
ARIMA[a] (p,d,q)	ARIMA (9,0,0)	ARIMA (3,0,0)
School closure	−0.84 (−1.69 to 0.01)	−1.24 (−2.40 to −0.08)[b]
Grocery and pharmacy percent change from baseline	0.01 (−0.01 to 0.03)	−0.01 (−0.04 to 0.01)
Residential percent change from baseline	0.10 (0.06 to 0.14)[b]	0.22 (0.16 to 0.27)[b]
0-6 years		
ARIMA (p,d,q)	ARIMA (7,0,0)	ARIMA (7,0,1)
School closure	−0.01 (−1.00 to 0.97)	−0.19 (−1.06 to 0.68)
Grocery and pharmacy percent change from baseline	0.02 (−0.004 to 0.04)	0.01 (−0.01 to 0.03)
Residential percent change from baseline	0.04 (−0.004 to 0.09)	−0.04 (−0.08 to 0.001)
7-12 years		
ARIMA (p,d,q)	ARIMA (2,0,8)	ARIMA (2,0,7)
School closure	−0.08 (−1.30 to 1.14)	0.81 (−0.27 to 1.88)
Grocery and pharmacy percent change from baseline	0.03 (0.002 to 0.06)[b]	0.02 (−0.001 to 0.05)
Residential percent change from baseline	0.003 (−0.05 to 0.06)	−0.12 (−0.17 to −0.07)[b]
13-18 years		
ARIMA (p,d,q)	ARIMA (0,1,1)	ARIMA (0,0,0)
School closure	−1.58 (−2.90 to −0.26)[b]	0.56 (−0.70 to 1.83)
Grocery and pharmacy percent change from baseline	−0.02 (−0.05 to 0.01)	0.03 (−0.004 to 0.05)
Residential percent change from baseline	0.07 (0.01 to 0.12)[b]	−0.09 (−0.16 to −0.03)[b]
19-24 years		
ARIMA (p,d,q)	ARIMA (1,0,0)	ARIMA (5,0,0)
School closure	0.99 (−0.31 to 2.29)	2.49 (1.63 to 3.34)[b]
Grocery and pharmacy percent change from baseline	0.02 (−0.01 to 0.05)	0.001 (−0.02 to 0.02)
Residential percent change from baseline	0.03 (−0.02 to 0.08)	−0.003 (−0.05 to 0.04)
25-44 years		
ARIMA (p,d,q)	ARIMA (0,1,7)	ARIMA (0,0,1)
School closure	0.65 (−0.52 to 1.83)	0.14 (−0.99 to 1.27)
Grocery and pharmacy percent change from baseline	0.05 (0.02 to 0.08)[b]	−0.02 (−0.04 to 0.01)
Residential percent change from baseline	0.05 (0.001 to 0.09)[b]	0.01 (−0.04 to 0.06)
45-64 years		
ARIMA (p,d,q)	ARIMA (6,1,2)	ARIMA (1,0,3)
School closure	−0.17 (−1.17 to 0.82)	0.96 (−0.22 to 2.15)
Grocery and pharmacy percent change from baseline	0.01 (−0.01 to 0.04)	0.01 (−0.02 to 0.03)
Residential percent change from baseline	0.10 (0.06 to 0.14)[b]	0.06 (−0.0001 to 0.11)
≥65 years		
ARIMA (p,d,q)	ARIMA (8,0,0)	ARIMA (0,0,0)
School closure	−1.30 (−2.06 to −0.55)[b]	0.68 (−0.58 to 1.93)
Grocery and pharmacy percent change from baseline	−0.002 (−0.02 to 0.01)	−0.02 (−0.05 to 0.01)

Age and variable	Clinic (estimation), coefficient (95% CI)	Hospital (estimation), coefficient (95% CI)
Residential percent change from baseline	0.15 (0.12 to 0.19)[b]	0.06 (−0.004 to 0.12)

[a]ARIMA: autoregressive integrated moving average.

[b]*P*<.05.

Table 4. Effect of school closure on the standardized incidence of COVID-19–related syndromes in clinics and community hospitals from January 1, 2021, to June 15, 2021.

Age and variable	Clinic (estimation), coefficient (95% CI)	Hospital (estimation), coefficient (95% CI)
Overall		
ARIMA[a] (p,d,q)	ARIMA (6,1,2)	ARIMA (1,0,2)
School closure	−0.42 (−1.34 to 0.50)	−0.45 (−1.65 to 0.76)
Grocery and pharmacy percent change from baseline	0.03 (0.01 to 0.06)[b]	−0.01 (−0.04 to 0.01)
Residential percent change from baseline	0.08 (0.05 to 0.12)[b]	0.05 (−0.02 to 0.11)
0-6 years		
ARIMA (p,d,q)	ARIMA (9,0,0)	ARIMA (7,0,1)
School closure	−0.90 (−1.73 to −0.06)[b]	−0.43 (−1.14 to 0.27)
Grocery and pharmacy percent change from baseline	0.01 (−0.01 to 0.03)	0.01 (−0.01 to 0.03)
Residential percent change from baseline	0.03 (−0.01 to 0.07)	−0.04 (−0.07 to 0.00)
7-12 years		
ARIMA (p,d,q)	ARIMA (2,0,8)	ARIMA (0,0,6)
School closure	−0.05 (−1.00 to 0.90)	0.61 (−0.30 to 1.52)
Grocery and pharmacy percent change from baseline	0.02 (−0.01 to 0.05)	0.01 (−0.02 to 0.04)
Residential percent change from baseline	0.01 (−0.03 to 0.06)	−0.10 (−0.14 to −0.05)[b]
13-18 years		
ARIMA (p,d,q)	ARIMA (2,0,6)	ARIMA (0,0,0)
School closure	−1.43 (−2.43 to −0.43)[b]	0.42 (−0.59 to 1.43)
Grocery and pharmacy percent change from baseline	0.02 (−0.01 to 0.05)	0.02 (−0.01 to 0.05)
Residential percent change from baseline	0.05 (0.005 to 0.10)[b]	−0.07 (−0.12 to −0.02)[b]
19-24 years		
ARIMA (p,d,q)	ARIMA (1,0,0)	ARIMA (5,0,0)
School closure	−0.39 (−1.37 to 0.59)	1.44 (0.45 to 2.42)[b]
Grocery and pharmacy percent change from baseline	0.03 (0.005 to 0.06)[b]	0.01 (−0.01 to 0.04)
Residential percent change from baseline	0.01 (−0.04 to 0.05)	−0.04 (−0.09 to 0.005)
25-44 years		
ARIMA (p,d,q)	ARIMA (0,1,10)	ARIMA (0,0,1)
School closure	−0.21 (−1.47 to 1.05)	0.10 (−0.77 to 0.97)
Grocery and pharmacy percent change from baseline	0.05 (0.03 to 0.08)[b]	−0.02 (−0.04 to 0.01)
Residential percent change from baseline	0.01 (−0.02 to 0.05)	0.00 (−0.04 to 0.04)
45-64 years		
ARIMA (p,d,q)	ARIMA (6,1,2)	ARIMA (5,0,0)
School closure	−0.01 (−1.13 to 1.11)	0.26 (−0.67 to 1.18)
Grocery and pharmacy percent change from baseline	0.03 (0.01 to 0.06)[b]	0.02 (−0.01 to 0.04)
Residential percent change from baseline	0.07 (0.03 to 0.12)[b]	0.02 (−0.03 to 0.06)
≥65 years		
ARIMA (p,d,q)	ARIMA (7,0,0)	ARIMA (0,0,0)
School closure	−1.29 (−2.47 to −0.11)[b]	−0.21 (−1.24 to 0.83)
Grocery and pharmacy percent change from baseline	0.01 (−0.01 to 0.03)	0.00 (−0.03 to 0.03)

Age and variable	Clinic (estimation), coefficient (95% CI)	Hospital (estimation), coefficient (95% CI)
Residential percent change from baseline	0.12 (0.09 to 0.14)[b]	0.02 (−0.03 to 0.07)

[a]ARIMA: autoregressive integrated moving average.
[b]*P*<.05.

Table 5. Effect of school closure on the standardized incidence of COVID-19–related syndromes in clinics and community hospitals from January 1, 2022, to June 6, 2022.

Age and variable	Clinic (estimation), coefficient (95% CI)	Hospital (estimation), coefficient (95% CI)
Overall		
ARIMA[a] (p,d,q)	ARIMA (8,0,0)	ARIMA (2,0,2)
School closure	−0.40 (−0.90 to 0.10)	−0.29 (−0.60 to 0.03)
Grocery and pharmacy percent change from baseline	0.03 (0.01 to 0.05)[b]	0.04 (0.02 to 0.06)[b]
Residential percent change from baseline	0.09 (0.04 to 0.14)[b]	0.14 (0.09 to 0.18)[b]
0-6 years		
ARIMA (p,d,q)	ARIMA (0,0,3)	ARIMA (5,0,0)
School closure	−0.23 (−0.63 to 0.18)	−0.09 (−0.18 to −0.01)[b]
Grocery and pharmacy percent change from baseline	0.01 (−0.01 to 0.03)	−0.01 (−0.01 to 0.000003)
Residential percent change from baseline	0.03 (−0.001 to 0.07)	0.03 (0.02 to 0.04)[b]
7-12 years		
ARIMA (p,d,q)	ARIMA (5,0,2)	ARIMA (6,0,2)
School closure	−0.05 (−0.54 to 0.45)	−0.25 (−0.50 to −0.01)[b]
Grocery and pharmacy percent change from baseline	0.00 (−0.02 to 0.02)	0.00 (−0.02 to 0.01)
Residential percent change from baseline	0.04 (−0.002 to 0.08)	0.07 (0.03 to 0.10)[b]
13-18 years		
ARIMA (p,d,q)	ARIMA (0,1,9)	ARIMA (10,0,0)
School closure	−0.35 (−0.74 to 0.04)	−0.05 (−0.32 to 0.23)
Grocery and pharmacy percent change from baseline	0.01 (−0.01 to 0.03)	0.01 (−0.01 to 0.02)
Residential percent change from baseline	−0.01 (−0.06 to 0.04)	0.02 (−0.02 to 0.05)
19-24 years		
ARIMA (p,d,q)	ARIMA (6,0,2)	ARIMA (7,0,0)
School closure	−0.31 (−0.80 to 0.17)	−0.38 (−0.74 to −0.03)[b]
Grocery and pharmacy percent change from baseline	−0.01 (−0.03 to 0.01)	−0.01 (−0.03 to 0.02)
Residential percent change from baseline	0.06 (0.02 to 0.10)[b]	0.09 (0.03 to 0.15)[b]
25-44 years		
ARIMA (p,d,q)	ARIMA (3,0,5)	ARIMA (7,0,0)
School closure	−0.23 (−0.68 to 0.22)	0.06 (−0.28 to 0.41)
Grocery and pharmacy percent change from baseline	0.02 (0.001 to 0.04)[b]	−0.01 (−0.03 to 0.01)
Residential percent change from baseline	0.07 (0.03 to 0.11)[b]	0.02 (−0.03 to 0.06)
45-64 years		
ARIMA (p,d,q)	ARIMA (6,1,0)	ARIMA (0,0,1)
School closure	−0.46 (−0.77 to −0.15)[b]	−0.22 (−0.48 to 0.05)
Grocery and pharmacy percent change from baseline	0.04 (0.02 to 0.05)[b]	0.04 (0.02 to 0.06)[b]
Residential percent change from baseline	0.10 (0.06 to 0.14)[b]	0.10 (0.06 to 0.13)[b]
≥65 years		
ARIMA (p,d,q)	ARIMA (7,0,0)	ARIMA (0,0,4)
School closure	−0.42 (−0.73 to −0.12)[b]	0.02 (−0.33 to 0.36)

Age and variable	Clinic (estimation), coefficient (95% CI)	Hospital (estimation), coefficient (95% CI)
Grocery and pharmacy percent change from baseline	0.04 (0.02 to 0.05)[b]	0.04 (0.01 to 0.06)[b]
Residential percent change from baseline	0.10 (0.06 to 0.14)[b]	0.08 (0.04 to 0.12)[b]

[a]ARIMA: autoregressive integrated moving average.

[b]*P*<.05.

Table 6. Effect of school closure on the standardized incidence of COVID-19–related syndromes in clinics and community hospitals from January 1, 2022, to June 20, 2022.

Age and variable	Clinic (estimation), coefficient (95% CI)	Hospital (estimation), coefficient (95% CI)
Overall		
ARIMA[a] (p,d,q)	ARIMA (8,0,0)	ARIMA (0,0,8)
School closure	−0.29 (−0.77 to 0.19)	−0.19 (−0.48 to 0.10)
Grocery and pharmacy percent change from baseline	0.03 (0.02 to 0.05)[b]	0.04 (0.02 to 0.06)[b]
Residential percent change from baseline	0.09 (0.04 to 0.14)[b]	0.13 (0.09 to 0.17)[b]
0-6 years		
ARIMA (p,d,q)	ARIMA (7,0,0)	ARIMA (10,0,0)
School closure	−0.08 (−0.51 to 0.35)	−0.05 (−0.20 to 0.09)
Grocery and pharmacy percent change from baseline	0.00 (−0.01 to 0.02)	−0.01 (−0.02 to 0.00)
Residential percent change from baseline	0.05 (0.01 to 0.10)[b]	0.02 (0.01 to 0.04)[b]
7-12 years		
ARIMA (p,d,q)	ARIMA (7,0,0)	ARIMA (8,0,2)
School closure	0.05 (−0.39 to 0.48)	−0.29 (−0.61 to 0.04)
Grocery and pharmacy percent change from baseline	−0.01 (−0.02 to 0.01)	0.003 (−0.01 to 0.02)
Residential percent change from baseline	0.04 (0.001 to 0.08)[b]	0.02 (−0.01 to 0.06)
13-18 years		
ARIMA (p,d,q)	ARIMA (2,0,1)	ARIMA (10,0,0)
School closure	−0.26 (−0.70 to 0.19)	−0.02 (−0.31 to 0.26)
Grocery and pharmacy percent change from baseline	0.01 (−0.01 to 0.03)	0.01 (−0.01 to 0.02)
Residential percent change from baseline	0.00 (−0.04 to 0.04)	0.01 (−0.02 to 0.05)
19-24 years		
ARIMA (p,d,q)	ARIMA (3,0,3)	ARIMA (7,0,0)
School closure	−0.31 (−0.74 to 0.13)	−0.30 (−0.68 to 0.08)
Grocery and pharmacy percent change from baseline	−0.01 (−0.03 to 0.01)	−0.01 (−0.03 to 0.01)
Residential percent change from baseline	0.06 (0.03 to 0.10)[b]	0.03 (−0.01 to 0.08)
25-44 years		
ARIMA (p,d,q)	ARIMA (5,0,3)	ARIMA (7,0,0)
School closure	−0.17 (−0.61 to 0.28)	0.10 (−0.23 to 0.44)
Grocery and pharmacy percent change from baseline	0.02 (0.001 to 0.04)[b]	−0.01 (−0.03 to 0.01)
Residential percent change from baseline	0.06 (0.02 to 0.10)[b]	0.01 (−0.03 to 0.05)
45-64 years		
ARIMA (p,d,q)	ARIMA (6,1,0)	ARIMA (1,0,1)
School closure	−0.42 (−0.73 to −0.11)[b]	−0.14 (−0.38 to 0.11)
Grocery and pharmacy percent change from baseline	0.04 (0.02 to 0.05)[b]	0.04 (0.02 to 0.06)[b]
Residential percent change from baseline	0.10 (0.06 to 0.14)[b]	0.08 (0.05 to 0.11)[b]
≥65 years		
ARIMA (p,d,q)	ARIMA (7,0,0)	ARIMA (0,0,4)
School closure	−0.39 (−0.68 to −0.09)[b]	0.01 (−0.32 to 0.35)
Grocery and pharmacy percent change from baseline	0.04 (0.02 to 0.05)[b]	0.04 (0.01 to 0.06)[b]

Age and variable	Clinic (estimation), coefficient (95% CI)	Hospital (estimation), coefficient (95% CI)
Residential percent change from baseline	0.11 (0.07 to 0.15)[b]	0.08 (0.04 to 0.12)[b]

[a]ARIMA: autoregressive integrated moving average.

[b]*P*<.05.

Discussion

Principal Findings

School closure was only significantly negatively associated with the overall standardized incidence of COVID-19–related syndromes in 2021 for 2 weeks after the intervention in community hospitals. However, within different age groups, the effects of school closure were not consistent. Among people aged 13-18 years and ≥65 years for 2 and 4 weeks after the intervention, school closure had a significantly negative association with the standardized incidence of COVID-19–related syndromes in 2021 in clinics. There were some suspected reasons for the effects of school closure. First, the scale of the epidemic was small in 2021, and the proportion of COVID-19–confirmed cases in children was low. School closure, including kindergarten closure, can prevent infection among students and children aged 2-6 years [26]. Second, both the COVID-19 alert level in 2021 and soft lockdown reduced population movements and COVID-19–confirmed cases especially among people aged ≥65 years. Social distancing actions contribute to the reduction of the number of new COVID-19 cases [27]. In Taiwan, some children live with not only their parents but also their grandparents. According to a 2017 report of the Senior Citizen Condition Survey, adults aged 55-64 years who take care of grandchildren account for 16.05% of the same age group, while daily care for grandchildren accounts for 10.79%. Adults aged 65 years or older who care for grandchildren account for 11.80% of the same age group [28]. School closure, which keeps students at home, has been linked to decreased COVID-19 incidence owing to this family structure. In community hospitals, school closure had a significantly positive association with the standardized incidence among people aged 19-24 years. We suspected that distance learning let college students have more freedom to arrange their time and have more opportunities for social contact. However, the relation between the contact rate and infection rate has to be explored. In a United States study, adults aged 20-34 years and 35-49 years had sustained SARS-CoV-2 transmission with reproduction numbers consistently above 1 as of October 29, 2020. The high transmission rates among adults aged 20-49 years are linked to both rebounding mobility over the summer and elevated transmission risk per venue visit [29].

In community hospitals in 2022, 2 weeks after the intervention among people aged 0-6, 7-12, and 19-24 years, the school closure had a significantly negative association with the standardized incidence of COVID-19–related syndromes. Because COVID-19 diagnoses were only given from hospitals or community screening stations before May 2022 and children died from COVID-19, the proportion of children visiting hospitals was high. Therefore, it is easier to observe the effects of school closure.

In 2022, school closure did not have an obvious effect on students after 4 weeks of the intervention. However, in clinics for people aged >45 years in 2022, school closure was significantly negatively associated with the standardized incidence of COVID-19–related syndromes. First, in the commuting zone of Taipei and New Taipei City, school closure was mandated for just 2 weeks, and the announcement was made by the Taipei government. Residents or students could still commute across cities and possibly become infected from the workplace or from crammed schools. Second, there was a general outbreak of epidemics, with the number of confirmed cases of COVID-19 rising after May 2022, which was tallied from Taiwan CDC [30]. Third, the alert level of COVID-19 was reduced from Level 3 to Level 2, and NPIs were less strict. Although school closure had no significant effect on students, it still had a protective effect on family members. A study in Sweden reported that exposure to open rather than closed schools resulted in a small increase in PCR-confirmed infections among parents (odds ratio 1.17, 95% CI 1.03-1.32) [31]. During school closure, students stay home, which reduces the risk of contact with schoolmates and indirectly reduces the risk of infection of family members, especially those who stay longer at home. School closure affects family routines, modified work schedules, remote work, and alternative childcare [32].

In terms of the hierarchy of medical facilities, children and young adults are more likely to visit clinics, while older people with comorbidities are more likely to visit community hospitals [33]. In addition, community hospitals are more likely to have cross-district patients. In Sentinel Plus, the proportion of patients diagnosed with COVID-19–related syndromes from community hospitals was low.

For people’s movement changes in 2021 and 2022 in different age groups, a longer grocery and pharmacy time was associated with a higher standardized incidence of COVID-19–related syndromes. This may be because early symptoms of infection are not obvious, and people may go to pharmacies to seek medication [34].

For people’s movement changes in 2022 in different age groups, a longer residence time was associated with a higher standardized incidence of COVID-19–related syndromes. First, the high rate of community infection and confirmed cases in 2022 could have easily caused household infections. This family structure increases the probability of household secondary transmission, and school closure, which keeps students at home, has been linked to increased COVID-19 incidence. A meta-analysis estimated the household secondary attack rate to be 16.6% (95% CI 14.0-19.3), which is higher than the secondary attack rates for SARS-CoV and MERS-CoV. The rates of household secondary attacks were higher for contacts with adults than for contacts with children [35]. Second, the alert level of COVID-19 was reduced from Level 3 to Level 2, relaxing some of the NPI policies.

During school closure in 2021, vaccinations were not administered. According to Taiwan CDC statistics [36], the vaccination rate for the first dose was 16.9% and that for the second dose was 0% among children aged 6-11 years. On May 23, 2022, the vaccination rate for the second dose was 82.1% among teenagers aged 12-17 years in Taipei City.

Since Taiwan did not implement a strict lockdown during the COVID-19 pandemic, Taiwanese citizens may have different fears, anxieties, and concerns compared to the citizens of other countries that implemented strict lockdown policies, leading to different attitudes and behaviors regarding infection risks [37]. Additionally, social media, as a platform for rapidly spreading information during disease outbreaks, may consequently influence public perceptions and emotions, potentially triggering behavior changes in response to the COVID-19 pandemic [38]. Under Taiwan's soft lockdown policy, social and family compliance with epidemic prevention policies has played a great role in controlling the local epidemic. Overall, school closure was effective in suppressing COVID-19 among students due to reduced contact. A recent Taiwanese study constructed parameters of interpersonal contact patterns from diary-approach surveys. The results showed that the most active age group was schoolchildren (aged 5-14 years), with an average of 16-18 contacts [39]. Reducing schoolchildren's contact rate can effectively reduce disease transmission. In addition, school closure has a spillover effect on elderly people who stay home. A systematic review of school closure studies showed substantial heterogeneity, with half of the studies at a lower risk of bias reporting a reduction in community transmission of up to 60% [40].

It is difficult to determine the specific impact of school closure on COVID-19 incidence when multiple nonpharmacological interventions are implemented simultaneously. Analysis of the 2022 time interval (the 10 months in which alert Level 2 lasted before the school closure policy was implemented) helps highlight the impact of school closure in Taipei City. Our outpatient syndromic surveillance system can divide the entire population into 7 age groups, allowing the differential impacts of school closure on different age groups to be explored. Specifically, because of the Chinese family structure with a higher proportion of grandparents, parents, and grandchildren living together in the same house, elderly people benefit from the protective effects of school closure.

Comparison With Prior Work

Although some studies have explored the effect of school closure on COVID-19 infection, we focused on the effect in 7 different age groups with different severities of the epidemic. Furthermore, we used community-based syndromic surveillance to monitor immediate changes in COVID-19–related syndromes. Moreover, the movement of people to grocery stores, pharmacies, and residences was also considered.

Staguhn et al [26] noted that school closure significantly impacted COVID-19 infection rates in the United States. However, not all states were included in this study. Only states with state-wide stay-at-home orders, stay-at-home order dates preceding school closure dates, less than 3 days between the date of 10 cases and the school closure date, and less than 3 days between the school closure date and the stay-at-home order were included in the study. Stay-at-home orders that coincided with school closure may have influenced the effects observed in the data.

Another study analyzed nasopharyngeal swab-confirmed SARS-CoV-2 cases in Milan from February 1 to April 5, 2021. Total school closure in Lombardy was imposed between March 5 and April 5, 2021. The school closure intervention produced a daily reduction in the risk of contagion by 4% in those aged 3-11 years and 12-19 years, and by 3% in those aged ≥20 years. The epidemic curve decreased first in school-aged children after the intervention and subsequently in the adult population [41].

Limitations

There are some limitations in this study. First, we used the standardized incidence of COVID-19–related syndromes as an indicator and not confirmed cases of COVID-19. Before the end of May 2022, COVID-19 cases were only diagnosed by COVID-19 PCR tests from community screening stations and emergency departments of hospitals. The COVID-19–related syndromic group was first established in November 2020. Although COVID-19 presenting symptoms changed with the evolution of coronavirus variants, we discussed with infectious disease physicians and family doctors a few months after the COVID-19 pandemic in 2020, based on their clinical observations and studies, to define susceptible patients at that time. Some symptoms may vary with variants; however, it remains difficult to make a differential diagnosis. For our purposes, if we could find one symptom of a COVID-19–related syndrome, we accounted for it in our model. In the future, a dynamic definition of syndrome groups is needed to update the clinical manifestations of the disease. Second, during the periods of school closure in 2021 and 2022, positive rates for influenza or respiratory viral infections were nearly zero, and positive cases for respiratory viral infections in 2021 throughout the year were very low (828 cases collected in contracted laboratories for viral infection symptoms) [42]. COVID-19–related syndromes are important diagnoses of potential infections in the community. Third, the data set of people's movement included national data, but Sentinel Plus was only for Taipei. Not everyone, especially children, had mobile devices. The accuracy of the children's movement changes needs to be further explored. Fourth, potential confounders, such as family structures, vaccination status, citizens' perceptions and emotions, and the influence of media, which can influence social networks, infection rates, immunity levels, and behavior changes, were not considered owing to the limitations of aggregated surveillance data. However, our intervention period of school closure was short, and confounding factors may not have changed much during the intervention period.

Conclusions

Nationwide school closure, encouragement to work from home, and COVID-19 alert Level 3 were implemented in 2021. Although vaccine coverage was zero in the first observation period in 2021 among the student and preschool populations, the effect of school closure was observed among them. In 2022, the COVID-19 alert was downgraded to Level 2, there was no consistent implementation of school closure policies in each county or city, and people could not work from home unless they requested unpaid family care leave. However, an effect of school closure was also found in the student population after 2 weeks of the intervention. We concluded that the early stage of the COVID-19 outbreak and the short intervention period (2 weeks) effectively prevented and postponed student infection. However, if the epidemic is widespread, school closure alone is not effective in reducing transmission.

Dynamic and real-time surveillance can offer practical policy recommendations to local health departments, enabling them to proactively address the risk of disease transmission within communities. Various policy tools are currently in use to control epidemics, including health education via social media and mass media, as well as measures like class suspension and school closure. By using a community-based surveillance system, aberrant signals can be accurately tailored to specific age groups and geographical locations, allowing for early detection and affording additional time for the implementation of further interventions. This proposed approach is not limited to COVID-19 and can be extended to manage other infectious diseases such as influenza and enterovirus.

Acknowledgments

We gratefully thank the Department of Health, Taipei City Government for providing administrative support for recruiting participatory clinics, and Taipei Medical Association for supporting this project. This work was supported by a grant from the Ministry of Science and Technology, Taiwan (grant number: MOST 109-2420-H-001-006) and a grant titled “Implementing clinic-based disease surveillance network” from Taipei City Government. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability

The data used are provided on the open repository Figshare [43].

Authors' Contributions

TCC and KJC designed the study. PCC reviewed the literature, refined the data, performed statistical analysis, and wrote the draft of the manuscript. TCC built up the surveillance system, and collected and stored the data. TCC, KJC, and HMC helped interpret the results and edit the manuscript. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Viner RM, Russell SJ, Croker H, Packer J, Ward J, Stansfield C, et al. School closure and management practices during coronavirus outbreaks including COVID-19: a rapid systematic review. Lancet Child Adolesc Health 2020 May;4(5):397-404 [FREE Full text] [doi: 10.1016/S2352-4642(20)30095-X] [Medline: 32272089]
2. Jing Q, Liu M, Zhang Z, Fang L, Yuan J, Zhang A, et al. Household secondary attack rate of COVID-19 and associated determinants in Guangzhou, China: a retrospective cohort study. Lancet Infect Dis 2020 Oct;20(10):1141-1150 [FREE Full text] [doi: 10.1016/S1473-3099(20)30471-0] [Medline: 32562601]
3. Chung P, Chan T. Impact of physical distancing policy on reducing transmission of SARS-CoV-2 globally: Perspective from government's response and residents' compliance. PLoS One 2021;16(8):e0255873 [FREE Full text] [doi: 10.1371/journal.pone.0255873] [Medline: 34375342]
4. Koppa V, West J. School reopenings, COVID-19, and employment. Econ Lett 2022 Mar;212:110310 [FREE Full text] [doi: 10.1016/j.econlet.2022.110310] [Medline: 35132288]
5. Total number of cases of COVID-19 in the United States as of April 26, 2023, by age group. Statista. URL: https://www.statista.com/statistics/1254271/us-total-number-of-covid-cases-by-age-group/ [accessed 2023-12-05]
6. Regional age and gender statistics table - COVID-19 - statistics by day of onset (in weeks). Taiwan Centers for Disease Control. URL: https://data.gov.tw/dataset/118038 [accessed 2023-12-05]
7. Goldstein E, Lipsitch M, Cevik M. On the Effect of Age on the Transmission of SARS-CoV-2 in Households, Schools, and the Community. J Infect Dis 2021 Feb 13;223(3):362-369 [FREE Full text] [doi: 10.1093/infdis/jiaa691] [Medline: 33119738]
8. Carsetti R, Quintarelli C, Quinti I, Piano Mortari E, Zumla A, Ippolito G, et al. The immune system of children: the key to understanding SARS-CoV-2 susceptibility? Lancet Child Adolesc Health 2020 Jun;4(6):414-416 [FREE Full text] [doi: 10.1016/S2352-4642(20)30135-8] [Medline: 32458804]

9. Perez-Lopez A, Hasan M, Iqbal M, Janahi M, Roscoe D, Tang P. Dramatic decrease of laboratory-confirmed influenza A after school closure in response to COVID-19. Pediatr Pulmonol 2020 Sep;55(9):2233-2234 [FREE Full text] [doi: 10.1002/ppul.24933] [Medline: 32598576]
10. Bin Nafisah S, Alamery AH, Al Nafesa A, Aleid B, Brazanji NA. School closure during novel influenza: A systematic review. J Infect Public Health 2018;11(5):657-661 [FREE Full text] [doi: 10.1016/j.jiph.2018.01.003] [Medline: 29396256]
11. Cauchemez S, Ferguson NM, Wachtel C, Tegnell A, Saour G, Duncan B, et al. Closure of schools during an influenza pandemic. Lancet Infect Dis 2009 Aug;9(8):473-481 [FREE Full text] [doi: 10.1016/S1473-3099(09)70176-8] [Medline: 19628172]
12. Lee P, Tsai T, Huang Y, Wu C, Hu Y, Lin T. Effectiveness of case isolation and class suspension in mitigation of enterovirus transmission in children. J Infect Public Health 2022 May;15(5):594-598 [FREE Full text] [doi: 10.1016/j.jiph.2022.04.010] [Medline: 35500544]
13. Luca GD, Kerckhove KV, Coletti P, Poletto C, Bossuyt N, Hens N, et al. The impact of regular school closure on seasonal influenza epidemics: a data-driven spatial transmission model for Belgium. BMC Infect Dis 2018 Jan 10;18(1):29 [FREE Full text] [doi: 10.1186/s12879-017-2934-3] [Medline: 29321005]
14. Ciavarella C, Fumanelli L, Merler S, Cattuto C, Ajelli M. School closure policies at municipality level for mitigating influenza spread: a model-based evaluation. BMC Infect Dis 2016 Oct 18;16(1):576 [FREE Full text] [doi: 10.1186/s12879-016-1918-z] [Medline: 27756233]
15. Simetin IP, Svajda M, Ivanko P, Dimnjakovic J, Belavic A, Istvanovic A, et al. COVID-19 incidence, hospitalizations and mortality trends in Croatia and school closures. Public Health 2021 Sep;198:164-170 [FREE Full text] [doi: 10.1016/j.puhe.2021.07.030] [Medline: 34461449]
16. COVID-19: Schools for more than 168 million children globally have been completely closed for almost a full year, says UNICEF. UNICEF. URL: https://www.unicef.org/press-releases/schools-more-168-million-children-globally-have-been-completely-closed [accessed 2023-12-05]
17. Hume S, Brown SR, Mahtani KR. School closures during COVID-19: an overview of systematic reviews. BMJ Evid Based Med 2023 Jun 31;28(3):164-174. [doi: 10.1136/bmjebm-2022-112085] [Medline: 37001966]
18. Viner R, Russell S, Saulle R, Croker H, Stansfield C, Packer J, et al. School Closures During Social Lockdown and Mental Health, Health Behaviors, and Well-being Among Children and Adolescents During the First COVID-19 Wave: A Systematic Review. JAMA Pediatr 2022 Apr 01;176(4):400-409. [doi: 10.1001/jamapediatrics.2021.5840] [Medline: 35040870]
19. Chan T, Tang J, Hsieh C, Chen KJ, Yu T, Tsai Y. Approaching precision public health by automated syndromic surveillance in communities. PLoS One 2021;16(8):e0254479 [FREE Full text] [doi: 10.1371/journal.pone.0254479] [Medline: 34358241]
20. COVID-19 Community Mobility Reports. Google. URL: https://www.google.com/covid19/mobility/ [accessed 2023-12-05]
21. Press Releases from Taiwan Centers for Disease Control. Taiwan Centers for Disease Control. URL: https://www.cdc.gov.tw/En [accessed 2023-12-05]
22. Yoneoka D, Kawashima T, Tanoue Y, Nomura S, Eguchi A. Distributed lag interrupted time series model for unclear intervention timing: effect of a statement of emergency during COVID-19 pandemic. BMC Med Res Methodol 2022 Jul 25;22(1):202 [FREE Full text] [doi: 10.1186/s12874-022-01662-1] [Medline: 35879679]
23. R: A language and environment for statistical computing. R Project. URL: https://www.r-project.org/ [accessed 2023-12-05]
24. Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, et al. forecast: Forecasting functions for time series and linear models. Rob J Hyndman. URL: https://pkg.robjhyndman.com/forecast/ [accessed 2023-12-05]
25. Wickham H. ggplot2: Elegant Graphics for Data Analysis. New York, NY: Springer; 2016.
26. Staguhn ED, Weston-Farber E, Castillo RC. The impact of statewide school closures on COVID-19 infection rates. Am J Infect Control 2021 Apr;49(4):503-505 [FREE Full text] [doi: 10.1016/j.ajic.2021.01.002] [Medline: 33428984]
27. Fan C, Lee S, Yang Y, Oztekin B, Li Q, Mostafavi A. Effects of population co-location reduction on cross-county transmission risk of COVID-19 in the United States. Appl Netw Sci 2021;6(1):14 [FREE Full text] [doi: 10.1007/s41109-021-00361-y] [Medline: 33623817]
28. Zimmerman FJ, Anderson NW. Association of the Timing of School Closings and Behavioral Changes With the Evolution of the Coronavirus Disease 2019 Pandemic in the US. JAMA Pediatr 2021 May 01;175(5):501-509 [FREE Full text] [doi: 10.1001/jamapediatrics.2020.6371] [Medline: 33616635]
29. Monod M, Blenkinsop A, Xi X, Hebert D, Bershan S, Tietze S, Imperial College COVID-19 Response Team. Age groups that sustain resurging COVID-19 epidemics in the United States. Science 2021 Mar 26;371(6536):eabe8372 [FREE Full text] [doi: 10.1126/science.abe8372] [Medline: 33531384]
30. Taiwan National Infectious Disease Statistics System. URL: https://nidss.cdc.gov.tw/ [accessed 2023-12-05]
31. Vlachos J, Hertegård E, B Svaleryd H. The effects of school closures on SARS-CoV-2 among parents and teachers. Proc Natl Acad Sci U S A 2021 Mar 02;118(9):e2020834118 [FREE Full text] [doi: 10.1073/pnas.2020834118] [Medline: 33574041]
32. Auger KA, Shah SS, Richardson T, Hartley D, Hall M, Warniment A, et al. Association Between Statewide School Closure and COVID-19 Incidence and Mortality in the US. JAMA 2020 Sep 01;324(9):859-870 [FREE Full text] [doi: 10.1001/jama.2020.14348] [Medline: 32745200]

33. Shih C, Weng C, Chen W, Yang H, Fan S. Consideration factors of older adults seeking medical treatment at outpatient services in Taiwan. BMC Health Serv Res 2021 Nov 09;21(1):1216 [FREE Full text] [doi: 10.1186/s12913-021-07251-0] [Medline: 34753473]
34. Wei-En T, Yong-Hao Z. A study of the correlation between transboundary infectious diseases and national healthcare system: Taking COVID-19 for example. Review of Global Politics 2021(74):111-133 [FREE Full text]
35. Madewell ZJ, Yang Y, Longini IM, Halloran ME, Dean NE. Household Transmission of SARS-CoV-2: A Systematic Review and Meta-analysis. JAMA Netw Open 2020 Dec 01;3(12):e2031756 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.31756] [Medline: 33315116]
36. COVID-19 Vaccine Statistics. Taiwan Centers for Disease Control. URL: https://www.cdc.gov.tw/Category/Page/9jFXNbCe-sFK9EImRRi2Og [accessed 2023-12-05]
37. Sharif Nia H, She L, Kaur H, Boyle C, Khoshnavay Fomani F, Hoseinzadeh E, et al. A Predictive Study Between Anxiety and Fear of COVID-19 With Psychological Behavior Response: The Mediation Role of Perceived Stress. Front Psychiatry 2022;13:851212 [FREE Full text] [doi: 10.3389/fpsyt.2022.851212] [Medline: 35392391]
38. Gozzi N, Tizzani M, Starnini M, Ciulla F, Paolotti D, Panisson A, et al. Collective Response to Media Coverage of the COVID-19 Pandemic on Reddit and Wikipedia: Mixed-Methods Analysis. J Med Internet Res 2020 Oct 12;22(10):e21597 [FREE Full text] [doi: 10.2196/21597] [Medline: 32960775]
39. Fu Y, Wang D, Chuang J. Representative contact diaries for modeling the spread of infectious diseases in Taiwan. PLoS One 2012;7(10):e45113 [FREE Full text] [doi: 10.1371/journal.pone.0045113] [Medline: 23056193]
40. Walsh S, Chowdhury A, Braithwaite V, Russell S, Birch JM, Ward JL, et al. Do school closures and school reopenings affect community transmission of COVID-19? A systematic review of observational studies. BMJ Open 2021 Aug 17;11(8):e053371 [FREE Full text] [doi: 10.1136/bmjopen-2021-053371] [Medline: 34404718]
41. Consolazio D, Sarti S, Terraneo M, Celata C, Russo AG. The impact of school closure intervention during the third wave of the COVID-19 pandemic in Italy: Evidence from the Milan area. PLoS One 2022;17(7):e0271404 [FREE Full text] [doi: 10.1371/journal.pone.0271404] [Medline: 35819972]
42. 2001-2021 genotypes of respiratory viruses. Taiwan Centers for Disease Control. URL: https://www.cdc.gov.tw/Category/MPage/pHp0os_Z-5NEVFdOCdrOQQ [accessed 2023-12-05]
43. COVID-19 related syndrome data from Sentinel plus in Taipei City, Taiwan. Figshare. URL: https://figshare.com/articles/dataset/COVID-19_related_syndrome_data_from_Sentinel_plus_in_Taipei_City_Taiwan/21587520 [accessed 2023-12-05]

Abbreviations

AIC: Akaike information criterion
ARIMA: autoregressive integrated moving average
ARIMAITS-DL: autoregressive integrated moving average interrupted time series- distributed lag
CDC: Centers for Disease Control
ICD-10: International Classification of Diseases, 10th Revision
NPI: nonpharmaceutical intervention
PCR: polymerase chain reaction

Edited by G Eysenbach; submitted 25.11.22; peer-reviewed by B Azhari, LK Chen, Y Yuan; comments to author 22.09.23; revised version received 27.09.23; accepted 30.11.23; published 15.12.23.

Please cite as:
Chung PC, Chen KJ, Chang HM, Chan TC
Evaluating the Effectiveness of School Closure in COVID-19–Related Syndromes From Community-Based Syndromic Surveillance: Longitudinal Observational Study
Interact J Med Res 2023;12:e44606
URL: https://www.i-jmr.org/2023/1/e44606
doi:10.2196/44606
PMID:38100192

Review

Closed Endotracheal Suction Systems for COVID-19: Rapid Review

Carmen Amaia Ramírez-Torres[1,2], PhD; Félix Rivera-Sanz[3], PhD; Teresa Sufrate-Sorzano[1,2], MsC; Azucena Pedraz-Marcos[4], PhD; Ivan Santolalla-Arnedo[1,2], PhD

[1]Health and Care Research Group, University of La Rioja, Logroño, Spain

[2]Research Unit on Health System Sustainability, Biomedical Center of La Rioja, Logroño, Spain

[3]Unit of Data, Biomedical Center of La Rioja, Logroño, Spain

[4]Department of Nursing, University Autonoma of Madrid, Madrid, Spain

Corresponding Author:
Carmen Amaia Ramírez-Torres, PhD
Health and Care Research Group
University of La Rioja
88, Duquesa de la Victoria
Logroño, 26004
Spain
Phone: 34 941299050
Email: caramit@unirioja.es

Abstract

Background: The increase in admissions to intensive care units (ICUs) in 2020 and the morbidity and mortality associated with SARS-CoV-2 infection pose a challenge to the analysis of evidence of health interventions carried out in ICUs. One of the most common interventions in patients infected with the virus and admitted to ICUs is endotracheal aspiration. Endotracheal suctioning has also been considered one of the most contaminating interventions.

Objective: This review aims to analyze the benefits and risks of endotracheal suctioning using closed suction systems (CSS) in COVID-19 patients.

Methods: A rapid review was carried out using the following databases: PubMed, MEDLINE, CINAHL, LILACS, the Cochrane Library, and IBECS. The data search included articles in English and Spanish, published between 2010 and 2020, concerning adult patients, and using the key words "endotracheal," "suction," and "closed system."

Results: A total of 15 articles were included. The benefits and risks were divided into 3 categories: patient, care, and organization. Relating to the patient, we found differences in cardiorespiratory variables and changes in the ventilator, for example, improvement in patients with elevated positive and end-expiratory pressure due to maladaptation and alveolar collapse. Relating to care, we found a shorter suctioning time, by up to 1 minute. Relating to organization, we found fewer microorganisms on staff gloves. Other conflicting results between studies were related to ventilator-associated pneumonia, bacterial colonization, or mortality.

Conclusions: Aside from the need for quality research comparing open suction systems and CSS as used to treat COVID-19 patients, closed endotracheal suctioning has benefits in terms of shorter stay in the ICU and reduced environmental contamination, preventing ventilator disconnection from the patient, reducing the suctioning time—though it does produce the greatest number of mucosal occlusions—and preventing interpatient and patient-staff environmental contamination. New evidence in the context of the SARS-CoV-2 virus is required in order to compare results and establish new guidelines.

(Interact J Med Res 2023;12:e42549) doi:10.2196/42549

KEYWORDS

endotracheal suctioning; closed suction system; rapid review; suction; mechanical ventilation; COVID-19; intensive care unit; health intervention; endotracheal; patient care; healthcare; ventilator; health benefit

Introduction

The effects of a disease such as COVID-19 have a global reach and can be a severe hindrance to society. Among patients diagnosed with COVID-19, 5% require admission to an intensive care unit (ICU), and, of these, 88% require mechanical ventilation (MV) to support their breathing [1].

COVID-19 is caused by the SARS-CoV-2 virus, transmitted by aerosols. During ICU admission, the patient is in the symptomatic phase of the disease, with a significant viral load, and can pose a significant health risk due to this airborne transmission, particularly to health care professionals because of the type of procedures used for patient stabilization and clinical recovery [2].

Understanding the infection mechanisms of SARS-CoV-2 requires studies of the procedures and interventions that cause greater risk of aerosol expansion. A systematic review by Jackson et al [2] found 14 procedures that are widely recognized as important generators of aerosols, including, most importantly, intubation and extubation, suction of the airways, bronchoscopy, and noninvasive ventilation.

Endotracheal suctioning is one of the most common procedures in patients intubated in an ICU. The intervention requires essential care in the form of oxygenation before suctioning, at the time of suction, and after suctioning. These procedures are performed by nurses. Endotracheal suctioning requires specialized staff, as, though it is a common procedure, it can occasionally cause harm to the patient. The types of harm directly associated with endotracheal suctioning include 6 that are particularly important when managing critical patients: asynchrony with the ventilator, hypoxia, hemodynamic alterations, atelectasis, pain, and damage to the tracheal mucosa [3-5].

There are 2 different systems for performing endotracheal suctioning, the more common open suction system (OSS) and the closed suction system (CSS). There are currently arguments for and against both systems [3].

CSS prevent the diffusion of aerosols in the outside air, thus reducing the risk of contamination for hospital staff in terms of environmental pollution. Although theoretically, it seems the best option, no national nor international studies have yet been published that evaluate the benefits and risks for patients with COVID-19 [2].

The objective of this review was to analyze the benefits and risks of closed endotracheal suctioning. The specific objectives were to (1) describe the benefits and risks of CSS with respect to OSS in patients connected to a mechanical ventilator and (2) evaluate which benefits are useful for the treatment of COVID-19 patients connected to a mechanical ventilator.

Methods

A narrative rapid review was carried out according to the Cochrane Rapid Reviews Method Group criteria, which define a rapid review as "a form of knowledge synthesis that accelerates the process of conducting a traditional systematic review through streamlining or omitting specific methods to produce evidence for stakeholders in a resource-efficient manner" [6].

Setting the Research Question—Topic Refinement

The formulation of the question followed the objectives of the review [7].

The MeSH term used for the search strategy was "endotracheal suction." To carry out a more advanced search, we added the term "closed system.“

Setting the Eligibility Criteria

We included articles following the inclusion and exclusion criteria (Textbox 1). We considered all literature containing the keywords, as long as the article contained information about CSS in adult patients.

The systematic search was carried out between November 1, 2020, and December 30, 2020.

Although the criteria indicate articles from 2010 onwards, we did include previous articles that we considered to contain essential information for our analysis.

Textbox 1. Summary of the inclusion and exclusion criteria.

Inclusion criteria

- Published between 2010 and 2020
- Published in English or Spanish
- Included a closed suction system
- Included intensive care unit (ICU) patients

Exclusion criteria

- Published prior to 2000
- Not a pediatric study

Search Procedure

Search Strategy

An electronic bibliographic search was carried out using the following databases: PubMed, MEDLINE, CINAHL, LILACS, the Cochrane Library, and IBECS. Some relevant articles were also selected from the bibliographic references of the articles found through the systematic search.

The selected keywords were "closed endotracheal suction system" and "COVID-19."

Data Collection

We first followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) process for data extraction [8]. The articles were first selected by title; we then narrowed the selection by reading abstracts and finally by reading the texts in full, dividing the work among researchers in the group. Of the 157 articles selected, after eliminating duplicates and applying the inclusion and exclusion criteria, we were left with 25 articles to read in full.

The 25 articles were read by 2 researchers, considering the objective and design of the study, and 2 other researchers selected those they considered suitable for the results. After this selection, the researchers read the articles and noted the most relevant aspects, establishing categories for the benefits and risks found. Finally, the researchers focused on the quantitative and qualitative nature of the results found.

Results

Articles

At the end of the selection process, we included 15 articles (Figure 1). It is important to highlight that they varied in terms of their methodology, from meta-analysis to randomized clinical trials and observational studies. All the articles met the criteria outlined in the methodology, using 1 reviewer to examine the final selection of articles and another reviewer to read the excluded articles in full.

The articles included were published between 2003 and 2020 (Table 1). All the included articles that were dated before 2010 were considered of special relevance and found through the search articles. Of the 15 articles, 9 were developed in collaboration with or within the European Union. The studies varied in design: 1 in vitro trial, 2 meta-analyses, 2 reviews, 2 quasi-experimental studies, 3 clinical trials with small patient samples, and 5 observational studies.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram flow. ICU: intensive care unit.

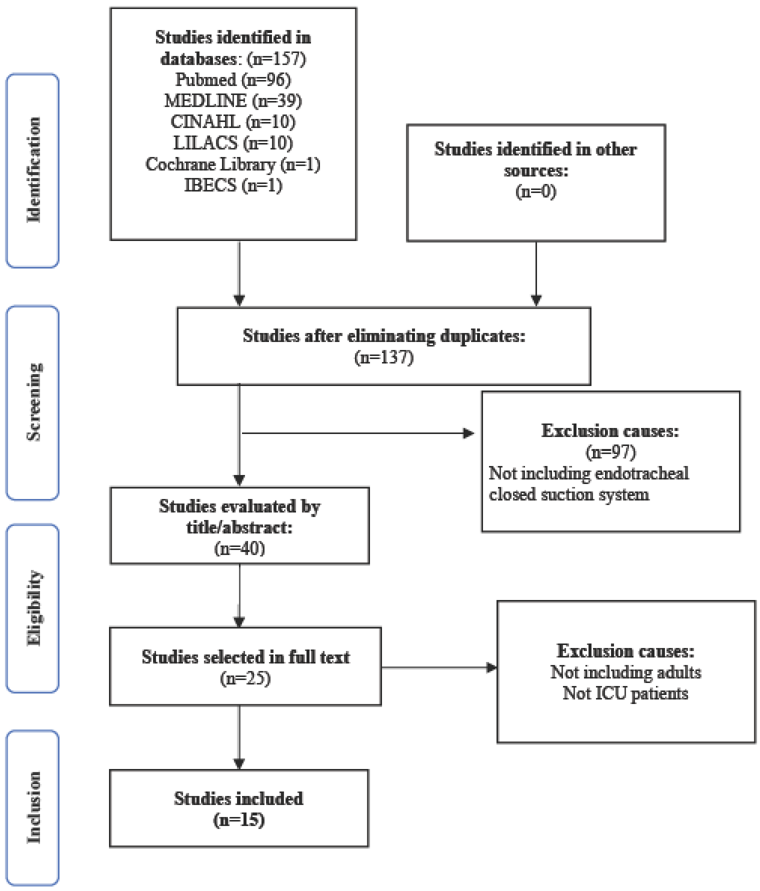


Table 1. Details of the included articles.

Article	Aim	Design	Participants
Jongerden et al (2007) [9]	To review the effectiveness of CSS[a] and OSS[b] in terms of cross contamination and economic cost	Meta-analysis	15 randomized clinical trials
Subirana et al (2007) [10]	To compare the use of CSS and OSS in patients on ventilators for more than 24 hours	Review	16 trials (1684 patients)
Elmansoury and Said (2017) [11]	To compare the use of CSS and OSS in patients on ventilators for more than 24 hours	Randomized clinical trial	141 patients
Siempos et al (2008) [12]	To evaluate if CSS prevent VAP[c]	Meta-analysis	9 randomized clinical trials
Zeitoun et al (2003) [13]	To verify the incidence of VAP with the use of CSS	Clinical trial	47 patients
Dave et al (2011) [14]	To evaluate the effectiveness of tracheal suctioning with a CSS	In vitro model	—[d]
Faradita Aryani and Tanner (2018) [15]	To compare the use of CSS and OSS in ventilated ICU[e] patients	Systematic review	5 studies
Jongerden et al (2012) [16]	To assess changes in heart rate, average arterial blood pressure, and peripheral oxygen saturation after endotracheal suctioning with a CSS	Randomized prospective observations	197 observations of endotracheal suctioning
Åkerman et al (2014) [17]	To compare the use of CSS with OSS in cases of VAP, bacterial contamination, and adverse circumstances	Observational cohorts	126 patients
Özden and Görgülü (2015) [18]	To compare the effects on the hemodynamics of patients undergoing open heart surgery	Quasi-experimental	120 patients
Adi et al (2013) [19]	To evaluate ETT[f] compared with new or unused ETTs in terms of changes in inspiratory resistance or peak inspiratory pressure	Observational	16 ETTs
de Fraga Gomes Martins et al (2019) [20]	To compare the suction volume, respiratory mechanics, and hemodynamics of patients treated with OSS/CSS and with inspiratory pause	Randomized clinical trial	31 patients
De Seta et al (2020) [21]	To establish a step-by-step protocol for patients with tracheotomy who require MV[g]	Observational	15 patients
Ricard et al (2011) [22]	To compare the contamination of gloves and the airway while using OSS and CSS	Quasi-experimental	19 cases of endotracheal suctioning
Vargas and Servillo (2020) [23]	To evaluate the use of an improvised CSS in a case of COVID-19	Observational	12 patients

[a]CSS: closed suction systems.

[b]OSS: open suction systems.

[c]VAP: ventilator-associated pneumonia.

[d]Not applicable.

[e]ICU: intensive care unit.

[f]ETT: endotracheal tube.

[g]MV: mechanical ventilation.

For all of the articles included, we identified the benefits and risks of CSS and OSS and classified them according to whether they were primary or secondary outcomes. A summary of the findings related to patient, care, and organization can be found in Table 2.

Table 2. Summary of the results of closed suction systems versus open suction systems.

Outcomes	Closed suction system	Open suction system
Patient-related outcomes		
VAP[a]	No differences [11]	Nonsignificant reduction in studies with small samples [24]; increases risk by facilitating microaspiration from the upper to the lower section [12]
Mortality	No differences [11]	—[b]
Cardiorespiratory variables	No differences in HR[c], ABP[d], and SpO_2[e] [16]	HR and ABP slightly more stable [18]; better SpO_2 recovery after pre-oxygenation [18]; better ABP and hypoxemia during heart surgery [18]
Bacterial colonization	No differences in the most common microorganisms [17]	Increased colonization (*Pseudomona aeuroginosa*, *Escherichia coli*, *Staphylococcus aereus*, and *Acinetobacter spp)* [9-11,21]
Care-related outcomes		
Changes to the ventilator	No significant differences relating to PEEP[f] [12]; less time spent connected to MV[g] [11]	Better Ri[h] and PIP[i] after first suctioning [19]; improvement in patients with elevated PEEP due to maladaptation and alveolar collapse [16]
Nursing care–related	No differences in 10-second to 15-second suctions [18]; more effective in removing secretions [16]	Variable number of suctions: every 3 hours or the minimum possible [18]; improvement in vital constants with pre-oxygenation [16]; expiratory pause, increased volume of suctioned secretions [20]; shortest suctioning time, up to 1 minute less [24]; increased number of ETT[j] occlusions requiring replacement and obstructions [17]
Organization-related outcomes		
Cost-effectiveness	Less cost- effective [9,11,16]	Less use of gloves, masks, and glasses [9]; greater cost or prolonged use (>72 hours) [11,24]
Environmental or cross contamination	No differences [9,17]	Fewer microorganisms found on staff gloves [21]
Time spent in the intensive care unit	No differences [11]	—

[a]VAP: ventilator-associated pneumonia.
[b]Not applicable.
[c]HR: heart rate.
[d]ABP: arterial blood pressure.
[e]SpO_2: peripheral oxygen saturation.
[f]PEEP: positive and end-expiratory pressure.
[g]MV: mechanical ventilation.
[h]Ri: inspiratory resistance.
[i]PIP: peak inspiratory pressure.
[j]ETT: endotracheal tube.

Primary Outcomes

The baseline condition of the patients or their pathologies was the variable that most influenced the analysis of the benefits and risks of endotracheal suctioning [10].

Ventilator-Associated Pneumonia

Elmansoury and Said [11] carried out an analysis in 2 groups: 1 intervention group (n=66) with a CSS and 1 control group (n=75) with an OSS for 6 months with the possible incidence of ventilator-associated pneumonia (VAP). They found no statistically significant differences: 30.13 VAP per 1000 ventilator days in the control group and 17.48 VAP per 1000 ventilator days in the intervention group. Jongerden et al [9] and Subirana et al [10] found no statistically significant differences between VAP with CSS and VAP with OSS (odds ratio [OR]=0.96, 95% CI 0.76-1.21; n=1377, risk ratio [RR]=0.88, 95% CI 0.70-1.12). A slight reduction was found when using a CSS in studies with a small sample size (n=9); for example, Zeitoun et al [13] found no significant differences, but the frequency of VAP in cases treated with OSS was 11 of 24 cases, while the frequency of VAP in cases treated with CSS was 10-14 of 23 cases. However, in the study by Dave et al

[14], the researchers concluded that CSS facilitate the microaspiration of fluid from the upper zone to the lower zone, increasing the risk of VAP. In addition, the variation in VAP definition criteria and the absence of a clear description make adequate comparison impossible. Siempos et al [12] and Jongerden et al [9] included articles that defined VAP using more quantitative or qualitative results, for example different temperatures, time with MV, or colony-forming unit [9,12,13].

Mortality

No significant differences in mortality were found in any of the included studies. For example, Jongerden et al [9] and Subirana et al [10] found no statistically significant differences between CSS and OSS (OR=1.02, 95% CI 0.84-1.25 and RR=1.02, 95% CI 0.84-1.23, respectively) [10,11,24]. The systematic review by Faradita Aryani and Tanner [15] included 435 prospective studies, concluding that none of the studies found differences regarding increased VAP or resultant mortality.

Secondary Outcomes

Cardiorespiratory Variables

In only 1 of the studies, the variables were more stable with the use of closed systems. Even so, heart rate (HR) was almost imperceptible, and there were no significant differences in arterial blood pressure (ABP) [10,13]. Jongerden et al [16], with a total of 165 patients using CSS and OSS, measured physiological parameters including HR, ABP, and peripheral oxygen saturation (SpO_2), without noting any differences. In this study, they found notable—although not significant—differences in SpO_2 recovery after pre-oxygenation in patients using CSS (96%-99%) and OSS (95%-98%) [10]. Likewise, in terms of oxygen saturation, no differences were found in the study carried out by Åkerman et al [17]. However, Özden and Görgülü [18] concluded that HR with OSS increased at 5 minutes and 15 minutes after the procedure and hypoxemia can in fact be avoided using a CSS while also improving ABP in postoperative patients, particularly after heart surgery.

Relating to Changes to the Ventilator

Adi et al [19] examined aspects relating to the ventilator, such as inspiratory resistance (Ri) and peak inspiratory pressure (PIP); they estimated endotracheal tube (ETT) obstruction at extubation, taking into account patients with more than 12 hours of MV and obtaining an improvement in both Ri and PIP after the first suction with a CSS. Dave et al [14], using a simulation model without patients, concluded that using a CSS does not achieve positive results in terms of maintaining positive and end-expiratory pressure (PEEP). de Fraga Gomes Martins et al [20] and Jongerden et al [16] did not achieve the same results, but both did recommend the use of CSS in patients who require elevated PEEP to prevent alveolar collapse in order to avoid asynchrony to MV. There were no significant differences in the time ICU patients were connected to MV (weighted mean difference [WMD]=0.44, 95% CI 0.92-1.80) [10]. Subirana et al [10] mentioned the time patients were connected to MV only, without conducting an analysis. However, Siempos et al [12] mentioned CSS was associated with longer MV duration (WMD=0.65 days, 95% CI 0.28-1.03) [9].

Bacterial Colonization

Some studies demonstrated an increase in colonization while using a CSS, with a 49% increased risk in comparison with OSS (OR=2.88, 95% CI 1.52-5.52) [9,10,13]. Åkerman et al [17], in their cohort study, included 126 patients: 61 using an OSS and 65 using a CSS. Both groups showed colonization with similar gram-negative bacteria, the most common being *Pseudomona aeuroginosa*, *Escherichia coli*, and *Staphylococcus aereus*; other studies connected colonization to *Acinetobacter spp.* without general VAP-related differences. However, Elmansoury and Said [11] found greater incidence of *Acinetobacter spp.* and *Pseudomonas aeuroginosa* (causative of VAP) with CSS as well as no incidence of methicillin-resistant *Staphylococcus aureus* and *Staphylococcus aereus.*

Nursing Care-Related

Suctioning Techniques

The lack of description of suctioning techniques and their characteristics in the articles makes comparison impossible [10,13].

The application of oxygenation prior to endotracheal suctioning is an important variable. For example, Jongerden et al [16] found differences in SpO_2 recovery when pre-oxygenation was used with a CSS. Özden and Görgülü [18] described the suctioning protocol: 1 minute of pre-oxygenation at 100%, universal precautions, for 10 seconds to 15 seconds, using the smallest possible level of suction (<120 mm Hg), and oxygenation at 100% for another minute. In the randomized clinical trial by de Fraga Gomes Martins et al [20], they provided a detailed description of the process: pre-oxygen at 100% for 1 minute before suctioning 3 times for 10 seconds. This procedure was followed for the control group, while an expiratory pause was included for the intervention group. The authors concluded that the expiratory pause resulted in an increase in the volume of secretions suctioned [20].

Suctioning System

The type of device used in CSS was only described in 1 of the articles, and we decided to discard studies on single-use devices because this means disconnecting the patient from the ventilator [11,19].

Suctioning Time and Frequency

Suctioning time is a necessary variable because it has an impact on workload and it can alter hemodynamics for more or less time [10]. The 2012 observational study by Jongerden et al [16] found no significant differences between one system and another; the suctioning processes lasted 10 seconds to 15 seconds and had the same effect on the vital constants. Work overload is another aspect that must be taken into account. It has been observed that using a CSS can be up to 1 minute faster: 2.5 minutes for OSS in comparison with 1.5 minutes for CSS [10].

In terms of the number of suctions, differences in recommendations range from performing endotracheal suctioning every 3 hours, to a minimal number and only when strictly necessary [22]. Frequent suctions can provoke hemodynamic instability, damage to the tracheal mucosa, hemorrhage, and infection [19].

Suctioning Results

The quantity of suctioned secretions has not been studied in depth despite the importance of these results [10,20]. OSS are considered more effective in removing secretions, but the articles only discussed experiments in vitro or in animal models [16]. In the study by Åkerman et al [17], 3 ETT occlusions and 3 severe obstructions were reported in patients using an OSS, while in the CSS group, only 1 occlusion was reported.

Organization-Related

Economic Benefit

Jongerden et al [9] found that the cost of CSS is between 14 and 100 times the cost of OSS, but there is less need for personal protective equipment (gloves, masks, and glasses) when using CSS.

There is some debate on this issue, although most studies associate higher costs with the use of closed systems, with the exception of systems connected for at least 72 hours (24 hours is recommended) [9,10]. However, prolonged use has the disadvantage of increasing bacterial colonization [9,10].

Environmental or Cross Contamination

The fewer the disconnections, the less likely that pathogens will be spread into the environment. No differences in transmissions were found between patients in the same place and receiving care from the same staff [9,17]. However, Ricard et al [22] highlighted that there was in fact less risk of glove contamination with microorganisms using a CSS.

Time Spent in the ICU

No differences were observed in relation to time spent in the ICU [10]. The 2007 meta-analysis by Jongerden et al [9] included 15 clinical trials using CSS and OSS and concluded that none of the benefits associated with CSSs are scientifically proven.

COVID-19

It is important to set out the benefits of a CSS in order to manage the benefits and risks faced by COVID-19 patients.

The protocol followed by De Seta et al [21] in COVID-19 patients with tracheotomy recommended mucus management using 3 elements: humidifiers, bacterial and viral filters, and CSS. These 3 elements are intended to prevent the spread of SARS-CoV-2 viral aerosols [21].

In the article by Vargas and Servillo [23], we found an alternative to CSS that arose from the shortage of CSS during the global pandemic. This system is composed of an OSS with the addition of the sterile sheath that is commonly used when performing ultrasound scans. However, there are no results on the effectiveness of CSS in COVID-19 patients [23].

Discussion

Principal Findings

This rapid review established 3 different categories for comparing the benefits and risks of the use of CSS versus OSS. The review compared 11 different outcomes. Using this classification, we can establish comparability indicators for future studies despite the fact that the indicators described are not significantly conclusive. Consideration should be given to the proposed benefits, which could make a significant difference to the procedures for treating COVID-19.

As seen in the Results section, there is ample variability in the conclusions of the studies. After analysis, we concluded that there is a need for the development and implementation of clinical practice guidelines on suctioning. This is because, in all cases, the implementation of suctioning guidelines that include a protocol for the technique improves outcomes for patients [25].

Among the benefits of CSS, the most discussed is the benefit and risk with respect to VAP, referred to in al the included articles. Although there was no significant evidence in most of the studies that CSS protect against VAP, all possible measures should be taken to avoid co-infection, as this leads to increased morbidity and mortality [26]. The fact that CSS can increase the risk of VAP should provide motivation for further studies that take into account actions to prevent this increase in colonizations, such as aspiration of subglottic secretions or oral hygiene.

One of the most positive aspects of CSSs is the avoidance of asynchrony and discomfort, maintaining PEEP and avoiding hypoxemia during suctioning [27,28].

This is because patients with COVID-19 require close ventilatory support [29].

However, it is important to highlight the key risks of CSS in order to apply possible measures of prevention. ETT obstruction is a proven complication as CSS fail to suction the same amount of sputum as OSS [9]. This is especially important with COVID-19 since obstruction of an ETT requires new intubation or the application of the Ambu ventilation procedure, which increases the risk of generating aerosols [2,30,31].

The economic cost of CSS is higher than that of OSS, although studies have shown that extended use of CSS can improve the cost-effectiveness. However, CSS are increasingly popular, and the benefits include reduced time spent performing tracheal suction, which frees nurses up for other important activities. We must bear in mind that there are different types of CSS, health care staff require training to use them, and protocols for correct use must be established by the unit or the manufacturer [10,32].

Another important benefit is the reduction of cross contamination or infection of the staff themselves. According to the World Health Organization [33], around 14% of infections worldwide occur among health care staff, which, in consideration of possible future virus outbreaks, is further incentive to improve their protection through the use of resources such as CSS [34].

There are differences in the samples obtained from contaminated gloves worn by health care staff using suction systems: 9 of 9 gloves were contaminated during tracheal suction using OSS, and 3 of 10 gloves when using CSS [22]. For COVID-19, contamination depends on the stage of infection as well as the interventions performed, with suctioning standing out as one of the most infectious interventions. Reducing the exposure time of staff each time a suctioning procedure is performed can be another major benefit [34].

Limitations

The major limitation of this review is the methodological quality of the studies included and the inability to carry out reliable comparisons. Most of the studies applied their selection criteria on the basis of convenience and inaccurately described the intervention carried out.

In addition, there are limitations relating to language and document access, as only fully accessible documents were included. The majority of study samples are small. It should be noted that fighting the COVID-19 global pandemic has required huge economic and human resources, which has reduced the resources applied to research and the production of specific literature on the subject.

Comparison With Prior Work

Despite these limitations, the strengths of the review include it serving as a starting point for future research and the fact that it was carried out at the time of the pandemic and includes quality and meaningful results, such as clinical trials. Moreover, the results found are similar to other reviews conducted for the same purpose.

Conclusions

This review suggests that CSS have some benefits for patients with COVID-19. However, the variation in design of the reviewed studies means that there are no comparative results. Further experimental research on CSS and OSS used on patients is required. This review is the first summary of the indicators of relevance to this area of practice and offers future researchers the outcome measures for comparing CSS and OSS. The first step is to establish a protocol study and evidence-based practical guidelines for COVID-19 patients. In order to carry out future experimental studies, we need to unify or specify the criteria in order to understand their possible influences. Key factors include the baseline situation of the patient, standardized criteria for VAP, or diagnosis at admission to the ICU. Furthermore, with endotracheal aspiration, this study shows that we must take into account the specific nursing care procedure adopted because it can be a very influential factor in the outcomes for CSS versus OSS.

Acknowledgments

This study received support for publication from the Research group in Healthcare (GRUPAC)-University of La Rioja and the Health System Sustainability Research Unit (GISOSS)-Biomedical Research Centre of La Rioja (CIBIR)

Conflicts of Interest

None declared.

References

1. World Health Organization. Management of severe / critical cases of COVID-19 with non-invasive or mechanical ventilation. World Health Organization. 2020 Jun 1. URL: https://apps.who.int/iris/handle/10665/332340 [accessed 2022-12-27]
2. Jackson T, Deibert D, Wyatt G, Durand-Moreau Q, Adisesh A, Khunti K, et al. Classification of aerosol-generating procedures: a rapid systematic review. BMJ Open Respir Res 2020 Oct;7(1):1 [FREE Full text] [doi: 10.1136/bmjresp-2020-000730] [Medline: 33040021]
3. AARC. AARC clinical practice guideline. Endotracheal suctioning of mechanically ventilated adults and children with artificial airways. American Association for Respiratory Care. Respir Care 1993 May;38(5):500-504. [Medline: 10145834]
4. Favretto D, Silveira RDC, Canini SDS, Garbin L, Martins F, Dalri M. Endotracheal suction in intubated critically ill adult patients undergoing mechanical ventilation: a systematic review. Rev Lat Am Enfermagem 2012;20(5):997-1007 [FREE Full text] [doi: 10.1590/s0104-11692012000500023] [Medline: 23174846]
5. Pedersen CM, Rosendahl-Nielsen M, Hjermind J, Egerod I. Endotracheal suctioning of the adult intubated patient--what is the evidence? Intensive Crit Care Nurs 2009 Feb;25(1):21-30. [doi: 10.1016/j.iccn.2008.05.004] [Medline: 18632271]
6. Garritty C, Gartlehner G, Nussbaumer-Streit B, King VJ, Hamel C, Kamel C, et al. Cochrane Rapid Reviews Methods Group offers evidence-informed guidance to conduct rapid reviews. J Clin Epidemiol 2021 Feb;130:13-22 [FREE Full text] [doi: 10.1016/j.jclinepi.2020.10.007] [Medline: 33068715]
7. da Costa Santos CM, de Mattos Pimenta CA, Nobre MRC. The PICO strategy for the research question construction and evidence search. Rev Lat Am Enfermagem 2007;15(3):508-511 [FREE Full text] [doi: 10.1590/s0104-11692007000300023] [Medline: 17653438]
8. Tricco A, Lillie E, Zarin W, O'Brien K, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]

9. Jongerden IP, Rovers MM, Grypdonck MH, Bonten MJ. Open and closed endotracheal suction systems in mechanically ventilated intensive care patients: a meta-analysis. Crit Care Med 2007 Jan;35(1):260-270. [doi: 10.1097/01.CCM.0000251126.45980.E8] [Medline: 17133187]
10. Subirana M, Solà I, Benito S. Closed tracheal suction systems versus open tracheal suction systems for mechanically ventilated adult patients. Cochrane Database Syst Rev 2007 Oct 17;2007(4):CD004581 [FREE Full text] [doi: 10.1002/14651858.CD004581.pub2] [Medline: 17943823]
11. Elmansoury A, Said H. Closed suction system versus open suction. Egyptian Journal of Chest Diseases and Tuberculosis 2017 Jul;66(3):509-515. [doi: 10.1016/j.ejcdt.2016.08.001]
12. Siempos II, Vardakas KZ, Falagas ME. Closed tracheal suction systems for prevention of ventilator-associated pneumonia. Br J Anaesth 2008 Mar;100(3):299-306 [FREE Full text] [doi: 10.1093/bja/aem403] [Medline: 18250225]
13. Zeitoun SS, de Barros ALBL, Diccini S. A prospective, randomized study of ventilator-associated pneumonia in patients using a closed vs. open suction system. J Clin Nurs 2003 Jul;12(4):484-489. [doi: 10.1046/j.1365-2702.2003.00749.x] [Medline: 12790861]
14. Dave MH, Frotzler A, Madjdpour C, Koepfer N, Weiss M. Massive aspiration past the tracheal tube cuff caused by closed tracheal suction system. J Intensive Care Med 2011;26(5):326-329. [doi: 10.1177/0885066610392516] [Medline: 21300670]
15. Faradita Aryani D, Tanner J. Does open or closed endotracheal suction affect the incidence of ventilator associated pneumonia in the intensive care unit? A systematic review. Enfermería Clínica 2018 Feb;28:325-331. [doi: 10.1016/S1130-8621(18)30179-7]
16. Jongerden I, Kesecioglu J, Speelberg B, Buiting A, Leverstein-van Hall MA, Bonten M. Changes in heart rate, mean arterial pressure, and oxygen saturation after open and closed endotracheal suctioning: a prospective observational study. J Crit Care 2012 Dec;27(6):647-654. [doi: 10.1016/j.jcrc.2012.02.016] [Medline: 22520496]
17. Åkerman E, Larsson C, Ersson A. Clinical experience and incidence of ventilator-associated pneumonia using closed versus open suction-system. Nurs Crit Care 2014 Jan;19(1):34-41. [doi: 10.1111/nicc.12010] [Medline: 24400607]
18. Özden D, Görgülü RS. Effects of open and closed suction systems on the haemodynamic parameters in cardiac surgery patients. Nurs Crit Care 2015 May;20(3):118-125. [doi: 10.1111/nicc.12094] [Medline: 24991700]
19. Adi NA, Tomer NT, Bergman GB, Kishinevsky EK, Wyncoll DW. Effects of prolonged mechanical ventilation with a closed suction system on endotracheal tube resistance and its reversibility by a closed suction cleaning system. Anaesth Intensive Care 2013 Nov;41(6):728-735 [FREE Full text] [doi: 10.1177/0310057X1304100607] [Medline: 24180713]
20. de Fraga Gomes Martins L, da Silva Naue W, Skueresky AS, Bianchi T, Dias AS, Forgiarini LA. Effects of combined tracheal suctioning and expiratory pause: a crossover randomized clinical trial. Indian J Crit Care Med 2019 Oct;23(10):454-457 [FREE Full text] [doi: 10.5005/jp-journals-10071-23263] [Medline: 31749553]
21. De Seta D, Carta F, Puxeddu R. Management of tracheostomy during COVID-19 outbreak: Heat and moisture exchanger filter and closed suctioning system. Oral Oncol 2020 Jul;106:104777 [FREE Full text] [doi: 10.1016/j.oraloncology.2020.104777] [Medline: 32402655]
22. Ricard J, Eveillard M, Martin Y, Barnaud G, Branger C, Dreyfuss D. Influence of tracheal suctioning systems on health care workers' gloves and equipment contamination: a comparison of closed and open systems. Am J Infect Control 2011 Sep;39(7):605-607. [doi: 10.1016/j.ajic.2010.10.031] [Medline: 21514008]
23. Vargas M, Servillo G. Closed-suction system for intubated COVID-19 patients with the use of an ultrasound probe cover. Anesthesiology 2020 Sep;133(3):687-689 [FREE Full text] [doi: 10.1097/ALN.0000000000003431] [Medline: 32501955]
24. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
25. Maggiore SM, Lellouche F, Pignataro C, Girou E, Maitre B, Richard JM, et al. Decreasing the adverse effects of endotracheal suctioning during mechanical ventilation by changing practice. Respir Care 2013 Oct;58(10):1588-1597 [FREE Full text] [doi: 10.4187/respcare.02265] [Medline: 23466423]
26. Sharov KS. SARS-CoV-2-related pneumonia cases in pneumonia picture in Russia in March-May 2020: Secondary bacterial pneumonia and viral co-infections. J Glob Health 2020 Dec;10(2):020504 [FREE Full text] [doi: 10.7189/jogh.10.020504] [Medline: 33110587]
27. Cereda M, Villa F, Colombo E, Greco G, Nacoti M, Pesenti A. Closed system endotracheal suctioning maintains lung volume during volume-controlled mechanical ventilation. Intensive Care Med 2001 Apr;27(4):648-654. [doi: 10.1007/s001340100897] [Medline: 11398690]
28. Morrow BM. Closed-system suctioning: why is the debate still open? Indian J Med Sci 2007 Apr;61(4):177-178. [Medline: 17401253]
29. Roesthuis L, van den Berg M, van der Hoeven H. Advanced respiratory monitoring in COVID-19 patients: use less PEEP!. Crit Care 2020 May 15;24(1):230 [FREE Full text] [doi: 10.1186/s13054-020-02953-z] [Medline: 32414399]
30. Sugimoto R, Kenzaka T, Fujikawa M, Kawasaki S, Nishisaki H. Humidifier use and prone positioning in a patient with severe COVID-19 pneumonia and endotracheal tube impaction due to highly viscous sputum. Cureus 2020 Jun 15;12(6):e8626 [FREE Full text] [doi: 10.7759/cureus.8626] [Medline: 32550095]

31. Bryant J, Tobias J. Enclosure with augmented airflow to decrease risk of exposure to aerosolized pathogens including coronavirus during endotracheal intubation. Can the reduction in aerosolized particles be quantified? Paediatr Anaesth 2020 Aug;30(8):900-904 [FREE Full text] [doi: 10.1111/pan.13934] [Medline: 32464695]
32. Coppadoro A, Bellani G, Foti G. Non-pharmacological interventions to prevent ventilator-associated pneumonia: a literature review. Respir Care 2019 Dec;64(12):1586-1595 [FREE Full text] [doi: 10.4187/respcare.07127] [Medline: 31551284]
33. WHO Director-General's opening remarks at the media briefing on COVID-19 - 14 December 2020. World Health Organization. 2020 Dec 14. URL: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---14-december-2020 [accessed 2022-12-27]
34. Ahn J, An S, Sohn Y, Cho Y, Hyun J, Baek Y, et al. Environmental contamination in the isolation rooms of COVID-19 patients with severe pneumonia requiring mechanical ventilation or high-flow oxygen therapy. J Hosp Infect 2020 Nov;106(3):570-576 [FREE Full text] [doi: 10.1016/j.jhin.2020.08.014] [Medline: 32828864]

Abbreviations

ABP: arterial blood pressure
CIBIR: Biomedical Research Centre of La Rioja
CSS: closed suction system
ETT: endotracheal tube
GISOSS: Health System Sustainability Research Unit
GRUPAC: Research group in Healthcare
HR: heart rate
ICU: intensive care unit
MV: mechanical ventilation
OR: odds ratio
OSS: open suction system
PEEP: positive and end-expiratory pressure
PIP: peak inspiratory pressure
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
Ri: inspiratory resistance
RR: risk ratio
SpO2: peripheral oxygen saturation
VAP: ventilator-associated pneumonia
WMD: weighted mean difference

Edited by A Mavragani; submitted 08.09.22; peer-reviewed by B Angulo Nalda, M Kapsetaki; comments to author 25.11.22; revised version received 12.12.22; accepted 13.12.22; published 10.01.23.

Please cite as:
Ramírez-Torres CA, Rivera-Sanz F, Sufrate-Sorzano T, Pedraz-Marcos A, Santolalla-Arnedo I
Closed Endotracheal Suction Systems for COVID-19: Rapid Review
Interact J Med Res 2023;12:e42549
URL: https://www.i-jmr.org/2023/1/e42549
doi:10.2196/42549
PMID:36548950

Original Paper

Cardiovascular Risk Assessment Among Adolescents and Youths Living With HIV: Evaluation of Electronic Health Record Findings and Implications

Sitaji Gurung[1], MPH, MD, PhD; Kit N Simpson[2], DrPH; Christian Grov[3], MPH, PhD; H Jonathon Rendina[4], MPH, PhD; Terry T K Huang[5], MPH, PhD; Henna Budhwani[6], MPH, PhD; Stephen Scott Jones[7], MS; Tyra Dark[8], MA, PhD; Sylvie Naar[8], PhD

[1]Department of Health Sciences, New York City College of Technology (City Tech), The City University of New York, Brooklyn, NY, United States

[2]Department of Healthcare Leadership and Management, Medical University of South Carolina, Charleston, SC, United States

[3]Department of Community Health and Health Policy, The City University of New York Graduate School of Public Health and Health Policy, New York, NY, United States

[4]Milken Institute School of Public Health, The George Washington University, Washington, DC, United States

[5]Center for Systems and Community Design, The City University of New York Graduate School of Public Health and Health Policy, New York, NY, United States

[6]College of Nursing, Florida State University, Tallahassee, FL, United States

[7]Whitman-Walker Institute, The George Washington University, Washington, DC, United States

[8]College of Medicine, Florida State University, Tallahassee, FL, United States

Corresponding Author:
Sitaji Gurung, MPH, MD, PhD
Department of Health Sciences
New York City College of Technology (City Tech)
The City University of New York
285 Jay Street, A811D
Brooklyn, NY, 11201
United States
Phone: 1 718 260 5673
Email: Sitaji.Gurung65@citytech.cuny.edu

Abstract

Background: The HIV epidemic remains a major public health concern, particularly among youths living with HIV. While the availability of antiretroviral therapy has significantly improved the health outcomes of people living with HIV, there is growing evidence that youths living with HIV may be at increased risk of cardiovascular disease. However, the underlying mechanisms linking HIV and cardiovascular disease among youths living with HIV remain poorly understood. One potential explanation is that HIV-related biomarkers, including detectable viral load (VL) and low cluster of differentiation 4 (CD4) lymphocyte counts, may contribute to increased cardiovascular risk. Despite the potential importance of these biomarkers, the relationship between HIV-related biomarkers and cardiovascular risk among youths living with HIV has been understudied.

Objective: To address this gap, we examined whether detectable VL and low CD4 lymphocyte counts, both of which are indications of unsuppressed HIV, were associated with cardiovascular risk among youths living with HIV.

Methods: We analyzed electronic health record data from 7 adolescent HIV clinics in the United States (813 youths living with HIV). We used multivariable linear regression to examine the relationship between detectable VL and CD4 lymphocyte counts of ≤200 and cardiovascular risk scores, which were adapted from the gender-specific Framingham algorithm.

Results: In our study, nearly half of the participants (366/766, 47.8%) had detectable VL, indicating unsuppressed HIV, while 8.6% (51/593) of them had CD4 lymphocyte counts of ≤200, suggesting weakened immune function. We found that those with CD4 lymphocyte counts of ≤200 had significantly higher cardiovascular risk, as assessed by Cardiac Risk Score2, than those with CD4 lymphocyte counts of >200 (P=.002). After adjusting for demographic and clinical factors, we found that for every 1000-point increase in VL copies/mL, the probability of having cardiovascular risk (Cardiac Risk Score2) increased by 38%. When measuring the strength of this connection, we observed a minor effect of VL on increased cardiovascular risk (β=.134, SE

0.014; *P*=.006). We obtained similar results with Cardiac Risk Score1, but the effect of CD4 lymphocyte counts of ≤200 was no longer significant. Overall, our findings suggest that detectable VL is associated with increased cardiovascular risk among youths living with HIV, and that CD4 lymphocyte counts may play a role in this relationship as well.

Conclusions: Our study highlights a significant association between unsuppressed HIV, indicated by detectable VL, and increased cardiovascular risk in youths living with HIV. These findings emphasize the importance of implementing interventions that address both VL suppression and cardiovascular risk reduction in this population. By tailoring interventions to meet the unique needs of youths, we can promote overall well-being throughout the HIV care continuum and across the life span. Ultimately, these efforts have the potential to improve the health outcomes and quality of life of youths living with HIV.

International Registered Report Identifier (IRRID): RR2-10.2196/11185

(Interact J Med Res 2023;12:e41574) doi:10.2196/41574

KEYWORDS

cardiovascular risk; cluster of differentiation 4 lymphocyte; electronic health record; viral load; youths living with HIV

Introduction

Youths living with HIV is a high-priority population within those living with HIV [1]. Youths living with HIV will need to manage their diagnosis for years longer than their adult peers, and negative health consequences associated with living with HIV may appear earlier in the life course [2-6]. Research shows significant gaps in the HIV treatment cascade in younger populations compared to older groups [7-9]. Youths living with HIV aged 18-24 years have lower uptake of HIV testing and lower rates of treatment initiation compared to older peers, which continues to present a significant challenge to epidemic control [8-10]. Notably, racial disparities are more pronounced in youths [11,12]. Youths who identify as Black and Latinx represent 54% and 25% of new HIV diagnoses (79% in aggregate), compared to those who identify as White and other races, accounting for 16% and 5% of new HIV diagnoses [13]. There is growing evidence to suggest that Black individuals are at a higher risk for cardiovascular disease (CVD) compared to individuals in other racial and ethnic groups [14]. Several factors contribute to this increased risk. One of the most significant is systemic racism and discrimination, which can result in chronic stress and inflammation that can damage the heart and blood vessels over time [15]. According to the American Heart Association, Black individuals are more likely to have hypertension and less likely to have it under control compared to other racial and ethnic groups [16]. Black individuals are also more likely to develop diabetes and are at a higher risk of dying from diabetes-related complications. These conditions are major risk factors for CVD and contribute to the increased risk seen in Black individuals.

Cardiovascular health disparities are well documented among older adults with HIV [17-22], but there is a dearth of research on cardiovascular health metrics and prevention strategies for youths living with HIV in the United States. Studies have shown that older adults with HIV have a higher risk of developing CVD compared to the general population, which is likely due to a combination of factors, including HIV-related inflammation, the use of antiretroviral therapy (ART), and lifestyle factors, such as smoking and poor diet [21,23,24]. However, there is a lack of research on the cardiovascular health of youths living with HIV in the United States, despite the fact that this population is growing and faces unique challenges to their health. Research on cardiovascular health metrics is therefore critical to the overall understanding of the impact of HIV infection on behavioral factors and the relationship between HIV infection and cardiovascular risk. Research has suggested that indicators of CVD may appear earlier in individuals living with HIV, with some studies suggesting that these indicators can emerge as early as adolescence [4,6,25], potentially due to a combination of factors. One factor is that HIV infection can cause chronic inflammation, which has been linked to the development of atherosclerosis, a key contributor to CVD [26,27]. Another factor is that some ARTs used to manage HIV infection may have adverse effects on lipid metabolism, leading to the accumulation of fatty deposits in the blood vessels and an increased risk of CVD [28].

In addition to these factors, individuals living with HIV may also have a higher prevalence of traditional CVD risk factors, such as hypertension, diabetes, and smoking, which can further increase their risk of developing CVD [29]. Stigma and discrimination related to HIV infection can also contribute to poor mental health outcomes, which have been linked to an increased risk of CVD [30]. Given the potential for CVD indicators to emerge earlier in individuals living with HIV, it is important to develop effective prevention and treatment strategies targeted toward this population. This may include identifying and managing traditional CVD risk factors, optimizing ART to minimize potential negative effects on lipid metabolism, and addressing mental health concerns through counseling and other interventions. The American Heart Association's Life's Simple 7 (LS7) concept of ideal cardiovascular health is not well understood in youths living with HIV when compared to the Healthy People 2020-2030 goals [31-34]. A higher LS7 score indicates better cardiovascular health and is associated with a lower incidence of CVD [35-38]. Traditional assessments of individual cardiovascular risk factors are inadequate in capturing the overall risk posed by multiple factors at the population level [39]. It is important to consider risk factors that occur together [40,41] as cardiovascular risk factor profiles [42-44] and assess them using composite measures such as the Framingham Risk Score (FRS) [45]. The modified FRS is the most commonly used algorithm for cardiovascular risk assessment in the United States [45,46].

Reducing cardiovascular risk in youths living with HIV is a critical issue that requires attention. It is essential to conduct cardiovascular risk assessment in this population to optimize the early diagnosis and treatment of CVD during adolescence [47]. However, research focused on comorbidities in HIV has rarely explored the connection between detectable viral load (VL) and low cluster of differentiation 4 (CD4) lymphocyte counts with increased cardiovascular risk in a US-based population of youths living with HIV. Additionally, there are currently no published data on the prevalence of traditional cardiovascular risk factors, such as dyslipidemia, hypertension, diabetes, and smoking, among youths living with HIV aged 14-26 years, even though metabolic changes leading to atherosclerosis can begin early in life, and go undiagnosed for an extended period. Despite this, cardiac risk estimating algorithms like the FRS have not been applied to younger populations [48,49]. Therefore, the aim of this study was to develop cardiovascular risk profiles for a cohort of US-based youths living with HIV and compare their profiles based on VL and CD4 lymphocyte status.

Methods

Study Design and Population

The study analyzed deidentified electronic health record (EHR) data of adolescents for the Adolescent Medicine Trials Network protocols. The Adolescent Medicine Trials Network protocols were previously described in a publication [50], and the participating sites included in the EHR extraction protocol [51] provide HIV care to youths living with HIV. This care begins with a diagnosis, followed by linking them to an HIV care provider who can help manage their HIV on a regular basis.

Ethics Approval

The institutional review board (IRB) of Florida State University granted approval (STUDY00000549) for the analysis of deidentified data, which was carried out in compliance with the US Department of Health and Human Services 45 Code of Federal Regulations Part 46. The parent protocol did not involve an informed consent process, and a waiver of consent and HIPAA (Health Insurance Portability and Accountability Act) waiver were granted. To protect confidentiality, the data were deidentified and participants were identified only by a unique study ID. The data were analyzed in aggregate to compare the clinic site with individuals, and a centralized data extraction process was used at all sites. Each participating site received a signed statement confirming the IRB's approval before the study began, and a reliance agreement was obtained from each site indicating their commitment to providing deidentified EHR data in accordance with the IRB's approval.

The extracted EHR data included those who had standard of care and treatment visits at one of the participating sites between January 1, 2016, and December 31, 2016. The baseline data extraction cycle was launched in December 2017 and completed in April 2018. Once the data files were verified for each site, they were merged to produce a single analytic data set of 1093 youths living with HIV. To be included in these analyses, data on demographics, HIV biomarkers, vital statistics, cholesterol panel, and the International Statistical Classification of Disease and Related Health Problems, Tenth Revision (ICD-10), diagnosis codes were required. Data downloads from Baltimore, Maryland; Birmingham, Alabama; Los Angeles, California; Memphis, Tennessee; San Diego, California; Tampa, Florida; and Washington, District of Columbia; were included in the analyses. Because Brooklyn, New York; Miami, Florida; and Philadelphia, Pennsylvania; did not provide gender and systolic blood pressure data, these 3 sites were excluded from further analyses.

Measures

Demographics

We defined age as a continuous variable in years as reported by the site. Gender, race, and ethnicity definitions varied within a site's data download. For gender, we constructed a binary variable of man and woman. Race and ethnicity were categorized into Black, Latinx, White, and other.

HIV and Immunologic Biomarkers

Earliest VL values in "copies/mL" by date were used from each site. VL was log-transformed for analysis as a continuous outcome. The detectable VL threshold as reported by the site was coded as a dichotomous variable. Based on the guidelines for the use of antiretroviral agents in HIV-1 infected adults and adolescents, we defined immunocompromised as a CD4 T lymphocyte count of less than or equal to 200 [52,53]. CD4 lymphocyte count was divided by 100 to enhance the interpretability of the coefficients to correspond to a 100-cell increase rather than a 1-cell increase.

ART Medication

All prescribed HIV medications were coded into a dichotomous variable, where 0=no prescribed medication and 1=at least one prescribed medication.

Substance Abuse Diagnosis

We used ICD-10 codes F1010, F1210, F1220, F1290, F1510, F1511, F1520, F1590, F17200, F17210, F1910, and F1920 to identify alcohol and drug dependence. Substance abuse was treated as a dichotomous variable.

BMI

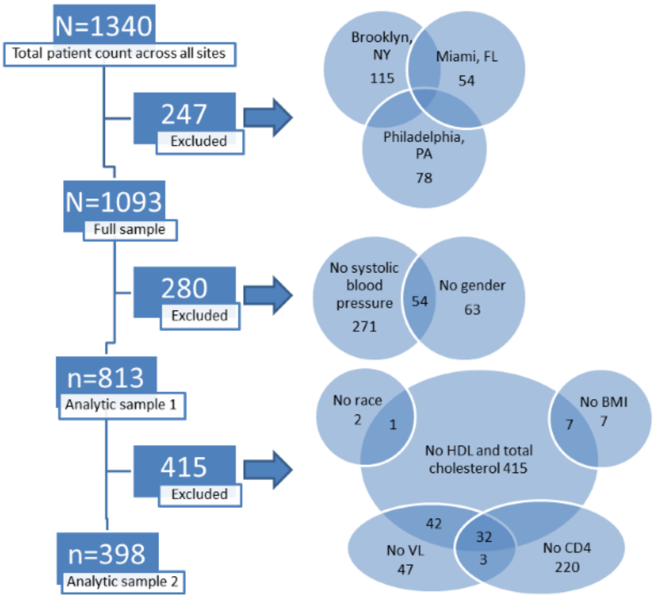
BMI was calculated based on weight and height (kg/m^2). If multiple values for a given patient were present in EHR, values from the earliest date were used.

Cardiac Risk Scores

We adopted the gender-specific cardiovascular risk algorithm developed by the Framingham Heart Study [43] to construct a cardiovascular risk variable referred to as Cardiac Risk Score1 using clinic-based predictors that are routinely obtained in primary care and do not require laboratory testing. The variables required for constructing the Cardiac Risk Score1 include age, gender, systolic blood pressure, use of antihypertensive medication, smoking status, and diabetes status. As gender and systolic blood pressure data were required for the development of the Cardiac Risk Score1, those without these data were excluded from the first analytic sample.

We also constructed a second cardiovascular risk variable referred to as Cardiac Risk Score2 that used routinely obtained clinic-based predictors including laboratory testing. The variables required for constructing the Cardiac Risk Score2 include age, gender, systolic blood pressure, antihypertensive medication use, current smoking, diabetes, total cholesterol, and high-density-lipoprotein (HDL) cholesterol. As gender, total cholesterol, HDL cholesterol, and systolic blood pressure data were required for the development of the Cardiac Risk Score2, those without this data were excluded from the second analytic sample. The CONSORT (Consolidated Standards of Reporting Trials)-style diagram illustrates a comprehensive assessment of all missingness in the construction of Cardiac Risk Score1 and Cardiac Risk Score2 (Figure 1). We assigned an age value of 21 years. Following the convention of the existing CVD prediction algorithm developed based on data obtained from the Framingham Heart Study, we constructed a separate gender-specific multivariable risk function algorithm for men versus women [43]. The syntax for both Cardiac Risk Score1 and Cardiac Risk Score2 can be found in the Multimedia Appendix 1.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials)-style diagram illustrating a comprehensive assessment of all missingness in the construction of Cardiac Risk Score1 and Cardiac Risk Score2. CD4: cluster of differentiation 4; HDL: high-density lipoprotein; VL: viral load.



Data Analysis

We included youths living with HIV who had health records of systolic blood pressure, antihypertensive medication use, current smoking, and diabetes status in 2016, in the first analytic sample for Cardiac Risk Score1 (n=813). Subsequently, we constructed a second analytic sample for Cardiac Risk Score2 (n= 398) by including total cholesterol and HDL cholesterol in addition to the clinic-based predictors mentioned above. As not all patients had records for these additional variables, the sample size of Cardiac Risk Score2 was reduced to 398.

We performed a descriptive analysis of the demographic and clinical characteristics of the sample, followed by bivariate analyses using the chi-square test and independent samples *t* test with detectable VL and low CD4 lymphocyte count as a dichotomous variable. Then, we ran 3 multivariable linear regression models with Cardiac Risk Score1 as a continuous outcome variable to determine the association between detectable VL and cardiovascular risk, impaired immune function (CD4 lymphocyte count ≤200) and cardiovascular risk, and the combined effect of detectable VL and low CD4 lymphocyte count with cardiovascular risk. We repeated these regression models with Cardiac Risk Score2 as the outcome variable. Covariates included age, race and ethnicity, gender, being prescribed ART medication, substance dependence, and BMI.

The linear regression coefficients were evaluated using unstandardized beta (B) as the point estimate, CI was 95% for the variability around that point estimate, and *P* value for statistical significance. Standardized beta (β) was used to estimate the effect size of independent variables on the dependent variable, categorized as small (.10), medium (.30), and large (.50) according to Cohen's criteria. SPSS Statistics (version 25; IBM Corp) was used for all analyses.

Results

Overview

Table 1 provides a summary of the demographic and clinical characteristics of the analytic sample of the study, which includes 813 youths living with HIV. The mean age of the sample was 21 (SD 2.6) years. The sample was predominantly of racial and ethnic minorities, with 70.3% (570/811) of them identifying as Black and 10.9% (88/811) of them as Latinx. Additionally, the majority of the sample were men (563/813, 69.2%). The study also presents the prevalence of detectable VL and impaired immune health, indicating that 47.8% (366/766) of those with VL data had detectable VL and 8.6% (51/593) of those with CD4 lymphocyte data had a baseline count ≤200. The proportion of youths living with HIV who were currently taking ART was high at 88.6% (720/813). The mean BMI of the sample was 25 (SD 6.7) kg/m^2, and the mean Cardiac Risk Scores 1 and 2 were 0.062 (SD 0.039) and 0.672 (SD 0.401), respectively.

Table 1. Demographic and clinical characteristics of the sample (N=813).

Characteristics	Values
Race and ethnicity (n=811)	
Black	570 (70.3)
Latinx	88 (10.9)
White	81 (10)
Other	72 (8.9)
Gender (N=813)	
Man	563 (69.2)
Woman	250 (30.8)
ART[a] medication (N=813)	
Currently on ART medication	720 (88.6)
Not reportedly on ART medication	93 (11.4)
VL[b] (n=766)	
Detectable VL	366 (47.8)
Undetectable VL	400 (52.2)
CD4[c] lymphocyte count (n=593)	
Baseline count ≤200	51 (8.6)
Baseline count >200	542 (91.4)
Substance abuse diagnosis (N=813)	
Diagnosed with substance abuse	99 (12.2)
Not diagnosed with substance abuse	714 (87.8)
ATN[d] clinical site (N=813)	
Baltimore, Maryland	90 (11.1)
Birmingham, Alabama	61 (7.5)
Los Angeles, California	84 (10.3)
Memphis, Tennessee	187 (23)
San Diego, California	101 (12.4)
Tampa, Florida	219 (26.9)
Washington, District of Columbia	71 (8.7)
Age (years), mean (SD); median (range)	21 (2.6); 21 (14-26)
BMI (kg/m^2; N=806), mean (SD); median (range)	25 (6.7); 23.3 (4.8-69.5)
Cardiac Risk Score1[e], mean (SD); median (range)	0.062 (0.040); 0.061 (0.01-0.26)
Cardiac Risk Score2[f] (n=398), mean (SD); median (range)	0.672 (0.401); 0.569 (0.17-4.05)

[a]ART: antiretroviral therapy.

[b]VL: viral load.

[c]CD4: cluster of differentiation 4.

[d]ATN: Adolescent Medicine Trials Network.

[e]Cardiac Risk Score1: defined as the risk score for patients with systolic blood pressure, smoking, diabetes, and antihypertensive medication use.

[f]Cardiac Risk Score2: defined as the risk score for patients with systolic blood pressure, smoking, diabetes, antihypertensive medication use, total cholesterol, and high-density-lipoprotein cholesterol.

Bivariate Correlates of HIV and Immunologic Biomarkers

Demographic and Clinical Correlates

Table 2 presents findings on the demographic and clinical correlates of the study population; the table highlights several general trends observed in the data. First, patients who had detectable VL were more likely to be Black than those who had undetectable VL. The proportion of Black patients was 74% (271/366) among those with detectable VL, compared to 67% (268/400) among those with undetectable VL. This difference was statistically significant (χ^2_3=9.9; *P*<.05). Second, patients with detectable VL were more likely to be men than those without detectable VL. Specifically, 72.7% (266/366) of patients with detectable VL were men, compared to 65.2% (261/400) of patients without detectable VL. This difference was also statistically significant (χ^2_1=4.9; *P*=.03). Finally, patients with detectable VL were less likely to be on ART medication than those with undetectable VL. Specifically, 87.7% (321/366) of those with detectable VL were on ART medication, compared to 92.5% (370/400) of those with undetectable VL. This difference was statistically significant (χ^2_1=5.0; *P*=.03). These findings suggest that demographic and clinical factors are important correlates of VL in this population.

The 10-year risk of developing CVD was estimated using 2 cardiovascular risk assessment variables, Cardiac Risk Score1 and Cardiac Risk Score2. In bivariate analyses, we found significant associations between VL and cardiovascular risk as measured by both cardiac risk scores. Specifically, the results show that the association between VL and cardiovascular risk was statistically significant for Cardiac Risk Score1 (*P*=.001) and Cardiac Risk Score2 (*P*=.007). However, no significant differences were found between the demographic characteristics and the immunologic biomarker, CD4 lymphocyte count, as displayed in Table 3. When comparing patients who had a CD4 lymphocyte count of ≤200 with those who had a CD4 lymphocyte count of >200, no significant difference was found in Cardiac Risk Score1. However, the results show that with Cardiac Risk Score2, the risk of CVD was significantly increased in patients who had a CD4 lymphocyte count of ≤200 compared with those who had a CD4 lymphocyte count of >200 (*P*=.002). These findings suggest that VL is associated with an increased risk of CVD as measured by both Cardiac Risk Score1 and Cardiac Risk Score2, and that CD4 lymphocyte count may be an important factor in predicting cardiovascular risk using Cardiac Risk Score2.

Table 2. Bivariate analysis by detectable viral load (VL; N=766).

Characteristics	Total sample	Detectable VL (n=366)	Undetectable VL (n=400)	*F* test (*df*)	Chi-square (*df*)	*P* value
Race and ethnicity (n=764), n (%)				N/A[a]	9.881 (3)	.02
Black	539 (70.5)	271 (74.5)	268 (67)			
Latinx	85 (11.1)	43 (11.8)	42 (10.5)			
White	73 (9.6)	27 (7.4)	46 (11.5)			
Other	67 (8.8)	23 (6.3)	44 (11)			
Gender (N=766), n (%)				N/A	4.912 (1)	.03
Man	527 (68.8)	266 (72.7)	261 (65.2)			
Woman	239 (31.2)	100 (27.3)	139 (34.8)			
ART[b] medication (N=766), n (%)				N/A	4.975 (1)	.03
Currently on ART medication	691 (90.2)	321 (87.7)	370 (92.5)			
Not reportedly on ART medication	75 (9.8)	45 (12.3)	30 (7.5)			
CD4[c] lymphocyte count (n=593), n (%)				N/A	32.326 (1)	<.001
Baseline count ≤ 200	50 (8.6)	42 (15.8)	8 (2.5)			
Baseline count > 200	532 (91.4)	224 (84.2)	308 (97.5)			
Substance abuse diagnosis (N=766), n (%)				N/A	0.333 (1)	.56
Diagnosed with substance abuse	97 (12.7)	49 (13.4)	48 (12)			
Not diagnosed with substance abuse	669 (87.3)	317 (86.6)	352 (88)			
Age (years), mean (SD)	21 (2.6)	21 (2.5)	21 (2.7)	2.990 (764)	N/A	.58
BMI (kg/m^2; n=762), mean (SD)	25 (6.8)	25 (6.6)	26 (6.9)	3.626 (760)	N/A	.13
Cardiac Risk Score1[d], mean (SD)	0.062 (0.040)	0.067 (0.042)	0.058 (0.037)	2.049 (764)	N/A	.001
Cardiac Risk Score2[e] (n=394), mean (SD)	0.672 (0.672)	0.735 (0.452)	0.626 (0.348)	4.577 (392)	N/A	.007

[a]N/A: not applicable.

[b]ART: antiretroviral therapy.

[c]CD4: cluster of differentiation 4.

[d]Cardiac Risk Score1: defined as the risk score for patients with systolic blood pressure, smoking, diabetes, and antihypertensive medication use.

[e]Cardiac Risk Score2: defined as the risk score for patients with systolic blood pressure, smoking, diabetes, antihypertensive medication use, total cholesterol, and high-density-lipoprotein cholesterol.

Table 3. Bivariate analysis by low cluster of differentiation 4 (CD4) lymphocyte count (N=593).

Characteristics	Total sample	CD4 lymphocyte count≤200 (n=51)	CD4 lymphocyte count>200 (n=542)	*F* test (*df*)	Chi-square (*df*)	*P* value
Race and ethnicity (n=591), n (%)				N/A[a]	6.548 (3)	.09
Black	431 (72.9)	44 (88)	387 (71.5)			
Latinx	63 (10.7)	3 (6)	60 (11.1)			
White	48 (8.1)	1 (2)	47 (8.7)			
Other	49 (8.3)	2 (4)	47 (8.7)			
Gender (N=593), n (%)				N/A	0.002 (1)	.97
Man	420 (70.8)	36 (70.6)	384 (70.8)			
Woman	173 (29.2)	15 (29.4)	158 (29.2)			
ART[b] medication (N=593), n (%)				N/A	1.397 (1)	.24
Currently on ART medication	529 (89.2)	48 (94.1)	481 (88.7)			
Not reportedly on ART medication	64 (10.8)	3 (5.9)	61 (11.3)			
VL[c] (n=582), n (%)				N/A	32.326 (1)	<.001
Detectable VL	266 (45.7)	42 (84)	224 (42.1)			
Undetectable VL	316 (54.3)	224 (42.1)	308 (57.9)			
Substance abuse diagnosis (N=593), n (%)				N/A	0.445 (1)	.51
Diagnosed with substance abuse	86 (14.5)	9 (17.6)	77 (14.2)			
Not diagnosed with substance abuse	507 (85.5)	42 (82.4)	465 (85.8)			
Age (years), mean (SD)	21 (2.6)	21 (2.5)	21 (2.5)	0.042 (591)	N/A	.63
BMI (kg/m^2; n=591), mean (SD)	25 (6.8)	24 (5.8)	25 (7)	3.357 (589)	N/A	.09
Cardiac Risk Score1[d], mean (SD)	0.062 (0.039)	0.066 (0.046)	0.060 (0.037)	1.223 (591)	N/A	.35
Cardiac Risk Score2[e] (n=343), mean (SD)	0.672 (0.401)	0.864 (0.741)	0.621 (0.338)	8.712 (341)	N/A	.002

[a]N/A: not applicable.

[b]ART: antiretroviral therapy.

[c]VL: viral load.

[d]Cardiac Risk Score1: defined as the risk score for patients with systolic blood pressure, smoking, diabetes, and antihypertensive medication use.

[e]Cardiac Risk Score2: defined as the risk score for patients with systolic blood pressure, smoking, diabetes, antihypertensive medication use, total cholesterol, and high-density-lipoprotein cholesterol.

Multivariable Analyses of Cardiac Risk Scores

The multivariable regression analysis of Cardiac Risk Score1 and Cardiac Risk Score2 scores is presented in Tables 4 and 5. For Cardiac Risk Score1, Model 1A showed a small effect of VL on increased cardiovascular risk (β=.067, SE 0.001; *P*=.008), while no significant association was found for CD4 lymphocyte count (Model 1B). Model 1C also demonstrated a significant positive association between VL and cardiovascular risk (*P*=.007), but no significant association was found for CD4 lymphocyte count. The findings of Model 1A and Model 1C were consistent.

For Cardiac Risk Score2, Model 2A showed a 38% increase in the likelihood of having cardiovascular risk for a 1000-point increase in VL (B=.038, 95% CI 0.011-0.066). Model 2B showed no significant association between CD4 lymphocyte count and cardiovascular risk. Model 2C demonstrated a significant positive association between VL and cardiovascular risk (*P*=.01), while that with the CD4 lymphocyte count was not significant (*P*=.54). The findings of Model 2A and Model 2C were consistent. The standardized beta value for the effect of VL on cardiovascular risk was only over .1 in both Cardiac Risk Score1 and Cardiac Risk Score2, indicating a small effect size.

Table 4. Multivariable linear regression Model 1 by Cardiac Risk Scores. Cardiac Risk Score1: defined as the risk score for patients with systolic blood pressure, smoking, diabetes, and antihypertensive medication use.

Multivariable	Model 1A with VL[a] findings					Model 1B with CD4[b] lymphocyte findings					Model 1C with VL and CD4 lymphocyte findings				
	Unstandardized			Standardized		Unstandardized			Standardized		Unstandardized			Standardized	
	B	SE	95% CI	β	*P* value	B	SE	95% CI	β	*P* value	B	SE	95% CI	β	*P* value
Age	.001	0.000	0.000 to 0.002	.072	.005	.002	0.000	0.001 to 0.003	.123	<.001	.002	0.000	0.001 to 0.003	.135	<.001
Race and ethnicity	.002	0.001	0.000 to 0.004	.056	.03	.002	0.001	0.000 to 0.004	.043	.12	.001	0.001	–0.001 to 0.004	.036	.22
Gender	.061	0.002	0.057 to 0.065	.710	<.001	.058	0.002	0.053 to 0.063	.691	<.001	.059	0.003	0.054 to 0.064	.696	<.001
ART[c] medication	.001	0.003	–0.005 to 0.008	.010	.70	–.001	0.003	–0.008 to 0.006	–.007	.81	.001	0.004	–0.006 to 0.008	.006	.83
Substance abuse diagnosis	.015	0.003	0.009 to 0.020	.124	<.001	.017	0.003	0.011 to 0.023	.159	<.001	.015	0.003	0.008 to 0.021	.136	<.001
BMI	.000	0.000	0.000 to 0.001	.057	.03	.000	0.000	0.000 to 0.001	.055	.06	.000	0.000	0.000 to 0.001	.053	.08
VL	.002	0.001	0.000 to 0.003	.067	.008	N/A[d]	N/A	N/A	N/A	N/A	.002	0.001	0.001 to 0.004	.088	.007
CD4 lymphocyte count	N/A	N/A	N/A	N/A	N/A	.000	0.000	–0.001 to 0.000	–.028	.32	.000	0.000	–0.001 to 0.001	.015	.65

[a]VL: viral load.

[b]CD4: cluster of differentiation 4.

[c]ART: antiretroviral therapy.

[d]N/A: not applicable.

Table 5. Multivariable linear regression Model 2 by Cardiac Risk Scores. Cardiac Risk Score2: defined as the risk score for patients with systolic blood pressure, smoking, diabetes, antihypertensive medication use, total cholesterol, and high-density-lipoprotein cholesterol.

Multivariable	Model 2A with VL[a] findings					Model 2B with CD4[b] lymphocyte findings					Model 2C with VL and CD4 lymphocyte findings				
	Unstandardized			Standardized		Unstandardized			Standardized		Unstandardized			Standardized	
	B	SE	95% CI	β	*P* value	B	SE	95% CI	β	*P* value	B	SE	95% CI	β	*P* value
Age	.012	0.008	–0.003 to 0.027	.081	.10	.023	0.008	0.007 to 0.039	.152	.004	.024	0.008	0.008 to 0.040	.158	.003
Race and ethnicity	.039	0.021	–0.002 to 0.081	.091	.06	.040	0.022	–0.002 to 0.082	.096	.06	.045	0.022	0.002 to 0.088	.105	.04
Gender	.164	0.044	0.077 to 0.251	.188	<.001	.163	0.045	0.074 to 0.252	.194	<.001	.170	0.046	0.080 to 0.259	.201	<.001
ART[c] medication	–.030	0.073	–0.174 to 0.114	–.020	.68	–.080	0.070	–0.218 to 0.057	–.059	.25	–.069	0.071	–0.209 to 0.070	–.050	.33
Substance abuse diagnosis	.059	0.055	–0.049 to 0.166	.053	.29	.074	0.054	–0.032 to 0.180	.071	.17	.055	0.054	–0.053 to 0.162	.053	.32
BMI	.014	0.003	0.008 to 0.019	.245	<.001	.013	0.003	0.007 to 0.018	.236	<.001	.013	0.003	0.007 to 0.018	.240	<.001
VL	.038	0.014	0.011 to 0.066	.134	.006	N/A[d]	N/A	N/A	N/A	N/A	.040	0.016	0.009 to 0.071	.146	.01
CD4 lymphocyte count	N/A	N/A	N/A	N/A	N/A	–.012	0.007	–0.025 to 0.001	–.095	.07	–.005	0.007	–0.019 to 0.010	–.035	.54

[a]VL: viral load.

[b]CD4: cluster of differentiation 4.

[c]ART: antiretroviral therapy.

[d]N/A: Not applicable.

Discussion

This study was among the first to examine the creation of cardiovascular risk profiles and assess their connections with HIV biomarkers in youths living with HIV aged 14-26 years. The results showed that detectable VL was statistically significantly linked to cardiovascular risk, even after adjusting for demographic and clinical factors, underscoring the need for integrating cardiovascular health and HIV care in regular clinical practice. The study also revealed that higher plasma VL was associated with a slight yet statistically significant increase in cardiovascular risk, regardless of CD4 lymphocyte count, ART exposure, and other demographic and clinical factors. These findings contribute to our understanding of the broad spectrum of the VL-cardiovascular risk relationship. In contrast, CD4 lymphocyte count did not demonstrate a connection to cardiovascular risk in any of the adjusted models.

There is currently no universally accepted method for evaluating cardiovascular risk and conveying this risk to youths living with HIV. The conventional approach of expressing increased risk of CVD as a probability of an event over the subsequent 10 years may discourage patients from adhering to healthier lifestyles and preventive care [54]. Using average age as a constant factor (the same for everyone) in our risk scores could pave the way for the development of an EHR-integrated monitoring device that assesses cardiovascular risk in young HIV-positive populations and enables effective risk communication to youths living with HIV. This approach could also have significant implications for clinical decision-making regarding treatment thresholds in youths living with HIV. Since age is a key factor in predicting absolute risk [46], the use of our risk scores could help classify youths living with HIV who are at risk of cardiovascular health issues. Our study provides a foundation for future research that can further investigate this approach and apply it to HIV-negative youth populations. As we obtained similar results from both risk scores, the simpler version could serve as a cost-effective means of assessing high-risk HIV-positive youth for cardiovascular health conditions using readily available clinic-based predictors. However, it is crucial for researchers to periodically assess potential biases that may not have been apparent in this study.

Previous studies have indicated that the FRS is not effective in categorizing lifetime risk in younger individuals [55] and have recommended relative risk estimates instead of age-dependent absolute risk estimates for those with low short-term risk [56]. A study that examined the ability of the FRS and the Adult Treatment Panel III to predict long-term risk for coronary heart disease death in young men aged 18-39 years found that neither method identified individuals under 30 years of age as high risk despite significant risk factor burden [57]. In this study, age was kept constant, meaning all individuals aged 18-29 years were given the risk estimate of a 30-year-old [57]. However, using continuous risk scores in our study could offer advantages over arbitrary classifications of high cardiovascular risk. In a cohort of youths living with HIV with lower event rates than the original Framingham cohort, identifying only high-risk individuals based on a >20% absolute risk in 10 years, despite significant risk factor burden, may not be the most effective strategy for estimating and communicating cardiovascular risk to younger individuals.

The FRS, on the other hand, has its own set of advantages. Given its strengths, we opted to adapt the FRS as a tool to assess CVD risk among youths living with HIV, despite the existence of other risk assessment algorithms. FRS was developed using a large, community-based cohort of individuals from Framingham, Massachusetts [58]. This population was representative of the US population at the time the study was conducted, which is important because HIV-positive individuals in the United States may have different risk profiles compared to those in other regions. FRS is based on traditional CVD risk factors, such as age, gender, blood pressure, total cholesterol, HDL cholesterol, and smoking status [58]. These risk factors are commonly assessed in clinical settings and are readily available in EHRs, making FRS a feasible tool for use in routine clinical care for youths living with HIV. FRS is a simple tool that is easy to use and understand, which may be particularly important for health care providers who may not have extensive training in CVD risk assessment. Other risk assessment algorithms may also be appropriate depending on the specific characteristics of the population being assessed. Nonetheless, there is still a need for further research to establish a definitive and widely accepted standard for assessing the risk of CVD in young adults living with HIV.

Despite the use of a multiclinic sample, there are potential limitations to the current findings. First, the study design was cross-sectional, preventing the establishment of causal inferences. Future research should adopt a longitudinal approach to observe changes over time and test these associations. Second, caution should be exercised when generalizing our findings to the wider population of youths living with HIV in the United States or other locations. For instance, the high proportion of Hispanics in Tampa and a large Black population in Memphis, Tennessee (over 50% of the population), may limit the generalizability of our results to areas with different demographic compositions. Third, we were unable to provide the lower limit detection threshold of VL due to the use of different laboratories and assays across clinical sites, resulting in varied detection limits. Fourth, the use of diagnostic codes as a proxy for substance abuse may have underestimated the prevalence of alcohol and drug use problems. This limitation is further compounded by the high number of missing data, particularly relating to cholesterol, which led to the exclusion of youth from 3 of the highest prevalence areas in the United States. Overall, these limitations suggest that caution should be exercised when interpreting the current findings, and further studies that prioritize more complete EHR data are necessary to address these limitations.

Furthermore, there are several limitations to grouping substances together in our study using EHR data. Combining substances into 1 group may mask potential differences in the cardiovascular effects of individual substances, limiting our ability to identify specific risk factors associated with each substance. This could lead to overgeneralization of findings and inaccurate conclusions about the effects of substance abuse on cardiovascular health. Another limitation of our study is that missing values were assumed as nonsubstance users. This assumption may lead to an underestimation of the prevalence of substance abuse in our sample, as there are issues around documentation and disclosure of alcohol and drug use in clinical settings, which may result in incomplete or inaccurate information in EHR data. This limitation could have implications for the generalizability of our findings, as it may not reflect the true prevalence of substance abuse in the population under study.

Additionally, assuming missing values indicate nonsubstance use could introduce bias if the reasons for missing data are related to substance abuse. For example, individuals who are actively using substances may be less likely to disclose their use or may miss appointments or follow-up visits, leading to missing data. This bias could lead to inaccurate estimates of the prevalence and effects of substance abuse on cardiovascular health. Therefore, while grouping substances and assuming missing values as nonsubstance users may be practical approaches for analyzing EHR data, they do have limitations that need to be considered when interpreting the results of our study.

Despite the limitations, our study is innovative in using EHRs to investigate the independent effects of HIV biomarkers on cardiovascular risk in a US-based cohort of youths living with HIV who received routine medical care. The use of EHRs allowed for the inclusion of HIV outcome data from clinic-based patients, which reduced study costs, and the use of ICD-10 codes to identify clinical covariates for analysis. Our findings have important implications for the development of a multivariable risk assessment method tailored to youths living with HIV. Future research should investigate the inclusion of VL in cardiovascular risk equations for young people to predict CVD risk. Our study emphasizes the importance of recognizing HIV infection as an additional risk factor for CVD and providing preventive CVD care for youths living with HIV in routine practice. Multicomponent interventions that target both VL suppression and cardiovascular risk reduction among youths living with HIV are warranted [25].

Our study also highlights the need for a preventative health life course approach in the care of youths living with HIV. A recently published cohort study found that the risk of CVD remained consistently higher among people with HIV, regardless of age or diagnosis timing [59]. It is important to note that HIV alone, without consideration of other comorbidities and risk factors, may underestimate the burden of CVD among young people living with HIV [1,25]. Therefore, interventions designed with a preventative health life course approach should consider those with particularly elevated cardiovascular risk, especially if risk-enhancing factors related to HIV (eg, low CD4 lymphocyte count or a history of prolonged viremia) are present [60]. Similarly, a syndemics approach is necessary to combat the growing burden of CVD among young adults with HIV.

Syndemics refer to the interaction between 2 or more epidemics that mutually reinforce each other and increase the burden of disease in a population [61]. In the case of youths living with HIV, the interaction between HIV and CVD epidemics is evident, as HIV infection increases the risk of developing CVD, while CVD risk factors are prevalent among individuals with

HIV [4,6,25]. Therefore, a comprehensive approach that addresses the interrelated factors contributing to the high burden of CVD in youths living with HIV is necessary. This approach should include interventions that target HIV treatment and management, CVD prevention, and the social determinants of health. By addressing these factors in a coordinated manner, health care providers can reduce the burden of CVD and improve the overall health outcomes of young adults with HIV.

Acknowledgments

This work was supported by the NIH Adolescent Medicine Trials Network for HIV/AIDS Interventions (ATN 154) as part of the FSU/CUNY *Scale It Up* Program (U19HD089875; principal investigator: SN), as well as through a career development award from the National Institute of Mental Health (K01MH116737; principal investigator: HB). The content is solely the responsibility of the authors and does not represent the official views of the funding agencies. This research was presented at the 2020 Conference on Retroviruses and Opportunistic Infections (CROI). We acknowledge the contributions of the *SIU* sites, particularly Aditya Gaur, MD, of St. Jude Children's Research Hospital; Allison Agwu, MD, of the Johns Hopkins University School of Medicine; Marvin Belzer, MD, of Children's Hospital Los Angeles; Patricia Emmanuel, MD, of University of South Florida; Lawrence D'Angelo, MD, of Children's National Health System; Stephen Spector, MD, of the UC San Diego School of Medicine; Jeffrey Birnbaum, MD, of SUNY Downstate Medical Center; Lawrence Friedman, MD, of University of Miami Health System; Marne Castillo, PhD and Mary Tanney, RN, of the Children's Hospital of Philadelphia; and Sonia Lee, PhD, of the National Institute of Child Health and Human Development. We also thank Demetria Cain, PhD, Brett Millar, PhD, Rafael Dilones, BSc, Richard Zhao, BA, Hector Moran, BS, and Jesse Bradford-Rogers, MPH of PRIDE Hunter College.

Data Availability

The corresponding author can provide the data sets that were generated and analyzed during this study upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Syntax for Cardiac Risk Score1 and Cardiac Risk Score2.
[DOCX File , 13 KB - ijmr_v12i1e41574_app1.docx]

References

1. Budhwani H, Robles G, Starks TJ, MacDonell KK, Dinaj V, Naar S. Healthy choices intervention is associated with reductions in stigma among youth living with HIV in the United States (ATN 129). AIDS Behav 2021;25(4):1094-1102 [FREE Full text] [doi: 10.1007/s10461-020-03071-1] [Medline: 33098483]
2. Hoffmann U, Lu MT, Foldyna B, Zanni MV, Karady J, Taron J, REPRIEVE trial. Assessment of coronary artery disease with computed tomography angiography and inflammatory and immune activation biomarkers among adults with HIV eligible for primary cardiovascular prevention. JAMA Netw Open 2021;4(6):e2114923 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.14923] [Medline: 34185068]
3. Marcus JL, Leyden WA, Alexeeff SE, Anderson AN, Hechter RC, Hu H, et al. Comparison of overall and comorbidity-free life expectancy between insured adults with and without HIV infection, 2000-2016. JAMA Netw Open 2020;3(6):e207954 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.7954] [Medline: 32539152]
4. Patel K, Wang J, Jacobson DL, Lipshultz SE, Landy DC, Geffner ME, Pediatric HIV/AIDS Cohort Study (PHACS). Aggregate risk of cardiovascular disease among adolescents perinatally infected with the human immunodeficiency virus. Circulation 2014;129(11):1204-1212 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.113.001978] [Medline: 24366631]
5. Ryscavage P, Still W, Nyemba V, Stafford K. Prevalence of systemic hypertension among HIV-infected and HIV-uninfected young adults. Open Forum Infect Dis 2017;4(Suppl 1):S59 [FREE Full text] [doi: 10.1093/ofid/ofx162.138]
6. Arrive E, Viard JP, Salanave B, Dollfus C, Matheron S, Reliquet V, et al. Metabolic risk factors in young adults infected with HIV since childhood compared with the general population. PLoS ONE 2018;13(11):e0206745 [FREE Full text] [doi: 10.1371/journal.pone.0206745] [Medline: 30408056]
7. Chen M, Rhodes PH, Hall IH, Kilmarx PH, Branson BM, Valleroy LA, Centers for Disease Control Prevention (CDC). Prevalence of undiagnosed HIV infection among persons aged =13 years--national HIV surveillance system, United States, 2005-2008. MMWR Suppl 2012;61(2):57-64. [Medline: 22695465]
8. Wong VJ, Murray KR, Phelps BR, Vermund SH, McCarraher DR. Adolescents, young people, and the 90-90-90 goals: a call to improve HIV testing and linkage to treatment. AIDS 2017;31(Suppl 3):S191-S194 [FREE Full text] [doi: 10.1097/QAD.0000000000001539] [Medline: 28665876]

9. Zanoni BC, Mayer KH. The adolescent and young adult HIV cascade of care in the United States: exaggerated health disparities. AIDS Patient Care STDS 2014;28(3):128-135 [FREE Full text] [doi: 10.1089/apc.2013.0345] [Medline: 24601734]
10. Koenig LJ, Hoyer D, Purcell DW, Zaza S, Mermin J. Young people and HIV: a call to action. Am J Public Health 2016;106(3):402-405. [doi: 10.2105/AJPH.2015.302979] [Medline: 26794156]
11. Crepaz N, Dong X, Wang X, Hernandez AL, Hall HI. Racial and ethnic disparities in sustained viral suppression and transmission risk potential among persons receiving HIV care - United States, 2014. MMWR Morb Mortal Wkly Rep 2018;67(4):113-118 [FREE Full text] [doi: 10.15585/mmwr.mm6704a2] [Medline: 29389918]
12. Burkholder GA, Tamhane AR, Safford MM, Muntner PM, Willig AL, Willig JH, et al. Racial disparities in the prevalence and control of hypertension among a cohort of HIV-infected patients in the southeastern United States. PLoS One 2018;13(3):e0194940 [FREE Full text] [doi: 10.1371/journal.pone.0194940] [Medline: 29596462]
13. HIV and Youth. Centers for Disease Control and Prevention. 2018. URL: https://www.cdc.gov/healthyyouth/youth_hiv/hiv-information-and-youth.htm [accessed 2023-06-16]
14. Benjamin EJ, Muntner P, Alonso A, Bittencourt MS, Callaway CF, Carson AP, American Heart Association Council on Epidemiology Prevention Statistics Committee Stroke Statistics Subcommittee. Heart disease and stroke statistics-2019 update: a report from the American heart association. Circulation 2019;139(10):e56-e528 [FREE Full text] [doi: 10.1161/CIR.0000000000000659] [Medline: 30700139]
15. Clark R, Anderson NB, Clark VR, Williams DR. Racism as a stressor for African Americans. A biopsychosocial model. Am Psychol 1999;54(10):805-816. [doi: 10.1037//0003-066x.54.10.805] [Medline: 10540593]
16. Carnethon MR, Pu J, Howard G, Albert MA, Anderson CAM, Bertoni AG, American Heart Association Council on Epidemiology and Prevention, Council on Cardiovascular Disease in the Young, Council on Cardiovascular and Stroke Nursing, Council on Clinical Cardiology, Council on Functional Genomics and Translational Biology, Stroke Council. Cardiovascular health in African Americans: a scientific statement from the American heart association. Circulation 2017;136(21):e393-e423 [FREE Full text] [doi: 10.1161/CIR.0000000000000534] [Medline: 29061565]
17. Al-Kindi SG, ElAmm C, Ginwalla M, Mehanna E, Zacharias M, Benatti R, et al. Heart failure in patients with human immunodeficiency virus infection: epidemiology and management disparities. Int J Cardiol 2016;218:43-46 [FREE Full text] [doi: 10.1016/j.ijcard.2016.05.027] [Medline: 27232910]
18. Guaraldi G, Orlando G, Zona S, Menozzi M, Carli F, Garlassi E, et al. Premature age-related comorbidities among HIV-infected persons compared with the general population. Clin Infect Dis 2011;53(11):1120-1126 [FREE Full text] [doi: 10.1093/cid/cir627] [Medline: 21998278]
19. Krikke M, Arends JE, Van Lelyveld S, Hoepelman A, Visseren F. Greater carotid intima media thickness at a younger age in HIV-infected patients compared with reference values for an uninfected cohort. HIV Med 2017;18(4):275-283 [FREE Full text] [doi: 10.1111/hiv.12428] [Medline: 27477496]
20. Krikke M, van Lelyveld SFL, Tesselaar K, Arends JE, Hoepelman IM, Visseren FLJ. The role of T cells in the development of cardiovascular disease in HIV-infected patients. Atherosclerosis 2014;237(1):92-98. [doi: 10.1016/j.atherosclerosis.2014.08.054] [Medline: 25238214]
21. Quiros-Roldan E, Raffetti E, Focà E, Brianese N, Ferraresi A, Paraninfo G, et al. Incidence of cardiovascular events in HIV-positive patients compared to general population over the last decade: a population-based study from 2000 to 2012. AIDS Care 2016;28(12):1551-1558. [doi: 10.1080/09540121.2016.1198750] [Medline: 27321070]
22. Triant V. Epidemiology of coronary heart disease in HIV patients. Rev Cardiovasc Med 2014;15(0 1):S1-S8.
23. De Socio GVL, Martinelli L, Morosi S, Fiorio M, Roscini AR, Stagni G, et al. Is estimated cardiovascular risk higher in HIV-infected patients than in the general population? Scand J Infect Dis 2007;39(9):805-812. [doi: 10.1080/00365540701230884] [Medline: 17701720]
24. Drozd DR, Kitahata MM, Althoff KN, Zhang J, Gange SJ, Napravnik S, et al. Increased risk of myocardial infarction in HIV-infected individuals in North America compared with the general population. J Acquir Immune Defic Syndr 2017;75(5):568-576 [FREE Full text] [doi: 10.1097/QAI.0000000000001450] [Medline: 28520615]
25. Budhwani H, Bulls M, Naar S. Proof of concept for the FLEX intervention: feasibility of home based coaching to improve physical activity outcomes and viral load suppression among African American youth living with HIV. J Int Assoc Provid AIDS Care 2021;20:2325958220986264 [FREE Full text] [doi: 10.1177/2325958220986264] [Medline: 33406973]
26. Hsue PY, Waters DD. Time to recognize HIV infection as a major cardiovascular risk factor. Circulation 2018;138(11):1113-1115 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.118.036211] [Medline: 30354392]
27. Hunt PW. HIV and inflammation: mechanisms and consequences. Curr HIV/AIDS Rep 2012;9(2):139-147. [doi: 10.1007/s11904-012-0118-8] [Medline: 22528766]
28. Lang S, Mary-Krause M, Cotte L, Gilquin J, Partisani M, Simon A, Clinical Epidemiology Group of the French Hospital Database on HIV. Impact of individual antiretroviral drugs on the risk of myocardial infarction in human immunodeficiency virus-infected patients: a case-control study nested within the French hospital database on HIV ANRS cohort CO4. Arch Intern Med 2010;170(14):1228-1238. [doi: 10.1001/archinternmed.2010.197] [Medline: 20660842]

29. Triant VA, Lee H, Hadigan C, Grinspoon SK. Increased acute myocardial infarction rates and cardiovascular risk factors among patients with human immunodeficiency virus disease. J Clin Endocrinol Metab 2007;92(7):2506-2512 [FREE Full text] [doi: 10.1210/jc.2006-2190] [Medline: 17456578]
30. Earnshaw VA, Smith LR, Chaudoir SR, Amico KR, Copenhaver MM. HIV stigma mechanisms and well-being among PLWH: a test of the HIV stigma framework. AIDS Behav 2013;17(5):1785-1795 [FREE Full text] [doi: 10.1007/s10461-013-0437-9] [Medline: 23456594]
31. Kim S, Chang Y, Cho J, Hong YS, Zhao D, Kang J, et al. Life's simple 7 cardiovascular health metrics and progression of coronary artery calcium in a low-risk population. Arterioscler Thromb Vasc Biol 2019;39(4):826-833 [FREE Full text] [doi: 10.1161/ATVBAHA.118.311821] [Medline: 30700133]
32. Kinno M, Gardin JM. American Heart Association's Life's Simple 7: Avoiding Heart Failure and Preserving Cardiac Structure and Function. American College of Cardiology. 2016. URL: https://www.acc.org/latest-in-cardiology/articles/2016/05/27/01/55/ahas-lifes-simple-7 [accessed 2023-06-16]
33. Lloyd-Jones DM, Hong Y, Labarthe D, Mozaffarian D, Appel LJ, Van Horn L, American Heart Association Strategic Planning Task Force and Statistics Committee. Defining and setting national goals for cardiovascular health promotion and disease reduction: the American heart association's strategic impact goal through 2020 and beyond. Circulation 2010;121(4):586-613 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.109.192703] [Medline: 20089546]
34. Razavi AC, Gingras V, Michos ED, Navar AM, Brown SA, Delker E, et al. American heart association EPI|lifestyle scientific sessions: 2020 meeting highlights. J Am Heart Assoc 2020;9(12):e017252 [FREE Full text] [doi: 10.1161/JAHA.120.017252] [Medline: 32476542]
35. Makarem N, St-Onge MP, Liao M, Lloyd-Jones DM, Aggarwal B. Association of sleep characteristics with cardiovascular health among women and differences by race/ethnicity and menopausal status: findings from the American heart association go red for women strategically focused research network. Sleep Health 2019;5(5):501-508 [FREE Full text] [doi: 10.1016/j.sleh.2019.05.005] [Medline: 31302068]
36. Folsom AR, Shah AM, Lutsey PL, Roetker NS, Alonso A, Avery CL, et al. American heart association's life's simple 7: avoiding heart failure and preserving cardiac structure and function. Am J Med 2015;128(9):970-976.e2 [FREE Full text] [doi: 10.1016/j.amjmed.2015.03.027] [Medline: 25908393]
37. Ogunmoroti O, Oni E, Michos ED, Spatz ES, Allen NB, Rana JS, et al. Life's simple 7 and incident heart failure: the multi-ethnic study of atherosclerosis. J Am Heart Assoc 2017;6(6):e005180 [FREE Full text] [doi: 10.1161/JAHA.116.005180] [Medline: 28655734]
38. Lloyd-Jones DM. Cardiovascular health and protection against CVD: more than the sum of the parts? Circulation 2014;130(19):1671-1673 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.114.012869] [Medline: 25273999]
39. Wood DA. Guidelines on cardiovascular risk assessmentand management. Eur Heart J Suppl 2005;7:L5-L10 [FREE Full text] [doi: 10.1093/eurheartj/sui079]
40. Genest J, Cohn JS. Clustering of cardiovascular risk factors: targeting high-risk individuals. Am J Cardiol 1995;76(2):8A-20A. [doi: 10.1016/s0002-9149(05)80010-4] [Medline: 7604805]
41. Bøg-Hansen E, Lindblad U, Bengtsson K, Ranstam J, Melander A, Råstam L. Risk factor clustering in patients with hypertension and non-insulin-dependent diabetes mellitus. The Skaraborg hypertension project. J Intern Med 1998;243(3):223-232 [FREE Full text] [doi: 10.1046/j.1365-2796.1998.00286.x] [Medline: 9627160]
42. Lloyd-Jones DM, Leip EP, Larson MG, D'Agostino RB, Beiser A, Wilson PWF, et al. Prediction of lifetime risk for cardiovascular disease by risk factor burden at 50 years of age. Circulation 2006;113(6):791-798 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.105.548206] [Medline: 16461820]
43. D'Agostino RB, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care: the Framingham heart study. Circulation 2008;117(6):743-753 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.107.699579] [Medline: 18212285]
44. Paisible AL, Chang CCH, So-Armah KA, Butt AA, Leaf DA, Budoff M, et al. HIV infection, cardiovascular disease risk factor profile, and risk for acute myocardial infarction. J Acquir Immune Defic Syndr 2015;68(2):209-216 [FREE Full text] [doi: 10.1097/QAI.0000000000000419] [Medline: 25588033]
45. Batsis JA, Lopez-Jimenez F. Cardiovascular risk assessment--from individual risk prediction to estimation of global risk and change in risk in the population. BMC Med 2010;8:29 [FREE Full text] [doi: 10.1186/1741-7015-8-29] [Medline: 20500815]
46. Wilson PWF, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of coronary heart disease using risk factor categories. Circulation 1998;97(18):1837-1847 [FREE Full text] [doi: 10.1161/01.cir.97.18.1837] [Medline: 9603539]
47. Sainz T, Serrano-Villar S, Mellado MJ. Prevalence of elevated blood pressure in HIV-infected children, adolescents and young adults. Pediatr Infect Dis J 2016;35(7):824-825 [FREE Full text] [doi: 10.1097/INF.0000000000001178] [Medline: 27351363]
48. Franks PW, Hanson RL, Knowler WC, Sievers ML, Bennett PH, Looker HC. Childhood obesity, other cardiovascular risk factors, and premature death. N Engl J Med 2010;362(6):485-493 [FREE Full text] [doi: 10.1056/NEJMoa0904130] [Medline: 20147714]

49. Skinner AC, Perrin EM, Moss LA, Skelton JA. Cardiometabolic risks and severity of obesity in children and young adults. N Engl J Med 2015;373(14):1307-1317 [FREE Full text] [doi: 10.1056/NEJMoa1502821] [Medline: 26422721]
50. Naar S, Parsons JT, Stanton BF. Adolescent trials network for HIV-AIDS scale it up program: protocol for a rational and overview. JMIR Res Protoc 2019;8(2):e11204 [FREE Full text] [doi: 10.2196/11204] [Medline: 30707102]
51. Pennar AL, Dark T, Simpson KN, Gurung S, Cain D, Fan C, et al. Cascade monitoring in multidisciplinary adolescent HIV care settings: protocol for utilizing electronic health records. JMIR Res Protoc 2019;8(5):e11185 [FREE Full text] [doi: 10.2196/11185] [Medline: 31148543]
52. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. Department of Health and Human Services. 2016. URL: https://aidsetc.org/resource/guidelines-use-antiretroviral-agents-hiv-1-infected-adults-and-adolescents [accessed 2023-06-16]
53. Gurung S, Ventuneac A, Cain D, Mirzayi C, Ferraris C, Rendina HJ, et al. Alcohol and substance use diagnoses among HIV-positive patients receiving care in NYC clinic settings. Drug Alcohol Depend 2017;180:62-67 [FREE Full text] [doi: 10.1016/j.drugalcdep.2017.07.034] [Medline: 28881318]
54. Blaha MJ, Naazie IN, Cainzos-Achirica M, Dardari ZA, DeFilippis AP, McClelland RL, et al. Derivation of a coronary age calculator using traditional risk factors and coronary artery calcium: the multi-ethnic study of atherosclerosis. J Am Heart Assoc 2021;10(6):e019351 [FREE Full text] [doi: 10.1161/JAHA.120.019351] [Medline: 33663219]
55. Lloyd-Jones DM, Wilson PWF, Larson MG, Beiser A, Leip EP, D'Agostino RB, et al. Framingham risk score and prediction of lifetime risk for coronary heart disease. Am J Cardiol 2004;94(1):20-24. [doi: 10.1016/j.amjcard.2004.03.023] [Medline: 15219502]
56. Ridker PM, Cook N. Should age and time be eliminated from cardiovascular risk prediction models? Rationale for the creation of a new national risk detection program. Circulation 2005;111(5):657-658 [FREE Full text] [doi: 10.1161/01.CIR.0000154544.90488.52] [Medline: 15699285]
57. Berry JD, Lloyd-Jones DM, Garside DB, Greenland P. Framingham risk score and prediction of coronary heart disease death in young men. Am Heart J 2007;154(1):80-86 [FREE Full text] [doi: 10.1016/j.ahj.2007.03.042] [Medline: 17584558]
58. Dawber TR, Meadors GF, Moore FE. Epidemiological approaches to heart disease: the Framingham study. Am J Public Health Nations Health 1951;41(3):279-286 [FREE Full text] [doi: 10.2105/ajph.41.3.279] [Medline: 14819398]
59. Gooden TE, Gardner M, Wang J, Jolly K, Lane DA, Benjamin LA, et al. Incidence of cardiometabolic diseases in people with and without human immunodeficiency virus in the United Kingdom: a population-based matched cohort study. J Infect Dis 2022;225(8):1348-1356 [FREE Full text] [doi: 10.1093/infdis/jiab420] [Medline: 34417792]
60. Feinstein MJ, Hsue PY, Benjamin LA, Bloomfield GS, Currier JS, Freiberg MS, et al. Characteristics, prevention, and management of cardiovascular disease in people living with HIV: a scientific statement from the American heart association. Circulation 2019;140(2):e98-e124 [FREE Full text] [doi: 10.1161/CIR.0000000000000695] [Medline: 31154814]
61. Singer M. Introduction to Syndemics: A Critical Systems Approach to Public and Community Health. San Francisco, CA: Jossey-Bass; 2009.

Abbreviations

ART: antiretroviral therapy
CD4: cluster of differentiation 4
CONSORT: Consolidated Standards of Reporting Trials
CVD: cardiovascular disease
EHR: electronic health record
FRS: Framingham Risk Score
HDL: high-density lipoprotein
HIPAA: Health Insurance Portability and Accountability Act
ICD-10: International Statistical Classification of Disease and Related Health Problems Tenth Revision
IRB: institutional review board
LS7: American Heart Association's Life's Simple 7
VL: viral load

Edited by A Mavragani; submitted 31.07.22; peer-reviewed by K Haubrick, R Kakumanu, Q Mengbing; comments to author 12.12.22; revised version received 27.03.23; accepted 07.04.23; published 16.08.23.

Please cite as:
Gurung S, Simpson KN, Grov C, Rendina HJ, Huang TTK, Budhwani H, Jones SS, Dark T, Naar S
Cardiovascular Risk Assessment Among Adolescents and Youths Living With HIV: Evaluation of Electronic Health Record Findings and Implications
Interact J Med Res 2023;12:e41574
URL: https://www.i-jmr.org/2023/1/e41574
doi:10.2196/41574
PMID:37585242

Original Paper

Correlation Between the Social Network Structure and Well-Being of Health Care Workers in Intensive Care Units: Prospective Observational Study

Ryo Esumi[1,2*], MD; Asami Ito-Masui[1,3*], MD, PhD; Eiji Kawamoto[1,4*], MD, PhD; Mami Ito[3], RN; Tomoyo Hayashi[5], RN; Toru Shinkai[3], MD; Atsuya Hane[3], MD; Fumito Okuno[3,4], MD; Eun Jeong Park[1], PhD; Ryuji Kaku[4], MD, PhD; Motomu Shimaoka[1*], MD, PhD

[1]Department of Molecular Pathobiology and Cell Adhesion Biology, Mie University Graduate School of Medicine, Tsu, Japan

[2]Department of Emergency Medicine, National Hospital Organization Mie Chuo Medical Center, Tsu, Japan

[3]Department of Emergency and Disaster Medicine, Mie University Hospital, Tsu, Japan

[4]Department of Anesthesiology, Mie University Hospital, Tsu, Japan

[5]Department of Nursing, Mie University Hospital, Tsu, Japan

[*]these authors contributed equally

Corresponding Author:
Eiji Kawamoto, MD, PhD
Department of Molecular Pathobiology and Cell Adhesion Biology
Mie University Graduate School of Medicine
2-174 Edobashi
Tsu, 514-0001
Japan
Phone: 81 59 232 1111
Email: a-2@umin.net

Abstract

Background: Effective communication strategies are becoming increasingly important in intensive care units (ICUs) where patients at high risk are treated. Distributed leadership promotes effective communication among health care professionals (HCPs). Moreover, beyond facilitating patient care, it may improve well-being among HCPs by fostering teamwork. However, the impact of distributed leadership on the communication structure and well-being of HCPs remains unclear.

Objective: We performed a social network analysis (SNA) to assess the characteristics of each HCP in the network, identify the number of HCP connections, analyze 4 centralities that can measure an HCP's importance, and evaluate the impact of distributed leadership structure on the well-being and communication structure of the medical staff.

Methods: Wearable sensors were used to obtain face-to-face interaction data from the ICU medical staff at Mie University Hospital, Japan. Participants wore a badge on the front of their clothing during working hours to measure the total frequency of face-to-face interactions. We collected data about the well-being of medical staff using the Center for Epidemiological Studies–Depression (CES-D) questionnaire and measured 4 centralities using SNA analysis. A CES-D questionnaire was administered during the study to measure the well-being of the HCPs.

Results: Overall, 247 ICU workers participated in this clinical study for 4 weeks yearly in February 2016, 2017, and 2018. The distributed leadership structure was established within the ICU in 2017 and 2018. We compared these results with those of the traditional leadership structure used in 2016. Most face-to-face interactions in the ICU were among nurses or between nurses and other professionals. In 2016, overall, 10 nurses could perform leadership tasks, which significantly increased to 24 in 2017 (P=.046) and 20 in 2018 (P=.046). Considering the increased number of nurses who could perform leadership duties and the collaboration created within the organization, SNA in 2018 showed that the betweenness (P=.001), degree (P<.001), and closeness (P<.001) centralities significantly increased compared with those in 2016. However, the eigenvector centrality significantly decreased in 2018 compared with that in 2016 (P=.01). The CES-D scores in 2018 also significantly decreased compared with those in 2016 (P=.01). The betweenness (r=0.269; P=.02), degree (r=0.262; P=.03), and eigenvector (r=0.261; P=.03) centralities and CES-D scores were positively correlated in 2016, whereas the closeness centrality and CES-D scores were negatively correlated (*r*=−0.318; *P*=.01). In 2018, the degree (*r*=−0.280; *P*=.01) and eigenvector (*r*=−0.284; *P*=.01) centralities were negatively correlated with CES-D scores.

Conclusions: Face-to-face interactions of HCPs in the ICU were measured using wearable sensors, and nurses were found to be centrally located. However, the introduction of distributed leadership created collaboration and informal leadership in the organization, altering the social network structure of HCPs and increasing organizational well-being.

Trial Registration: University Hospital Medical Information Network (UMIN) UMIN000037046; https://center6.umin.ac.jp/cgi-open-bin/icdr_e/ctr_view.cgi?recptno=R000042211

(Interact J Med Res 2023;12:e50148) doi:10.2196/50148

KEYWORDS

social network analysis; Center for Epidemiological Studies–Depression; CES-D; distributed leadership; intensive care unit; wearable sensor; face-to-face interaction

Introduction

Background

Effective communication in the intensive care unit (ICU) is critical for medical safety. However, poor communication in health care can lead to poor understanding among health care professionals, resulting in medication errors and jeopardizing patient safety. According to the Joint Commission, which is a US health care evaluation agency, <60% of the adverse events in hospitals in the United States are associated with poor communication [1]. In addition, poor communication among health care professionals can result in surgical complications and death [2,3]. Effective communication strategies, characterized by clarity, accuracy, and urgency, are becoming increasingly important in wards where patients at high risk are treated, such as ICUs and emergency departments. Therefore, the details of communication structures among health care professionals in the ICU, particularly the connections among staff, are important for patient safety [4].

Leadership structures have received attention as one of the key factors influencing the communication structure of health care professionals. Leadership in health care is generally defined as "the process by which an individual influences a group to achieve a common goal" [5]. Leadership structures can be categorized into two main categories as follows: (1) traditional leadership structures, where a single health care professional oversees the chain of command, and (2) distributed leadership structures, where multiple health care professionals form a team to perform medical tasks. Traditional leadership structure implies that a single leader makes decisions for the medical team and directs the organization or group; however, it is difficult for a single medical professional to make all important decisions in today's diversified medical care. In contrast, a distributed leadership structure comprises ≥2 people (not necessarily all members of an organization) [6]. Although these 2 structures are frequently considered incompatible, they complement each other [5]. However, a distributed leadership structure would likely work to the advantage of managing complex tasks that a single health care professional cannot handle, and it could also greatly improve the communication structure of the health care workforce [7].

Effective communication among health care professionals is beneficial for patient care, helps foster teamwork among them, and improves their working relationships. Good communication among health care professionals can significantly contribute to clinical effectiveness and job satisfaction [8], whereas poor communication can worsen relationships among medical staff, can negatively affect job satisfaction and turnover, and is a major factor in reducing the well-being of health care providers [3,4,9]. Therefore, clarifying the detailed relationship between the communication structure among medical professionals and the well-being of medical staff in the ICU, which is a stressful workplace with many patients who are critically ill, is desirable; however, comprehensively measuring face-to-face interactions among medical staff is difficult.

Objective

New wearable technologies have recently been developed to objectively and comprehensively measure face-to-face interactions [10]. Therefore, we aimed to use these techniques to analyze the impact of distributed leadership structure on the amount of face-to-face communication in a prospective observational study. We also measured the well-being of medical staff using scores from the Center for Epidemiological Studies–Depression (CES-D) questionnaire and evaluated the impact of distributed leadership structure on the well-being of medical staff.

Methods

Study Design

This was a prospective observational cohort study. The total amount of face-to-face interactions among health care professionals working in the ICU of the tertiary medical center at Mie University and the well-being of the health care professionals were prospectively measured during the observation period.

Setting

Health care professionals working in the ICU of Mie University Hospital, Japan, participated in this clinical study during each February from 2016 to 2018 (February 2016, February 2017, and February 2018). Specifically, recruitment for the first term began in mid-January 2016, with the research period extending from February 19, 2016, to March 17, 2016. For the second term, participants were recruited in mid-January 2017, and data collection was conducted between February 14, 2017, and March 13, 2017. In the third term, recruitment started in early January

2018, and the research was conducted from February 8, 2018, to March 7, 2018. The ICU is an intensive care center for patients who are critically ill with severe burns, cardiac arrest, septic shock, and trauma. Staff members in the ICU (physicians, nurses, nursing assistants, pharmacists, clinical engineers, receptionists, and secretaries) were included in this study. Participants were recruited 1 month before the beginning of the study and were required to wear wearable sociometric sensor badges (Business Microscope; Hitachi Ltd) on the chest area of their scrubs during work hours only, including breaks, for 4 weeks. These sensors were worn for a duration of 1 month each year and were collected by data analysts at the end of the study period.

Participants

In this study, all medical staff working in the ICU of Mie University Hospital were included as participants. In 2016, a total of 75 staff members were approached with a written explanation 1 month before the beginning of the study, followed by 93 in 2017 and 85 in 2018. However, each year, consent could not be obtained from 2 nurses, who were therefore excluded from the study. To ensure privacy protection and the integrity of the study, the 2 nurses who did not consent were asked to wear a dummy badge with disabled functions. Therefore, a total of 73 staff members in 2016, a total of 91 in 2017, and a total of 83 in 2018 participated in the study. From all participating health care professionals, data about face-to-face interactions and evaluations of well-being using the CES-D questionnaire were collected. There were no dropouts during the study. Thus, by preserving the completeness of the data, we reduced the potential bias owing to loss to follow-up.

Assessments

The primary outcome of this study was to collect data about face-to-face interactions among health care professionals over a 1-month period and CES-D scores measured during the same time frame. By analyzing these data, we aimed to elucidate the correlation between face-to-face interactions among health care workers and their sense of well-being. In addition, the study focused on the leadership styles of nurses working in the ICU. In 2016, a traditional leadership model was in place, whereas in 2017 and 2018, the model shifted to a distributed leadership style. This transition allowed for the measurement of changes in communication patterns among health care professionals between 2016 and the subsequent years. In this study, exposure was defined as the distributed leadership structure in 2017 and 2018, compared to the conventional leadership structure in 2016.

Data Sources

Wearable Sensors and Data Collection

Wearable sociometric sensor badges (Business Microscope; Hitachi Ltd) were used, as reported in a feasibility study of an ICU [10] and another study of a corporate call center [11]. The badge worn in the participant's front pocket automatically and comprehensively collected the data set needed for the social network analysis (SNA) of health care professionals in the ICU. In addition, the badges do not interfere with medical devices, enabling communication and safe collection of behavioral data in real ICU environments. These badges use a 3-axis microelectromechanical accelerometer to capture the physical movements of the wearer and detect personal activities. The badges can also detect face-to-face interactions using 6 infrared, data-related transceivers in front of them [11] and obtain data about who meets with whom, when, and for how long. The wearable sociometric sensor badge captures the physical movements of ICU staff using an accelerometer built into the badge. The greater the frequency with which the acceleration signal exceeds the threshold value per unit of time, the more active the staff member's body is, indicating the activity level of the staff member. Each staff member's activity level was evaluated minute by minute; they were classified as being in face-to-face communication if a predetermined threshold was exceeded. On the basis of previous studies' results, the threshold was set at 2 Hz, a level that could distinguish between active (including conversations with gestures) and quiet (such as keyboard input) actions [11]. Therefore, an active state of face-to-face communication was determined if the threshold value was >2 Hz and the activity lasted for 1 minute. Only face-to-face communications of ≥1 minute were analyzed because a threshold value <1 minute could misinterpret a simple case of staff passing each other in the ICU as active face-to-face communication.

SNA (Face-to-Face Interaction Analysis)

Overview

SNA was performed using the data set obtained from the wearable sensors, as previously described [11]. It was performed using the UCINET software (Analytic Technologies Inc). The relationship between individuals and their counterparts was evaluated using a centrality measure in the SNA. Furthermore, the following centrality measurement indicators were used.

Betweenness Centrality

The betweenness centrality represents a person's relevance in the sharing of information between people. The ratio of the number of paths that include a person to the total number of paths is known as the person's betweenness centrality when following the shortest path of combinations across all members. Specifically, a person in the approver role, such as a leader, in a medical context receives a high value [12,13].

Degree Centrality

The degree centrality indicates the number of health care professionals to which a person is connected. Specifically, a person who communicates with many health care professionals among the staff members will have a higher value of degree centrality than someone who does not communicate [12,14].

Closeness Centrality

The closeness centrality represents the inverse of the mean distance value between 2 people. Particularly, staff members in the ICU who engage in unrestricted communication with everyone without a preference will have a high value of closeness centrality [15].

Eigenvector Centrality

People associated with important individuals will have a high eigenvector centrality value. More precisely, people connected to those with high betweenness centrality values show high eigenvector centrality values. In the medical field, well-connected people can spread information quickly; therefore, those with high eigenvector centrality values are considered as important in workplaces with many patients who are critically ill, such as the ICU [16].

CES-D Measures of Well-Being

The CES-D questionnaire, which was administered during the third week of the 1-month study, captured the frequency of emotions and behaviors over the past 7 days and was rated on a 4-point scale ranging from 0 (little or none) to 3 (most or all). CES-D comprises 20 items, with a total score that ranges from 0 to 60. High scores indicate great frequency of depressive experiences [17]. Furthermore, a person had to score the lowest ("almost never") and highest ("almost never") on all negative (eg, "I felt sad") and positive (eg, "I enjoyed life") items, respectively, to score 0 on CES-D. Therefore, implying that these individuals had no depressive experiences would be misleading. Moreover, these individuals also indicated a sense of well-being [18]. Consistent with previous reports, we measured staff well-being using CES-D [19,20].

Ethical Considerations

This study was conducted in accordance with the ethical standards of the Declaration of Helsinki, and the Human Research Ethics Committee of Mie University approved its protocol (approval 2978). Written informed consent was obtained from all participants. The few staff members who refused participation wore a mock badge, with sensor functions turned off, to preserve anonymity. Furthermore, this clinical trial was registered in the University Hospital Medical Information Network (UMIN) Clinical Trial Registration System (UMIN000037046).

Sample Size

The purpose of this study was to elucidate the effects of distributed leadership. However, previous studies had not reported about the impact of distributed leadership on CES-D scores, and thus, we estimated an effect size of 0.25 [21]. With a significance level set at .05 and a power of 0.8, the total sample size required over the course of 3 years was calculated to be 159 individuals. Consequently, it was necessary to recruit 53 participants annually. Participation in this study was voluntary, and thus, adequate explanations were provided to all medical personnel working in the ICU to invite participation. Therefore, the number of participants exceeded the required sample size each year, with a total of 73 individuals in 2016, a total of 91 in 2017, and a total of 83 in 2018, thus securing a statistically appropriate sample size for the study.

Statistical Analysis

Categorical data were compared among study years using chi-square test and reported as frequencies and percentages. Quantitative comparisons of face-to-face interactions among study years, as shown in Figure 1, used the chi-square goodness-of-fit test for comparisons among the 3 groups. While analyzing the face-to-face interactions according to occupation, shown as a heat map in Figure 2, the correlation coefficient was illustrated using Spearman rank correlation coefficient. Heat map representation shows the cumulative duration of face-to-face interactions (minute × person) between specific professions. Statistical analysis was performed using the Kruskal-Wallis test while analyzing non–normally distributed data. Descriptive data were expressed as medians and IQRs. Correlation coefficients (*r*) between CES-D and SNA centrality were expressed using Spearman coefficients. The Shapiro-Wilk test, a hypothesis test that evaluates whether a data set is normally distributed, was used to evaluate normality. Statistical significance was set at *P*<.05, and all analyses were performed using IBM SPSS Statistics for Windows (version 23.0; IBM Corp).

Figure 1. Face-to-face interactions (minute×person) among health care professionals were compared annually. A chi-square test was used for categorical variables. NS: not significantly different. *Statistical significance was set at *P*<.05.

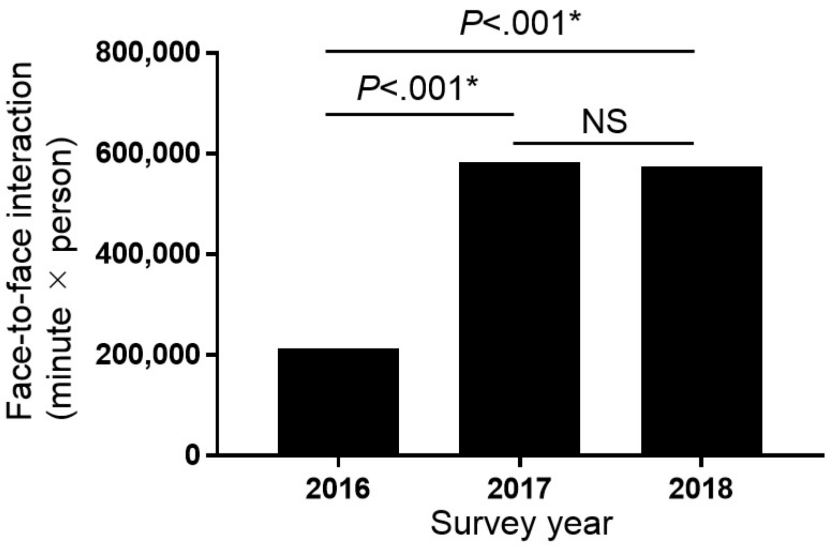
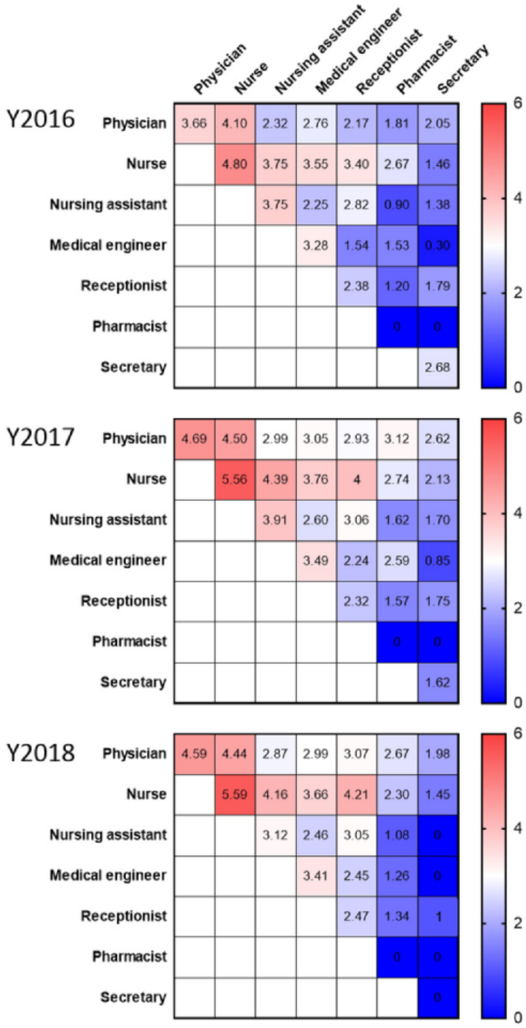


Figure 2. Heat map representation shows the cumulative duration of face-to-face interactions (minute×person) between specific professions. Data are displayed on a logarithmic scale that reflects the duration of the interactions multiplied by the number of people.



Results

Demographic Characteristics of All Health Care Providers

Overall, 247 health care professionals participated in the study across the 3 years from 2016 to 2018. Between 2016 and 2018, no significant differences were found in the number of health care professionals, such as physicians, nurses, clinical engineers, pharmacists, nursing assistants, secretaries, and receptionists (all *P*>.05; Table 1). However, the number of experienced leader nurses who could give instructions to other nurses significantly increased to 24 and 20 in 2017 and 2018, respectively, compared with 10 in 2016 (*P*=.046; Table 1). No change was found in the number of formal leaders across the 3 years in each work zone, with 1 physician and 2 nurses. However, the high number of nurses available to perform the leader's duties increased the number of informal leader nurses and nurses working cooperatively with the formal leader and other medical staff (*P*=.046).

Table 1. Demographic characteristics of all health care providers[a].

Type of occupation	Participants in 2016 (n=73), n (%)	Participants in 2017 (n=91), n (%)	Participants in 2018 (n=83), n (%)	*P* value[b]
Physician	18 (25)	20 (22)	23 (28)	.73
Nurse	38 (52)	53 (58)	45 (54)	.29
Experienced leader nurse	10 (14)	24 (26)[c]	20 (24)[c]	.046[c]
Clinical engineer	7 (10)	9 (10)	9 (11)	.85
Pharmacist	1 (1)	1 (1)	1 (1)	.99
Nursing assistant	5 (7)	4 (4)	2 (2)	.53
Secretary	2 (3)	2 (2)	2 (2)	.99
Receptionist	2 (3)	2 (2)	2 (2)	.99

[a]Chi-square test was used across the 3 years.

[b]Overall *P* value=.37.

[c]$P<.05$ represents statistical significance.

Total Frequency of Face-to-Face Interactions

All ICU staff worked an average of 160 hours during the 4 weeks of data collection. Overall, 212,872; 583,876; and 573,586 person-minute face-to-face interactions were obtained from a total of 73 staff members in 2016, a total of 91 in 2017, and a total of 83 in 2018, respectively. Face-to-face interactions significantly increased in 2017 and 2018 compared with those in 2016 ($P<.001$; Figure 1).

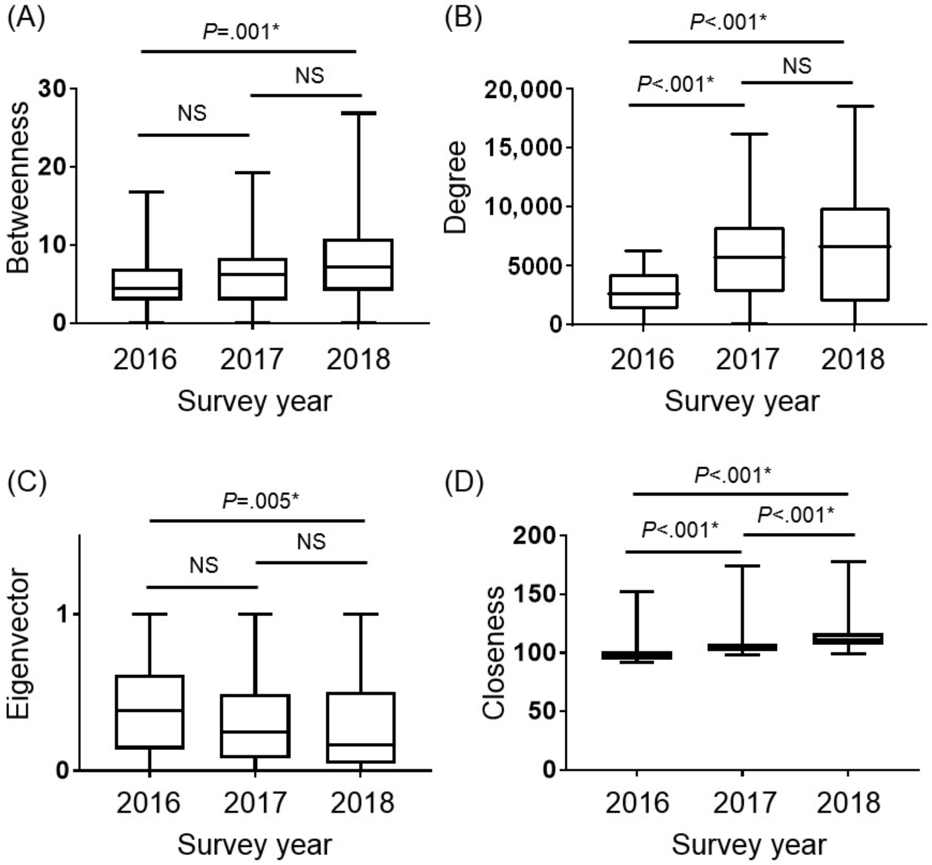
Pair Preferences for Face-to-Face Interactions Among ICU Staff Members

Next, we measured the frequency of face-to-face interactions according to occupation and examined which interprofessional communication dominated the frequency of face-to-face interactions in the ICU (Figure 2). First, we compared the total duration of face-to-face interactions per month according to occupation and by year. The duration of face-to-face interactions involving either physicians or nurses was longer than that of other occupations across the 3 years (2016-2018; Figure 2). In addition, across the 3 years, nurses had the highest frequency of face-to-face interactions with other nurses, and their duration of face-to-face interactions was longer than that among other professionals. Nurses also interacted actively and frequently with other professionals (Figure 2). Physicians communicated most frequently with nurses in 2016 during the 3-year measurement, whereas the communication frequency among physicians increased in 2017 and 2018. Therefore, these data indicate that nurses play a central role in face-to-face interactions in the ICU.

Comparison of SNA Centrality Across the 3 Years

We conducted a 3-year SNA of health care professionals to measure the centrality of SNA. Specifically, we measured 4 centralities as follows: the betweenness, degree, eigenvector, and closeness centralities. The high betweenness centrality for approvers in an organization (ie, leader physicians and nurses) significantly increased in 2018 compared with that in 2016 ($P=.001$; Figure 3A). Notably, this could be attributed to a significant increase in nurses' experience as leader nurses since 2017 (Table 1), the emergence of informal leaders to assist leaders in the organization, and more staff members communicating collaboratively with other staff members. The degree centrality, which represents the connection with many health care professionals, significantly increased in 2017 and 2018 compared with that in 2016 ($P<.001$; Figure 3B), and this could be because of the facilitated communication among staff members. Furthermore, the eigenvector centrality, which is high among staff members connected to many health care professionals, significantly decreased in 2018 compared with that in 2016 ($P=.005$; Figure 3C), indicating that more individuals played a role in rapidly disseminating information in the ICU in 2016 than in 2018. Specifically, information was transmitted by few leaders but received by many staff members, suggesting that leadership duties were undistributed. More interestingly, the closeness centrality, which is high for staff members who can communicate with everyone without any preference, significantly increased in 2017 and 2018 compared with that in 2016 ($P<.001$; Figure 3D). This result strongly supports the notion that informal leaders were created to assist those in the organization after 2017, when a distributed leadership structure was adopted, and that more staff members communicated collaboratively with others.

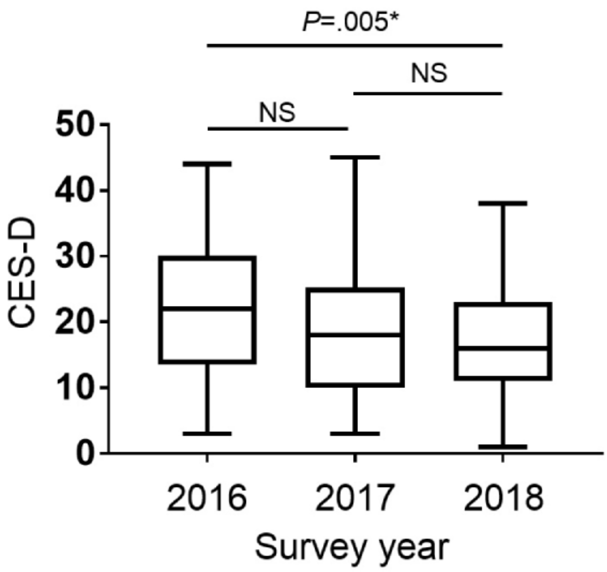
Figure 3. The face-to-face interaction data obtained from the wearable sensors were used to analyze the centrality of health care professionals and were compared with 3 years of data. Overall, 4 types of centralities are illustrated: (A) betweenness, (B) degree, (C) eigenvector, and (D) closeness centralities. Descriptive data are summarized as medians and IQRs. NS: not significantly different. *Statistical significance was set at $P<.05$.



Changes in Well-Being Over the 3 Consecutive Years

All participants (247/247, 100%) were asked to complete the CES-D questionnaire during the third week of the study to measure their well-being (Figure 4). High CES-D scores were positively correlated with the degree of depression, whereas the lower the CES-D scores, the happier the health care provider [19,20]. The median CES-D scores were 22 (IQR 14-30), 17 (IQR 10-25), and 16 (IQR 11-23) in 2016, 2017, and 2018, respectively. CES-D scores were significantly decreased in 2018 compared with those in 2016, suggesting an increase in the well-being of health care professionals in 2018 ($P=.005$). Generally, a score ≥16 indicates depression; therefore, these results show that although health care professionals were exposed to excessive stress, the situation gradually improved over the 3 consecutive years [17].

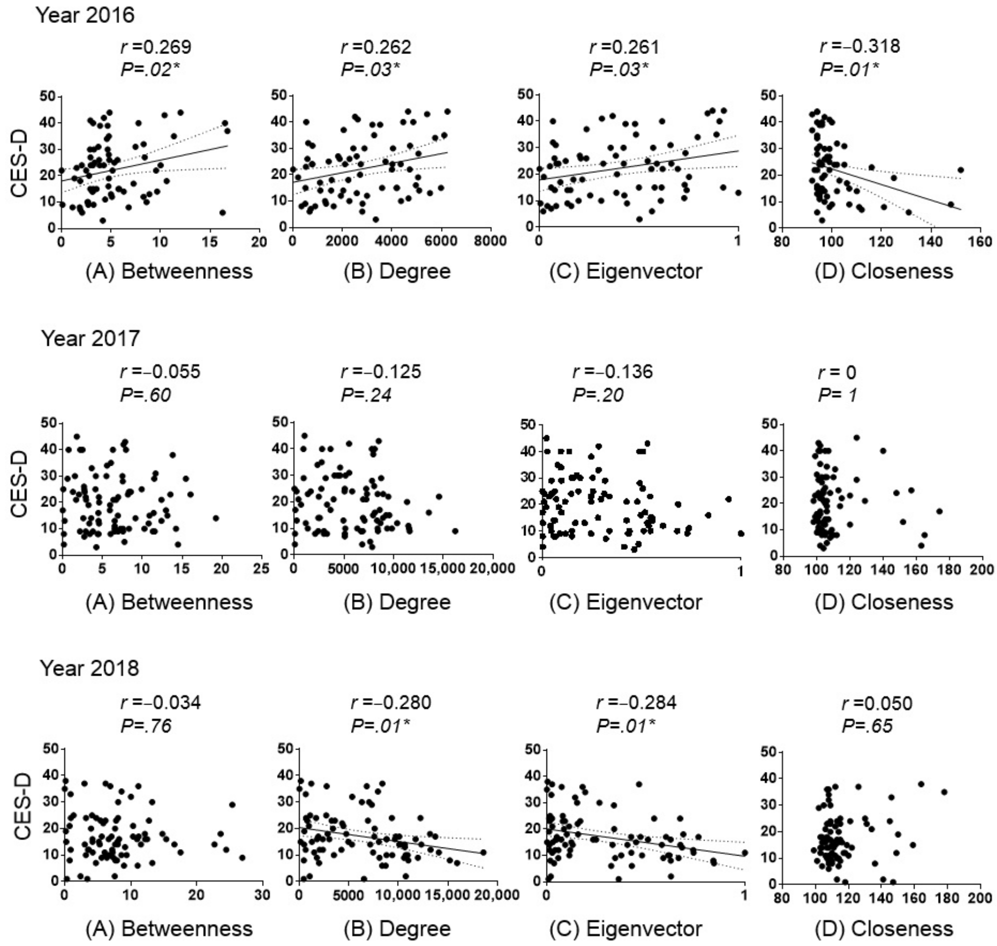
Figure 4. The Center for Epidemiological Studies–Depression (CES-D) questionnaire was administered during the third week of the 1-month study. Descriptive data are summarized as medians and IQRs. NS: not significantly different. *Statistical significance was set at $P<.05$.



Correlations Between the Well-Being of Health Care Professionals and SNA Centrality

Next, we analyzed the correlation between the well-being of all staff members and the centrality of the SNA (Figure 5). In 2016, the betweenness centrality increased and was positively correlated with health care professionals with high CES-D values (ie, those with low happiness and high depression levels; *r*=0.269; *P*=.02). The CES-D scores were also positively correlated with health care professionals with high degree centrality values (*r*=0.262; *P*=.03) and positively correlated with the eigenvector centrality (*r*=0.261; *P*=.03). These results strongly suggest that the leadership duties of some staff members placed excessive psychological burden and workload on others in 2016. Interestingly, the CES-D scores were also negatively correlated with the closeness centrality in 2016 (*r*=−0.318; *P*=.006). In addition, this result shows that people who could communicate with everyone in the organization without any preference had high levels of well-being, suggesting that effective communication positively influenced the well-being of staff members. Although no correlation was found between SNA centrality and CES-D scores in 2017 (Figure 5), the degree (*r*=−0.280; *P*=.01) and eigenvector (*r*=−0.284; *P*=.009) centralities were negatively correlated with CES-D scores in 2018. Moreover, opposite results were obtained in 2016 and 2018. Therefore, this result shows that staff members connected to many health care professionals have high level of well-being.

Figure 5. The correlation between the Center for Epidemiological Studies–Depression (CES-D) scores and the 4 centralities, which include (A) betweenness, (B) degree, (C) eigenvector, and (D) closeness centralities, is illustrated yearly. The correlation coefficients (r) between the CES-D scores and social network analysis centrality are expressed using Spearman coefficients. *Statistical significance was set at *P*<.05.



Demographic Characteristics of Health Care Providers Over the 3 Consecutive Years

We further conducted a subgroup analysis to ascertain whether similar results would be found among the staff members who participated in this study over the 3 consecutive years. Overall, 42 health care professionals participated in the study over the 3 consecutive years, of whom 7 (17%) were physicians, 26 (62%) were nurses, 4 (10%) were clinical engineers, 1 (2%) was a pharmacist, 2 (5%) were nursing assistants, and 2 (5%) were receptionists (Table 2). Among the 42 health care professionals, 6 (14%) nurses were available for leadership duties in 2016, significantly increasing to 19 (45%) in 2017 and 20 (48%) and 2018 (*P*=.02; Table 2).

Table 2. Demographic characteristics of participants over the 3 consecutive years[a].

Type of occupation	Participants in 2016 (n=42), n (%)	Participants in 2017 (n=42), n (%)	Participants in 2018 (n=42), n (%)	*P* value[b]
Physician	7 (17)	7 (17)	7 (17)	.99
Nurse	26 (62)	26 (62)	26 (62)	.99
Experienced leader nurse	6 (14)	19 (45)[c]	20 (48)[c]	.02[c]
Clinical engineer	4 (10)	4 (10)	4 (10)	.99
Pharmacist	1 (2)	1 (2)	1 (2)	.99
Nursing assistant	2 (5)	2 (5)	2 (5)	.99
Secretary	0 (0)	0 (0)	0 (0)	.99
Receptionist	2 (5)	2 (5)	2 (5)	.99

[a]Chi-square test was used across the 3 years.

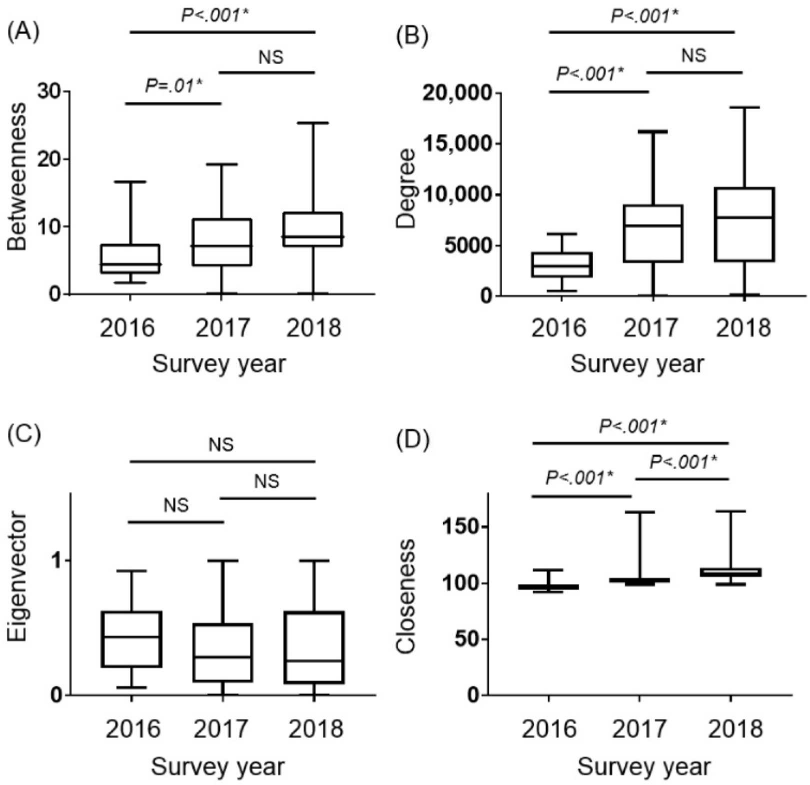
[b]Overall *P* value=.99.

[c]*P*<.05 represents statistical significance.

Comparison of SNA Centrality Among Health Care Professionals Who Participated in the Study Over the 3 Consecutive Years

In the subgroup analysis, the betweenness centrality was significantly increased in 2017 (*P*=.01) and 2018 (*P*<.001) compared with that in 2016 for the 42 health care professionals who participated in the study over the 3 consecutive years (Figure 6A). The degree centrality was also significantly increased in 2017 and 2018 compared with that in 2016 (*P*<.001; Figure 6B), whereas the eigenvector centrality did not differ across the 3 years of measurement (Figure 6C). The closeness centrality was significantly increased in 2017 and 2018 compared with that in 2016 (*P*<.001) and significantly increased in 2017 compared with that in 2018 (*P*<.001; Figure 6D). Subgroup analyses of the 42 health care professionals who participated in the study over the 3 consecutive years also showed increases in the betweenness, degree, and closeness centralities, similar to the analysis results of all health care professionals.

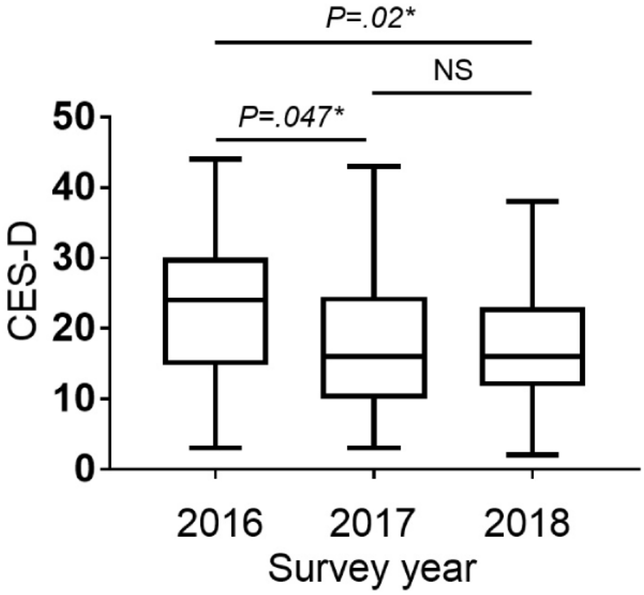
Figure 6. The face-to-face interaction data obtained from the wearable sensors were used to analyze the centrality of health care professionals and were compared among the 3 years of data. Overall, 4 types of centralities are illustrated as follows: (A) betweenness, (B) degree, (C) eigenvector, and (D) closeness centralities. Descriptive data are summarized as medians and IQRs. NS: not significantly different. *Statistical significance was set at *P*<.05.



Trends in the Well-Being of Health Care Professionals Who Participated in the Study Over the 3 Consecutive Years

The median CES-D scores for the 42 health care professionals who participated in the study over the 3 consecutive years were 24 (IQR 15-30), 16 (IQR 10-24), and 16 (IQR 12-23) in 2016, 2017, and 2018, respectively. Furthermore, the CES-D scores decreased in 2017 (*P*=.047) and 2018 (*P*=.02) compared with that in 2016 (Figure 7), whereas well-being significantly increased in the subgroup analyses.

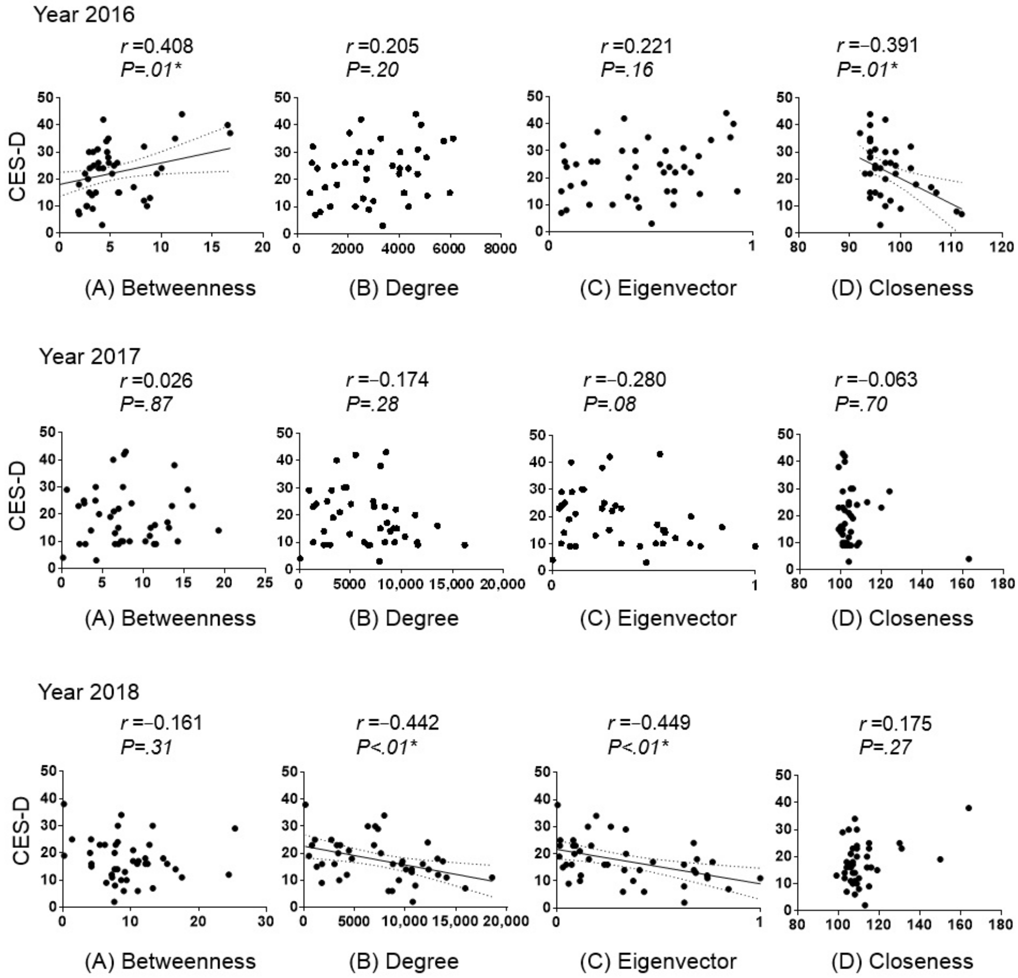
Figure 7. The Center for Epidemiological Studies–Depression (CES-D) questionnaire was administered during the third week of the 1-month study. Descriptive data are summarized as medians and IQRs. NS: not significantly different. *Statistical significance was set at *P*<.05.



Correlations Between Well-Being and SNA Centrality for Health Care Professionals Who Participated Over the 3 Consecutive Years

Next, we analyzed the correlation between well-being and SNA centrality for the 42 staff members who participated over the 3 consecutive years (Figure 8). In 2016, the betweenness centrality was positively correlated with CES-D scores (*r*=0.408; *P*=.007), whereas the closeness centrality was negatively correlated with CES-D scores (*r*=−0.391; *P*=.01). No correlation was found between CES-D scores and SNA centrality in 2017. In 2018, CES-D scores were negatively correlated with the eigenvector (*r*=−0.449; *P*=.003) and degree (*r*=−0.442; *P*=.003) centralities. Therefore, this result indicates that staff members who are connected to many health care professionals have high levels of well-being. In addition, staff members connected to health care professionals with high values of betweenness centrality (ie, health care professionals with high eigenvector centrality), such as leader nurses, showed high levels of well-being.

Figure 8. The correlation between the Center for Epidemiological Studies–Depression (CES-D) scores and the 4 centralities, which include (A) betweenness, (B) degree, (C) eigenvector, and (D) closeness centralities, is illustrated yearly. The correlation coefficients (r) between the CES-D scores and social network analysis centrality are expressed using Spearman coefficients. *Statistical significance was set at *P*<.05.



Discussion

Principal Findings

In this study, we found that nurses play a central role in face-to-face interactions in the ICU (Figures 1 and 2). Furthermore, since 2017, a distributed leadership structure has been implemented in the ICU, with the number of approvers (ie, leader physicians and nurses) and informal leaders in the organization significantly increasing in 2018 compared with that in 2016. Therefore, communication among staff members was facilitated, and communication between many health care professionals and others was improved. Interestingly, the activation of communication in the ICU also increased the well-being of health care professionals, particularly the well-being of medical staff, which increased with those who extensively communicated with the surrounding staff.

Distributed leadership was introduced to the ICU nursing staff members during the second and subsequent years of the study, dramatically altering the centrality of the communication structure among medical staff. The betweenness, degree, and closeness centralities increased after 2017 compared with those in 2016. This result supports the idea that a distributed leadership structure creates informal leaders who collaborate and facilitate interprofessional communication in the organization (Figure 3). Interestingly, the level of depression among ICU staff members improved as cooperation increased in the organization (Figure 4). Owing to the high number of patients who are critically ill in the ICU, a few staff members perform leadership duties, which decreases the well-being of some staff members and the organization's overall performance. Therefore, introducing a distributed leadership structure to overcome this situation and create effective collaboration could improve the ICU work environment. Similar results were shown in a study that conducted an SNA with similar wearable technology for employees at a company outside the health care industry [22]. In a previous study, the betweenness centrality in the SNA was highly negatively correlated with satisfaction with group interactions [22]. Therefore, overloading employees, such as with leadership duties, may reduce the well-being of some individuals, consequently reducing the satisfaction and well-being of the organization.

The comprehensive measurement of face-to-face interactions among health care professionals requires using new technologies rather than traditional, survey-like methods. Therefore, we addressed this issue using a wearable social measurement sensor badge (Hitachi Ltd) to measure face-to-face interactions among health care professionals in the ICU. Furthermore, SNA using the data set obtained from the wearable sensor revealed the structure of interprofessional interactions in the ICU and showed that nurses play a pivotal role in the interaction network (Figures 1 and 2). Similar sociometric wearable sensor technologies have also been used to quantify the dynamics of health care professionals' behavior in complex and unpredictable health care environments. For example, it has been used in studies of the interactions among health care professionals, patients, and caregivers in a general children's hospital ward in Italy [23]; among nurses in a surgical ICU at a university hospital in the United States [24]; and in the emergency department of a university hospital in the United States [25]. Therefore, the ability to monitor face-to-face interactions in health care settings using wearable sensors has been supported by several independent studies, including ours [10].

SNA is a research approach in the social sciences that focuses on analyzing interactions between individuals and social groups to explain social patterns of behavior and interest [26,27]. SNA studies in health care have examined various topics, such as patient satisfaction with the social support received by patients with chronic pain [28], studies of professional network structure [29], and health care provider job satisfaction and leadership; it has also been used to demonstrate various outcomes [30,31]. However, SNA remains underused in understanding professional communication and performance among health care providers owing to the difficulty in accurately measuring communication among them [32]. We accurately measured face-to-face interactions using a wearable sensor in this study, which allowed us to use SNA to prospectively analyze accurate face-to-face interactions among health care professionals according to occupation and over 3 consecutive years.

Distributed leadership or collaborative leadership is a novel theory in this field. Supporting and building relationships with health care professionals increase the likelihood that they will be positively influenced and motivated to work toward their goals. Furthermore, the emergence of supportive leaders who empathize on a personal level enables staff to be happy and more satisfied with their work [5]. The distributed leadership introduced in this study may have developed cooperativeness and enabled the organization to support informal leaders in communicating information to colleagues and related organizations, thus enabling all health care professionals to make well-informed decisions. In addition, informal leaders participated in the leadership process and were actively involved in identifying and communicating the needs of health care professionals, which may have altered the communication structure of health care professionals, reduced the psychological and physical burden on formal leaders, and increased the well-being of organizations.

Improved teamwork and communication are among the most important factors for health care professionals to improve their clinical effectiveness and job satisfaction [8]. Therefore, activating communication among health care professionals may improve patient life outcomes in the ICU [33-35]. As we found in this study, the introduction of distributed leadership may improve patient life outcomes by changing the structure of SNA among ICU staff members and increasing their well-being. However, further prospective clinical studies are needed to confirm this hypothesis.

In this study, the CES-D scale was used to measure well-being. Unfortunately, the CES-D scale does not measure well-being comprehensively, as it assesses the continuum from happiness to depression [20]. However, the CES-D scale has established reliability and validity in a wide range of research for the assessment of the presence and severity of depressive symptoms, and it is deemed suitable for capturing the negative aspects of mental health among health care workers, which is why it was adopted for this study. However, for a more extensive evaluation of well-being, it would be desirable to use additional measures that assess subjective happiness in conjunction with the CES-D scale. This approach would enable the assessment of not only the presence of depression but also positive aspects such as sense of happiness and quality of life. Future studies should aim for a more holistic understanding of well-being by using composite measures that multifacetedly evaluate psychological well-being [36].

The measurement of face-to-face interactions using wearable sensors revealed that nurses were central to ICU communication. Furthermore, in stressful workplaces, such as the ICU, the well-being of medical staff members in leadership positions is reduced owing to excessive workload and psychological burden.

However, distributed leadership can increase organizational well-being by changing the communication structure of health care professionals in the ICU and creating a cooperative environment in the workplace.

Comparison With Previous Studies

One of the impediments to the implementation of SNA in the field of health care research is the lack of effective techniques for objective and comprehensive data acquisition [31]. Currently, data sets for SNA are commonly obtained through manual acquisition of questionnaires, observations, and electronic medical and administrative records [37,38]. However, objectively, comprehensively, and continuously measuring social network connections among health care professionals with such manual data acquisition is difficult. Therefore, in this study, we overcame these difficulties by using wearable sensors to objectively and comprehensively collect data about the duration of face-to-face communication among health care professionals in the ICU. To the best of our knowledge, this is the first study to reveal the impact of distributed leadership on the centrality of health care worker communication networks and the well-being of ICU health care professionals. Therefore, comparing our findings with those of previous studies is not practical, given the novelty of our approach.

Limitations

This study had some limitations. First, this was an observational study; therefore, it could reveal a causal relationship between a distributed leadership structure and the well-being of ICU staff members. However, as nurses are central to the ICU communication structure, it is reasonable to assume that changes in the leadership structure of nurses and their colleagues would influence the level of organizational well-being. Furthermore, a previous study showed that the key elements for effective physician-nurse teamwork are the quality of collaboration, coordination, shared mental models, communication structure, and leadership structure [39]. Leadership structure largely influences ICU communication structure and teamwork [39]. Therefore, an improved ICU work environment resulting from implementing a distributed leadership structure can improve the well-being of health care professionals in the ICU. For example, implementing distributed leadership may effectively prevent burnout, which many ICU staff experience.

Second, this study did not measure communication other than face-to-face interactions, as measured using wearable badges. For example, digital web-based information exchange and communication via email and mobile phones, among others, which have rapidly become mainstream in today's society, were not included in the information obtained from wearable sensors. However, the most important human interactions in the ICU still occur offline, and face-to-face interactions are essential for building organizational trust [40]. Therefore, this study focused on offline, face-to-face interactions.

Conclusions

Our study revealed a relationship between the ICU communication structure and the well-being of health care professionals. Specifically, the well-being of health care professionals with high degree centrality values, who communicate better with other health care professionals, was increased, suggesting that the promotion of interdisciplinary medical care can stimulate communication within the ICU and increase the level of well-being in the workplace. In addition, the level of well-being of health care professionals with high betweenness centrality, such as leader physicians and nurses, was low. These results suggest that in stressful workplaces, including ICUs, the excessive emotional burden placed on some leader physicians and nurses weighs on them, resulting in poor organizational performance. However, a possible solution to this problem is introducing a distributed leadership structure, where leadership duties can be shared among multiple staff members to reduce the burden on 1 person.

In summary, this study shows that increasing the number of experienced leader nurses and promoting a distributed leadership structure can enhance the well-being of health care professionals in ICUs.

Acknowledgments

This study was supported by the Japan Society for the Promotion of Science (grants JP 22K06910, 22K09117, 22K09138, 19KK0224, 18K08916, and 21K09015).

Data Availability

The data sets analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

RE, AI-M, EK, and MS were involved in the conception and design of the study. RE, EK, MI, and TH were involved in data acquisition. EK, MI, TS, AH, FO, and EJP contributed to data analysis and interpretation. RE, AI-M, EK, RK, and MS were involved in the drafting of the manuscript.

Conflicts of Interest

None declared.

References

1. Müller M, Jürgens J, Redaèlli M, Klingberg K, Hautz WE, Stock S. Impact of the communication and patient hand-off tool SBAR on patient safety: a systematic review. BMJ Open 2018 Aug 23;8(8):e022202 [FREE Full text] [doi: 10.1136/bmjopen-2018-022202] [Medline: 30139905]
2. Mazzocco K, Petitti DB, Fong KT, Bonacum D, Brookey J, Graham S, et al. Surgical team behaviors and patient outcomes. Am J Surg 2009 May;197(5):678-685. [doi: 10.1016/j.amjsurg.2008.03.002] [Medline: 18789425]
3. Rosenstein AH, O'Daniel M. Impact and implications of disruptive behavior in the perioperative arena. J Am Coll Surg 2006 Jul;203(1):96-105. [doi: 10.1016/j.jamcollsurg.2006.03.027] [Medline: 16798492]
4. Rosenstein AH, O'Daniel M. Disruptive behavior and clinical outcomes: perceptions of nurses and physicians. Am J Nurs 2005 Jan;105(1):54-64; quiz 64. [doi: 10.1097/00000446-200501000-00025] [Medline: 15659998]
5. Al-Sawai A. Leadership of healthcare professionals: where do we stand? Oman Med J 2013 Jul;28(4):285-287 [FREE Full text] [doi: 10.5001/omj.2013.79] [Medline: 23904925]
6. Barattucci M, Lo Presti A, Bufalino G, Jønsson T, Teresi M, Pagliaro S. Distributed leadership agency and work outcomes: validation of the Italian DLA and its relations with commitment, trust, and satisfaction. Front Psychol 2020 Mar 31;11:512 [FREE Full text] [doi: 10.3389/fpsyg.2020.00512] [Medline: 32296370]
7. Currie G, Gulati K, Sohal A, Spyridonidis D, Busari JO. Distributing systems level leadership to address the COVID-19 pandemic. BMJ Lead 2022 Mar;6(1):39-44. [doi: 10.1136/leader-2020-000280] [Medline: 35537018]

8. Flin R, Fletcher G, McGeorge P, Sutherland A, Patey R. Anaesthetists' attitudes to teamwork and safety. Anaesthesia 2003 Mar;58(3):233-242 [FREE Full text] [doi: 10.1046/j.1365-2044.2003.03039.x] [Medline: 12603453]
9. Rosenstein AH, O'Daniel M. A survey of the impact of disruptive behaviors and communication defects on patient safety. Jt Comm J Qual Patient Saf 2008 Aug;34(8):464-471. [doi: 10.1016/s1553-7250(08)34058-6] [Medline: 18714748]
10. Ito-Masui A, Kawamoto E, Nagai Y, Takagi Y, Ito M, Mizutani N, et al. Feasibility of measuring face-to-face interactions among ICU healthcare professionals using wearable sociometric badges. Am J Respir Crit Care Med 2020 Jan 15;201(2):245-247. [doi: 10.1164/rccm.201904-0779LE] [Medline: 31545651]
11. Watanabe JI, Ishibashi N, Yano K. Exploring relationship between face-to-face interaction and team performance using wearable sensor badges. PLoS One 2014 Dec 15;9(12):e114681 [FREE Full text] [doi: 10.1371/journal.pone.0114681] [Medline: 25501748]
12. Blanchet K, James P. How to do (or not to do) ... a social network analysis in health systems research. Health Policy Plan 2012 Aug;27(5):438-446. [doi: 10.1093/heapol/czr055] [Medline: 21840934]
13. Freeman LC. Centrality in social networks conceptual clarification. Soc Netw 1978 Jan;1(3):215-239. [doi: 10.1016/0378-8733(78)90021-7]
14. Valente TW. Social Networks and Health: Models, Methods, and Applications. New York, NY: Oxford University Press; Mar 25, 2010.
15. Zhang J, Luo Y. Degree centrality, betweenness centrality, and closeness centrality in social network. 2017 Presented at: 2nd International Conference on Modelling, Simulation and Applied Mathematics (MSAM2017); March 26-27, 2017; Bangkok, Thailand. [doi: 10.2991/msam-17.2017.68]
16. Effken JA, Carley KM, Gephart S, Verran JA, Bianchi D, Reminga J, et al. Using ORA to explore the relationship of nursing unit communication to patient safety and quality outcomes. Int J Med Inform 2011 Jul;80(7):507-517 [FREE Full text] [doi: 10.1016/j.ijmedinf.2011.03.015] [Medline: 21536492]
17. Vilagut G, Forero CG, Barbaglia G, Alonso J. Screening for depression in the general population with the center for epidemiologic studies depression (CES-D): a systematic review with meta-analysis. PLoS One 2016 May 16;11(5):e0155431 [FREE Full text] [doi: 10.1371/journal.pone.0155431] [Medline: 27182821]
18. Joseph S, Wood A. Assessment of positive functioning in clinical psychology: theoretical and practical issues. Clin Psychol Rev 2010 Nov;30(7):830-838. [doi: 10.1016/j.cpr.2010.01.002] [Medline: 20137841]
19. Wood AM, Taylor PJ, Joseph S. Does the CES-D measure a continuum from depression to happiness? comparing substantive and artifactual models. Psychiatry Res 2010 May 15;177(1-2):120-123. [doi: 10.1016/j.psychres.2010.02.003] [Medline: 20207424]
20. Siddaway AP, Wood AM, Taylor PJ. The center for epidemiologic studies-depression (CES-D) scale measures a continuum from well-being to depression: testing two key predictions of positive clinical psychology. J Affect Disord 2017 Apr 15;213:180-186 [FREE Full text] [doi: 10.1016/j.jad.2017.02.015] [Medline: 28254608]
21. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Cambridge, MA: Academic Press; 1969.
22. Olguin Olguin D, Waber BN, Kim T, Mohan A, Ara K, Pentland A. Sensible organizations: technology and methodology for automatically measuring organizational behavior. IEEE Trans Syst Man Cybern B Cybern 2009 Feb;39(1):43-55. [doi: 10.1109/TSMCB.2008.2006638] [Medline: 19150759]
23. Isella L, Romano M, Barrat A, Cattuto C, Colizza V, Van den Broeck W, et al. Close encounters in a pediatric ward: measuring face-to-face proximity and mixing patterns with wearable sensors. PLoS One 2011 Feb 28;6(2):e17144 [FREE Full text] [doi: 10.1371/journal.pone.0017144] [Medline: 21386902]
24. Rosen MA, Dietz AS, Lee N, Wang IJ, Markowitz J, Wyskiel RM, et al. Sensor-based measurement of critical care nursing workload: unobtrusive measures of nursing activity complement traditional task and patient level indicators of workload to predict perceived exertion. PLoS One 2018 Oct 12;13(10):e0204819 [FREE Full text] [doi: 10.1371/journal.pone.0204819] [Medline: 30312326]
25. Yu D, Blocker RC, Sir MY, Hallbeck MS, Hellmich TR, Cohen T, et al. Intelligent emergency department: validation of sociometers to study workload. J Med Syst 2016 Mar;40(3):53. [doi: 10.1007/s10916-015-0405-1] [Medline: 26645317]
26. Wasserman S, Faust K. Social Network Analysis: Methods and Applications. Cambridge, UK: Cambridge University Press; 1994.
27. Merrill J, Hripcsak G. Using social network analysis within a department of biomedical informatics to induce a discussion of academic communities of practice. J Am Med Inform Assoc 2008;15(6):780-782 [FREE Full text] [doi: 10.1197/jamia.M2717] [Medline: 18756000]
28. Fernández-Peña R, Molina JL, Valero O. Satisfaction with social support received from social relationships in cases of chronic pain: the influence of personal network characteristics in terms of structure, composition and functional content. Int J Environ Res Public Health 2020 Apr 15;17(8):2706 [FREE Full text] [doi: 10.3390/ijerph17082706] [Medline: 32326411]
29. Cunningham FC, Ranmuthugala G, Plumb J, Georgiou A, Westbrook JI, Braithwaite J. Health professional networks as a vector for improving healthcare quality and safety: a systematic review. BMJ Qual Saf 2012 Mar;21(3):239-249 [FREE Full text] [doi: 10.1136/bmjqs-2011-000187] [Medline: 22129933]

30. Tasselli S. Social networks of professionals in health care organizations: a review. Med Care Res Rev 2014 Dec;71(6):619-660. [doi: 10.1177/1077558714557079] [Medline: 25380607]
31. Chambers D, Wilson P, Thompson C, Harden M. Social network analysis in healthcare settings: a systematic scoping review. PLoS One 2012;7(8):e41911 [FREE Full text] [doi: 10.1371/journal.pone.0041911] [Medline: 22870261]
32. Sabot K, Wickremasinghe D, Blanchet K, Avan B, Schellenberg J. Use of social network analysis methods to study professional advice and performance among healthcare providers: a systematic review. Syst Rev 2017 Oct 23;6(1):208 [FREE Full text] [doi: 10.1186/s13643-017-0597-1] [Medline: 29058638]
33. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. An evaluation of outcome from intensive care in major medical centers. Ann Intern Med 1986 Mar;104(3):410-418. [doi: 10.7326/0003-4819-104-3-410] [Medline: 3946981]
34. Zimmerman JE, Shortell SM, Rousseau DM, Duffy J, Gillies RR, Knaus WA, et al. Improving intensive care: observations based on organizational case studies in nine intensive care units: a prospective, multicenter study. Crit Care Med 1993 Oct;21(10):1443-1451. [Medline: 8403951]
35. Shortell SM, Zimmerman JE, Rousseau DM, Gillies RR, Wagner DP, Draper EA, et al. The performance of intensive care units: does good management make a difference? Med Care 1994 May;32(5):508-525. [doi: 10.1097/00005650-199405000-00009] [Medline: 8182978]
36. Lara-Cabrera ML, Betancort M, Muñoz-Rubilar A, Rodríguez-Novo N, Bjerkeset O, De Las Cuevas C. Psychometric properties of the WHO-5 well-being index among nurses during the COVID-19 pandemic: a cross-sectional study in three countries. Int J Environ Res Public Health 2022 Aug 16;19(16):10106 [FREE Full text] [doi: 10.3390/ijerph191610106] [Medline: 36011741]
37. Reeves S, Pelone F, Harrison R, Goldman J, Zwarenstein M. Interprofessional collaboration to improve professional practice and healthcare outcomes. Cochrane Database Syst Rev 2017 Jun 22;6(6):CD000072 [FREE Full text] [doi: 10.1002/14651858.CD000072.pub3] [Medline: 28639262]
38. Walters SJ, Stern C, Robertson-Malt S. The measurement of collaboration within healthcare settings: a systematic review of measurement properties of instruments. JBI Database System Rev Implement Rep 2016 Apr;14(4):138-197. [doi: 10.11124/JBISRIR-2016-2159] [Medline: 27532315]
39. Manser T. Teamwork and patient safety in dynamic domains of healthcare: a review of the literature. Acta Anaesthesiol Scand 2009 Feb;53(2):143-151. [doi: 10.1111/j.1399-6576.2008.01717.x] [Medline: 19032571]
40. Csikszentmihalyi M. Creativity: Flow and the Psychology of Discovery and Invention. New York, NY: Harper Perennial; 1997.

Abbreviations

CES-D: Center for Epidemiological Studies–Depression
HCP: health care professional
ICU: intensive care unit
SNA: social network analysis
UMIN: University Hospital Medical Information Network

Edited by T de Azevedo Cardoso; submitted 21.06.23; peer-reviewed by M Ullah; comments to author 21.07.23; revised version received 28.07.23; accepted 07.11.23; published 29.11.23.

Please cite as:
Esumi R, Ito-Masui A, Kawamoto E, Ito M, Hayashi T, Shinkai T, Hane A, Okuno F, Park EJ, Kaku R, Shimaoka M
Correlation Between the Social Network Structure and Well-Being of Health Care Workers in Intensive Care Units: Prospective Observational Study
Interact J Med Res 2023;12:e50148
URL: https://www.i-jmr.org/2023/1/e50148
doi:10.2196/50148
PMID:37935050

Original Paper

The Relation Between Students' Theoretical Knowledge and Practical Skills in Endodontics: Retrospective Analysis

Franziska Haupt[1], Dr med dent; Philipp Kanzow[1], MSc, Dr rer medic, PD Dr med dent

[1]Department of Preventive Dentistry, Periodontology and Cariology, University Medical Center Göttingen, Göttingen, Germany

Corresponding Author:
Philipp Kanzow, MSc, Dr rer medic, PD Dr med dent
Department of Preventive Dentistry, Periodontology and Cariology
University Medical Center Göttingen
Robert-Koch-Str 40
Göttingen, 37075
Germany
Phone: 49 551 3960870
Fax: 49 551 3960869
Email: philipp.kanzow@med.uni-goettingen.de

Abstract

Background: Dental undergraduate students are required to show sufficient practical skills prior to treating patients. Practical skills and the underlying theoretical knowledge are taught in preclinical courses. Usually, the learning outcome is assessed in written multiple-choice examinations (theoretical knowledge) and practical skills tests. However, students' assessment of practical skills is more time consuming and prone to bias than objective multiple-choice examinations.

Objective: This study aims to analyze the relation between students' theoretical knowledge and practical skills in endodontics. Furthermore, the predictive validity of a theoretical knowledge assessment on students' practical skills was assessed.

Methods: Examination results from all students who participated in the preclinical phantom course in Operative Dentistry (sixth semester of the undergraduate dental curriculum in Germany) between the 2015 summer term and the 2022 summer term were retrospectively evaluated (N=447). The effects of age, sex, previous course participation, and theoretical knowledge on students' practical skills were assessed, using Pearson correlations, Wilcoxon rank sum tests, and a linear regression analysis. Subsequently, students' theoretical knowledge and practical skills were compared via a Fisher exact test to identify a suitable pass mark for students' theoretical knowledge that was associated with sufficient practical skills (≥60%).

Results: Students' theoretical knowledge was significantly associated with practical skills ($P_{adjusted}$=.02; *r*=0.13). By using the current pass mark for theoretical knowledge (ie, 60%), a significant differentiation between insufficient practical skills (<60%) and sufficient practical skills (≥60%) was achieved (*P*=.02). However, for the discrimination between students with sufficient practical skills and students with insufficient practical skills, an adapted pass mark for theoretical knowledge would be more appropriate. The ideal pass mark amounted to 58% (*P*=.02).

Conclusions: Students' practical skills and theoretical knowledge are significantly correlated. By objectively measuring students' theoretical knowledge, a rough estimation of students' practical skills (ie, a differentiation between sufficient and insufficient practical skills) is possible.

(Interact J Med Res 2023;12:e46305) doi:10.2196/46305

KEYWORDS

curricula; curriculum; dental; dental education; dentist; dentistry; endodontics; endodontology; educational assessment; educational measurement; examination; knowledge assessment; practical skills; skill assessment; theoretical knowledge; undergraduate; undergraduate curriculum; undergraduate education

Introduction

Measuring the outcome of education (ie, theoretical knowledge and practical skills) is one of the major issues in dental education. Preclinical teaching within the German undergraduate dental curriculum consists of 6 semesters. During this time, students are taught theoretical knowledge and practical skills for a variety of subjects. Usually, written examinations and practical skills tests are performed to monitor the students' progress; their capability to apply the acquired knowledge; and, therefore, their ability to treat patients during the subsequent clinical part of the undergraduate dental curriculum. Theoretical knowledge is often objectively assessed via written examinations that use different multiple-choice item types. Practical skills are often assessed by simulating clinical situations that have to be mastered by students. Without any doubt, the implementation of practical skills tests is a complicated process that requires a large amount of time and personnel input [1]. Therefore, several previous studies aimed to assess the correlation between the theoretical knowledge and practical competence of medical students [2-5]. However, only a few studies have evaluated the relation between theoretical knowledge and practical skills among dental students [1,6]. These studies assessed the correlation between students' achievements in written examinations and objective structured clinical examinations and reported a significant but moderate correlation between both assessments' scores [1,6].

Ideally, written examinations predict students' competence and preparedness for further challenges and advanced practice at the end of a preclinical course. Moreover, the measurement process should prevent false-negative results (failing a student who is competent) and false-positive results (passing a student who is incompetent) [7]. Similarly, when applying theoretical knowledge as a predictor for students' practical skills, an optimal cutoff value has to be calculated, so that the number of false-negative results (theoretically failing but being practically capable) and false-positive ones (theoretically passing but being practically incompetent) is reduced to the greatest possible extent.

In the field of endodontics, both the transfer of theoretical knowledge and the acquisition of basic practical skills play important roles in dental education [8]. A survey regarding undergraduate endodontic teaching among dental schools in the United Kingdom reported that lectures, seminars, tutorials, and laboratory or practical learning were the most frequently applied teaching formats [9]. Furthermore, the *Undergraduate Curriculum Guidelines for Endodontology* of the European Society of Endodontology recommends practical supervision by endodontic specialists or by educators with special interest and training in endodontics [8], which emphasizes the need for practical skills acquisition. However, detailed theoretical knowledge concerning root canal anatomy is a prerequisite for successful endodontic treatment [10]. A survey among undergraduate dental students confirmed the relevance of sufficient anatomical knowledge, as 74% of the students did not feel competent in treating posterior and multirooted teeth with complex anatomies [11]. Moreover, sufficient knowledge about the correct use of endodontic instrumentation systems and their properties, which differ due to the different alloys that these systems are made of, is required for error avoidance during root canal treatment [12]. Without any doubt, theoretical knowledge and the acquisition of practical skills seem to be important factors related to successful endodontic treatment. However, to the best of our knowledge, the correlation between these competencies and the predictive validity of theoretical examinations on the practical capability of dental students in preclinical courses have not been evaluated so far.

This study aimed to analyze the relation between students' theoretical knowledge and practical skills in endodontics. Furthermore, the predictive validity of a theoretical knowledge assessment on students' practical skills was assessed, and an optimal cutoff value for theoretical knowledge was defined.

The null hypothesis was that students' theoretical knowledge does not impact practical skills.

Methods

Ethics Approval

This study was approved by the local ethics committee of University Medical Center Göttingen (approval number: 23/10/22). The data analyzed in this study were routinely generated during students' undergraduate dental education. Participating students did not receive any compensation. The local ethics committee allowed for the secondary analysis of the data set without additional consent. The data set was anonymized prior to this study.

The use of anonymized extracted human teeth in routine teaching practices was approved by the local ethics committee of University Medical Center Göttingen (approval number: 27/8/13). Prior to the collection of extracted teeth during routine care, patients received written information, and informed consent was obtained.

Participants

All students who were enrolled in the preclinical phantom course in Operative Dentistry (sixth semester of the undergraduate dental curriculum in Germany) between the 2015 summer term and the 2022 summer term were included in the retrospective analysis. Students who did not participate in both the practical skills test and the final written examination (eg, course dropout or absence from examinations due to illness) were excluded.

Assessment of Theoretical Knowledge

Theoretical knowledge was assessed in summative electronic examinations, using the CAMPUS examination software (Umbrella Consortium for Assessment Networks [13]). Examinations took place at the end of each term and consisted of 30 items. Among these, single-choice items with 5 answer options (Type A items), multiple-select items with 4 to 6 statements (Multiple-True-False items), multiple-select items with 5 to 8 answer options (students were aware of the total number of true answer options that should be selected [Pick-N items]), and open-ended items were used. Single-choice and open-ended items were scored dichotomously (ie, examinees received either 0 or 1 credit point per item). Multiple-True-False items were scored according to the Vorkauf method [14] (in the literature, the terms *Halbpunkt-Bewertung* [14], *Half-point Scoring* [15], and *Vorkauf Method* [16] are used); examinees received 1 credit point if all statements per item were marked correctly, 0.5 credit points if only 1 statement per item was marked incorrectly, and 0 credit points if more than 1 statement per item was marked incorrectly. Pick-N items were scored according to the method proposed by Bauer et al [17] (in the literature, the terms *Partial Scoring 50%* [18] and PS_{50} [18] are used); examinees received 1 credit point if all true answer options were marked, 0.5 credit points if at least half of the true answer options were marked, and 0 credit points if less than half of the true answer options were marked. Prior to the examinations, all items were reviewed by multiple educators using a checklist for content and formal correctness. The total examination time amounted to 90 seconds per item. A fixed pass mark of 60% was applied.

Each examination covered 3 topics (Cariology/Restorative Dentistry, Endodontics, and Periodontics), but only items on endodontics were considered in this study to allow for a comparison with practical skills in endodontics. Students' theoretical knowledge was calculated as a relative percentage score based on the number of gained credit points.

Assessment of Practical Skills

Practical skills were assessed once per term in a standardized practical skills test. Students were given 2.5 hours to perform an endodontic treatment on an extracted human premolar. Teeth were previously embedded in polymethyl methacrylate (Paladur [Kulzer GmbH]) at their physiological position in full-arch models. During the examination, models were mounted in their maxillary or mandibular position and placed in a phantom head (Phantomkopf PK-2 with face mask P-6 GM [Frasaco GmbH]). During the treatment, the use of a rubber dam was mandatory. Assessed treatment steps included (1) the preparation of an endodontic access cavity, (2) the determination of endodontic working length, (3) the preparation of root canals, (4) the cold obturation of root canals by using gutta-percha and sealer, and (5) the cleaning of the endodontic access cavity. Before (preoperative), during (verification of working length), and after the treatments, x-ray images were taken.

Each treatment step was rated by an endodontic specialist using a piloted spreadsheet. For each treatment step, up to 3 raw points were awarded based on students' performance, and a final practical achievement score was calculated as a relative percentage score. Again, a pass mark of 60% was applied.

Statistical Analysis

All statistical analyses were performed by using the software R (version 4.2.1; R Foundation for Statistical Computing) [19]. The level of significance was set at an α of .05.

Variables that potentially impacted students' practical skills (ie, age, sex, previous course participation, and theoretical knowledge) were tested univariately, using Pearson correlations (continuous variables) or Wilcoxon rank sum tests (categorical variables). *P* values were corrected for multiple testing according to the Bonferroni-Holm method. Subsequently, variables were simultaneously entered in a multiple linear regression model.

The number of students with a practical skills level of ≥60% (ie, sufficient) or <60% (ie, insufficient) was determined. By applying the current pass mark of 60% for theoretical knowledge, these students were further categorized as students with a theoretical knowledge mark of <60% or ≥60%. The distribution of students among the four emerging categories was determined by using a Fisher exact test.

To determine the best lower limit for theoretical knowledge, the receiver operator characteristic curve method was applied at 1% intervals. Again, participants were divided into groups of students with a practical skills level of ≥60% or <60%. For the construction of the curve, the number of students with a theoretical knowledge level below a sliding delimiter among both groups was calculated. The Youden index (θ [theoretical knowledge≥delimiter | practical skills≥60%] + θ [theoretical knowledge≥delimiter | practical skills<60%] – 1) and the negative log of Fisher *P* values were used to calculate an optimal theoretical knowledge cutoff value.

Results

Participants

A total of 447 students with paired measurements of practical skills and theoretical knowledge were included in this study. Descriptive data and the univariate effect of each variable on students' practical skills are shown in Table 1. Only the level of theoretical knowledge was significantly associated with practical skills ($P_{adjusted}$=.02), indicating a small effect size (r=0.13).

The effects of the assessed variables on students' practical skills were further analyzed via linear regression analysis, as shown in Table 2. The overall model was significant ($F_{4,442}$=2.442; P=.046; $R^2_{adjusted}$=0.013), indicating a small effect size (f^2=0.01) [20]. Again, only the level of theoretical knowledge was associated with students' practical skills (P=.006). Therefore, the null hypothesis must be rejected.

Table 1. Univariate effects of assessed variables on students' practical skills (practical achievement score).

Variable	Value	Adjusted *P* value[a]
Age at time of practical skills test (years), mean (SD)	25.06 (3.44)	>.99
Sex, n (%)		.38
Female	307 (68.7)	
Male	140 (31.3)	
Previous course participation, n (%)		>.99
No	412 (92.2)	
Yes	35 (7.8)	
Theoretical knowledge (score), mean (SD)	76.54 (18.5)	.02

[a]Continuous variables were assessed with *P* values from Pearson correlations. Categorical variables were assessed with *P* values from Wilcoxon rank sum tests.

Table 2. Linear regression model[a].

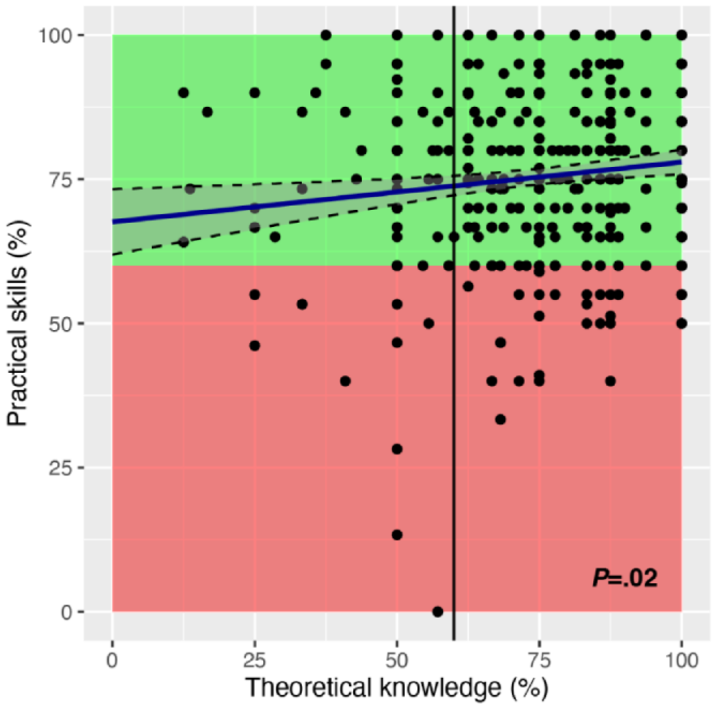
Variable	Estimate, B	*P* value
Age at time of practical skills test	−0.037	.85
Sex (female vs male)	−1.939	.19
Previous participation (yes vs no)	0.364	.89
Theoretical knowledge	0.102	.006

[a]Intercept=69.971 (*P*<.001).

Relation Between Theoretical Knowledge and Practical Skills

The relation between students' theoretical knowledge and practical skills is shown in Figure 1. By applying the current pass mark for theoretical knowledge, a significant differentiation was achieved (*P*=.02; Fisher exact test).

Figure 1. The relation between students' theoretical knowledge and practical skills. The regression line and 95% CI (dashed lines) are shown. To pass the practical skills test, a minimum practical achievement score of 60% was required (green-shaded area). For the theoretical knowledge assessment in written examinations, a pass mark of 60% was used (vertical line). The *P* value was obtained from a Fisher exact test.



Predictive Validity of Theoretical Knowledge on Practical Skills

Based on a fixed pass mark of 60% for practical skills, the area under the receiver operator characteristic curve amounted to 59.2% (95% CI 45.7%-72.1%; Figure 2).

The best lower limit (ie, pass mark) for theoretical knowledge was 58%, as indicated by the maximized negative log of Fisher *P* values (1.710) and a Youden index of 0.155 (Figure 3). The associated odds ratio amounted to 2.58 (95% CI 1.13-5.59), indicating that students with a theoretical knowledge mark below 58% are 1.22 times more likely to show insufficient practical skills (<60%).

Figure 2. The receiver operator characteristic curve and 95% CI (dashed lines) of theoretical knowledge marks. The putative pass mark for theoretical knowledge was used as a sliding delimiter (0%-100%). AUC: area under the curve.

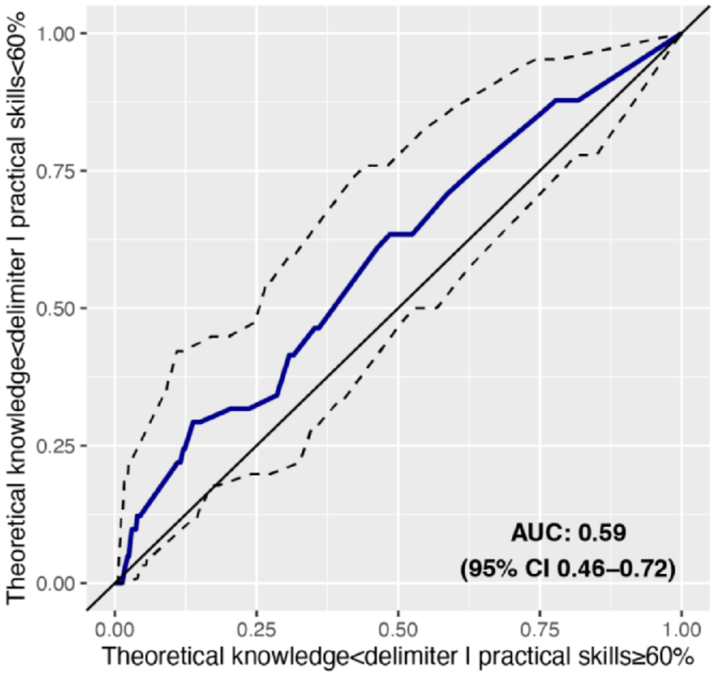
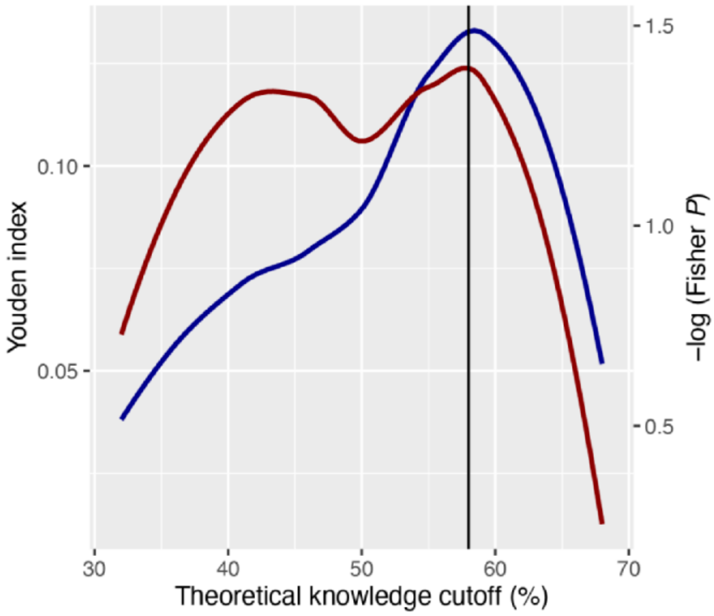


Figure 3. The relation between the Youden index and theoretical knowledge (blue line) and the relation between the negative log of Fisher *P* values and theoretical knowledge (red line) are shown. The putative pass mark for theoretical knowledge was used as a sliding delimiter (0%-100%). Smoothed lines are shown. The maximized Youden index and negative log of Fisher *P* values are indicated by the vertical line.



Discussion

Principal Findings

This study shows that a significant differentiation between students with sufficient practical skills and students with insufficient practical skills can be achieved by applying the currently used pass mark of 60% on written examinations assessing theoretical knowledge. Nevertheless, it must be mentioned that every test has its limitations and unavoidably results in a specific number of false decisions. Aiming to maximize the number of correctly categorized students (ie, true positives: theoretically passing and being practically capable; true negatives: theoretically failing and being practically incapable), we calculated the best lower limit for theoretical knowledge, which amounted to 58%. Furthermore, the results of this study indicate that students who show insufficient theoretical knowledge (ie, those who achieve a score below 58% in the written multiple-choice examination) are 1.22 times more likely to show insufficient practical skills (ie, achieving a score below 60% in the practical skills test).

Comparison to Prior Work

The assessment of clinical competence is one of the major issues in medical and dental education. Written examinations, especially those consisting of multiple-choice items, are widely used to objectively assess students' theoretical knowledge [2,4,6]. However, no single assessment tool that measures all facets of clinical competence has been established yet [21]. The Miller pyramid illustrates several stages of clinical competence, using the terms *knows*, *knows how*, *shows how*, and *does* [22]. Multiple-choice items in written examinations are suitable for assessing both basic facts and applied knowledge (ie, the two lower levels of the Miller pyramid) [21]. Higher levels of clinical competence need to be tested by using practical skills tests or by simulating clinical situations, such as in objective structured clinical examinations [21]. However, according to the Miller pyramid, it is suggested that practical skills are based on sufficient theoretical knowledge. Moreover, the implementation of practical skills tests is cost-intensive, requires many resources, and results in a large amount of personnel effort [1]. Due to these circumstances, previous studies investigated the relationship between the theoretical knowledge and practical skills of medical students [2,4,23-26]. Many of these studies found a significant but weak to moderate correlation between theoretical knowledge and practical skills [1,2,4,6,23,24]. However, some studies failed to show a significant correlation between students' theoretical knowledge and practical skills [25,26], leading to contradicting results.

Only a few studies have explored the relation between the theoretical knowledge and practical skills of undergraduate dental students [1,6]. Confirming the findings of other previous investigations in medicine, these studies found a moderate correlation between theoretical knowledge and practical performance [1,6]. Similarly, our study shows that theoretical knowledge is significantly associated with practical skills among undergraduate dental students performing root canal treatments ($P_{adjusted}$=.02). Two of the major advantages of being able to anticipate the future practical performance of dental students are the possibilities of early intervention and individual promotion. Students who do not perform well on written examinations may benefit from closer monitoring during the early stages of clinical practice. Likewise, students with excellent theoretical performances may be further encouraged and challenged by providing them with more complex cases.

Strengths and Limitations

The major strengths of this study are the large number of dental students who participated in the preclinical phantom course in Operative Dentistry between the 2015 summer term and the 2022 summer term and the inclusion of student-related variables (ie, age, sex, and previous course participation). However, different limitations are also present. First, the assessed predictive validity of students' theoretical knowledge on practical skills was based on the used written examinations and practical skills tests. Second, the practical skills tests used extracted human premolars, which potentially could have resulted in inequities. However, all teeth were assessed via x-ray images prior to the practical skills tests, and teeth were excluded if any anatomical difficulties were obvious. Thereby, similar levels of difficulty for the practical skills tests were ensured. The use of extracted human teeth in endodontic skills tests is recommended, as students' performance on tests involving artificial teeth does not predict their future performance during clinical treatments [27]. Third, the results reflect the competence of undergraduate preclinical course students in performing root canal treatments on extracted teeth of low-level difficulty. Further research is required to assess the relation between the theoretical knowledge and practical competence of more experienced students who treat patients and are confronted with more demanding tasks (ie, during clinical teaching). Fourth, the COVID-19 pandemic occurred in the middle of the study period. However, practical teaching was always fully carried out on site while ensuring sufficient physical distancing (eg, students were placed in 2 cohorts), and theoretical knowledge was partially taught via the internet, as outlined in a previous publication [28]. Thereby, all participating students completed the full practical curriculum, and the pandemic likely did not impact the presented results.

Future Directions

Although this study found statistically significant results, the weak correlation does not warrant an exact prediction of the practical skills test outcome. Even though the results confirm that the acquisition of sufficient theoretical knowledge is associated with adequate practical skills, the need for the integration of practical courses must be emphasized. Interestingly, the linear regression model of this study shows that previous but unsuccessful participation in the preclinical phantom course had no effect on the outcomes of the practical skills tests when compared to first-time participation in the course. Moreover, 2 previous studies regarding students' self-perceptions during practical courses reported that most dental students still do not feel confident and competent when performing nonsurgical root canal treatments, especially on premolars and molars [11,29]. This study confirms that theoretical knowledge and extensive practical training (beyond the preclinical course) in endodontics are required to comprehend the importance of each single step in endodontic treatment [30].

Conclusion

This study provided valuable information concerning the relation between students' theoretical knowledge and practical skills for performing endodontic treatments. By objectively measuring students' theoretical knowledge, a rough estimation of students' practical skills (ie, a differentiation between sufficient and insufficient practical skills) is possible.

Acknowledgments

The authors acknowledge support by the Open Access Publication Funds of Göttingen University. The funder had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Availability

The data sets generated during or analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

Both authors contributed to this study's conception and design. PK performed statistical analyses. FH und PK interpreted the data. Both authors drafted the manuscript.

Conflicts of Interest

None declared.

References

1. Eberhard L, Hassel A, Bäumer A, Becker F, Beck-Mubotter J, Bömicke W, et al. Analysis of quality and feasibility of an objective structured clinical examination (OSCE) in preclinical dental education. Eur J Dent Educ 2011 Aug;15(3):172-178. [doi: 10.1111/j.1600-0579.2010.00653.x] [Medline: 21762322]
2. Auewarakul C, Downing SM, Jaturatamrong U, Praditsuwan R. Sources of validity evidence for an internal medicine student evaluation system: an evaluative study of assessment methods. Med Educ 2005 Mar;39(3):276-283. [doi: 10.1111/j.1365-2929.2005.02090.x] [Medline: 15733163]
3. Dong T, Saguil A, Artino ARJ, Gilliland WR, Waechter DM, Lopreaito J, et al. Relationship between OSCE scores and other typical medical school performance indicators: a 5-year cohort study. Mil Med 2012 Sep;177(9 Suppl):44-46. [doi: 10.7205/milmed-d-12-00237] [Medline: 23029860]
4. Eftekhar H, Labaf A, Anvari P, Jamali A, Sheybaee-Moghaddam F. Association of the pre-internship objective structured clinical examination in final year medical students with comprehensive written examinations. Med Educ Online 2012;17:15958 [FREE Full text] [doi: 10.3402/meo.v17i0.15958] [Medline: 22547924]
5. Kirton SB, Kravitz L. Objective structured clinical examinations (OSCEs) compared with traditional assessment methods. Am J Pharm Educ 2011 Aug 10;75(6):111 [FREE Full text] [doi: 10.5688/ajpe756111] [Medline: 21931449]
6. Ali K, Jerreat M, Zahra D, Tredwin C. Correlations between final-year dental students' performance on knowledge-based and clinical examinations. J Dent Educ 2017 Dec;81(12):1444-1450. [doi: 10.21815/JDE.017.105] [Medline: 29196332]
7. van der Vleuten C. Validity of final examinations in undergraduate medical training. BMJ 2000 Nov 11;321(7270):1217-1219 [FREE Full text] [doi: 10.1136/bmj.321.7270.1217] [Medline: 11073517]
8. De Moor R, Hülsmann M, Kirkevang LL, Tanalp J, Whitworth J. Undergraduate curriculum guidelines for endodontology. Int Endod J 2013 Dec;46(12):1105-1114. [doi: 10.1111/iej.12186] [Medline: 24117830]
9. Al Raisi H, Dummer PMH, Vianna ME. How is endodontics taught? A survey to evaluate undergraduate endodontic teaching in dental schools within the United Kingdom. Int Endod J 2019 Jul;52(7):1077-1085. [doi: 10.1111/iej.13089] [Medline: 30706491]
10. Vertucci FJ. Root canal morphology and its relationship to endodontic procedures. Endod Topics 2005 Mar;10(1):3-29. [doi: 10.1111/j.1601-1546.2005.00129.x]
11. Davey J, Bryant ST, Dummer PMH. The confidence of undergraduate dental students when performing root canal treatment and their perception of the quality of endodontic education. Eur J Dent Educ 2015 Nov;19(4):229-234. [doi: 10.1111/eje.12130] [Medline: 25490882]
12. Zupanc J, Vahdat-Pajouh N, Schäfer E. New thermomechanically treated NiTi alloys - a review. Int Endod J 2018 Oct;51(10):1088-1103. [doi: 10.1111/iej.12924] [Medline: 29574784]
13. CAMPUS | UCAN ASSESS. Institute for Communication and Assessment Research. URL: https://www.ucan-assess.org/campus/?lang=en [accessed 2022-10-24]

14. Vorkauf H. Teilpunktbewertung bei K'-Items [Partial credit scoring of Multiple-True-False items]. In: Jahresbericht 1986 der Gruppe Medizinalprüfungen und der Gruppe Statistik und EDV. Bern, Switzerland: Institut für Ausbildungs- und Examensforschung, Medizinische Fakultät der Universität Bern; 1987:44-48.
15. Schmidt D, Raupach T, Wiegand A, Herrmann M, Kanzow P. Relation between examinees' true knowledge and examination scores: systematic review and exemplary calculations on Multiple-True-False items. Educ Res Rev 2021 Nov;34:100409 [FREE Full text] [doi: 10.1016/j.edurev.2021.100409]
16. Kanzow P, Schuelper N, Witt D, Wassmann T, Sennhenn-Kirchner S, Wiegand A, et al. Effect of different scoring approaches upon credit assignment when using Multiple True-False items in dental undergraduate examinations. Eur J Dent Educ 2018 Nov;22(4):e669-e678. [doi: 10.1111/eje.12372] [Medline: 29934980]
17. Bauer D, Holzer M, Kopp V, Fischer MR. Pick-N multiple choice-exams: a comparison of scoring algorithms. Adv Health Sci Educ Theory Pract 2011 May;16(2):211-221 [FREE Full text] [doi: 10.1007/s10459-010-9256-1] [Medline: 21038082]
18. Schmidt D, Raupach T, Wiegand A, Herrmann M, Kanzow P. Relation between examinees' true knowledge and examination scores: systematic review and exemplary calculations on Pick-N items. Educ Res Rev 2022 Nov;37:100483 [FREE Full text] [doi: 10.1016/j.edurev.2022.100483]
19. The R project for statistical computing. R Foundation for Statistical Computing. URL: https://www.r-project.org [accessed 2022-11-09]
20. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155] [Medline: 19565683]
21. Wass V, Van der Vleuten C, Shatzer J, Jones R. Assessment of clinical competence. Lancet 2001 Mar 24;357(9260):945-949. [doi: 10.1016/S0140-6736(00)04221-5] [Medline: 11289364]
22. Miller GE. The assessment of clinical skills/competence/performance. Acad Med 1990 Sep;65(9 Suppl):S63-S67. [Medline: 2400509]
23. Sandoval GE, Valenzuela PM, Monge MM, Toso PA, Triviño XC, Wright AC, et al. Analysis of a learning assessment system for pediatric internship based upon objective structured clinical examination, clinical practice observation and written examination. J Pediatr (Rio J) 2010;86(2):131-136. [doi: 10.2223/JPED.1986] [Medline: 20231951]
24. Tijani KH, Giwa SO, Abiola AO, Adesanya AA, Nwawolo CC, Hassan JO. A comparison of the objective structured clinical examination and the traditional oral clinical examination in a Nigerian university. J West Afr Coll Surg 2017;7(3):59-72 [FREE Full text] [Medline: 30525003]
25. Schwartz RW, Donnelly MB, Sloan DA, Johnson SB, Strodel WE. The relationship between faculty ward evaluations, OSCE, and ABSITE as measures of surgical intern performance. Am J Surg 1995 Apr;169(4):414-417. [doi: 10.1016/s0002-9610(99)80187-1] [Medline: 7694980]
26. Johnson G, Reynard K. Assessment of an objective structured clinical examination (OSCE) for undergraduate students in accident and emergency medicine. J Accid Emerg Med 1994 Dec;11(4):223-226 [FREE Full text] [doi: 10.1136/emj.11.4.223] [Medline: 7894806]
27. Bitter K, Gruner D, Wolf O, Schwendicke F. Artificial versus natural teeth for preclinical endodontic training: A randomized controlled trial. J Endod 2016 Aug;42(8):1212-1217. [doi: 10.1016/j.joen.2016.05.020] [Medline: 27469437]
28. Kanzow P, Krantz-Schäfers C, Hülsmann M. Remote teaching in a preclinical phantom course in Operative Dentistry during the COVID-19 pandemic: Observational case study. JMIR Med Educ 2021 May 14;7(2):e25506 [FREE Full text] [doi: 10.2196/25506] [Medline: 33941512]
29. Murray CM, Chandler NP. Undergraduate endodontic teaching in New Zealand: students' experience, perceptions and self-confidence levels. Aust Endod J 2014 Dec;40(3):116-122. [doi: 10.1111/aej.12084] [Medline: 25307451]
30. Picart G, Pouhaër M, Dautel A, Pérard M, Le Clerc J. Dental students' observations about teaching of endodontic access cavities in a French dental school. Eur J Dent Educ 2022 Aug;26(3):499-505 [FREE Full text] [doi: 10.1111/eje.12726] [Medline: 34808027]

Edited by A Mavragani; submitted 06.02.23; peer-reviewed by J Jeong, M Pang; comments to author 14.03.23; revised version received 14.03.23; accepted 30.03.23; published 18.04.23.

Please cite as:
Haupt F, Kanzow P
The Relation Between Students' Theoretical Knowledge and Practical Skills in Endodontics: Retrospective Analysis
Interact J Med Res 2023;12:e46305
URL: https://www.i-jmr.org/2023/1/e46305
doi:10.2196/46305
PMID:36994987

Original Paper

Development of a Pilot Introductory Advanced Cardiovascular Resuscitation Course for Senior Medical Students in Switzerland: Student-Driven Implementation Study

Tara Herren[1*], MSc; Loris Fivaz[1*], MSc; Eva Dufeil[1*], MSc; Eric Golay[1], MAS; Ely Braun[1], DAS; Emilie Straub[1], EMT-P; Philippe Nidegger[1], EMT-P; Olivier Grosgurin[1], MD; Birgit Andrea Gartner[1], MD; Mélanie Suppan[2], MD; Laurent Suppan[1], MD

[1]Division of Emergency Medicine, Department of Anesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine, University of Geneva Hospitals and Faculty of Medicine, Geneva, Switzerland

[2]Division of Anesthesiology, Department of Anesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine, University of Geneva Hospitals and Faculty of Medicine, Geneva, Switzerland

[*]these authors contributed equally

Corresponding Author:
Laurent Suppan, MD
Division of Emergency Medicine
Department of Anesthesiology, Clinical Pharmacology, Intensive Care and Emergency Medicine
University of Geneva Hospitals and Faculty of Medicine
Rue Gabrielle-Perret-Gentil 4
Geneva, 1211
Switzerland
Phone: 41 223723311
Email: laurent.suppan@hcuge.ch

Abstract

Background: Cardiac arrest is the most time-critical emergency medical students and junior physicians may face in their personal or professional life. However, many studies have shown that most of them lack the necessary knowledge and skills to efficiently perform resuscitation. This could be related to the fact that advanced cardiovascular resuscitation courses are not always part of the undergraduate medical curriculum.

Objective: The aim of this study was to describe the development, pilot implementation, and assessment of an advanced cardiovascular resuscitation course designed to enable senior medical students to manage the initial resuscitation phase in case of cardiac arrest.

Methods: An introductory advanced cardiovascular resuscitation course was developed on the initiative of fifth-year medical students, in collaboration with the prehospital emergency medical service team of the Geneva University Hospitals. The 60 slots available to the 157 members of the fifth-year promotion of the University of Geneva Faculty of Medicine were filled in less than 8 hours. This unexpected success prompted the creation of a first questionnaire, which was sent to all fifth-year students to determine the overall proportion of students interested in attending an advanced cardiovascular resuscitation course. This questionnaire was also used to assess basic life support education and experience among course participants. A postcourse questionnaire was used to gather feedback regarding the course and to assess student confidence regarding the resuscitation skills they had been taught.

Results: Out of 157 fifth-year medical students, 73 (46%) completed the first questionnaire. Most thought that the current curriculum did not provide them with enough knowledge and skills regarding resuscitation and 85% (62/73) wished to attend an introductory advanced cardiovascular resuscitation course. All the participants who would have wanted to follow the full Advanced Cardiovascular Life Support course before graduating were set back by its cost (10/10, 100%). Of the 60 students who had registered for the training sessions, 56 (93%) actually attended. The postcourse questionnaire was completed by 42 (87%) students (out of 48 who had registered on the platform). They unanimously answered that an advanced cardiovascular resuscitation course should be part of the standard curriculum.

Conclusions: This study demonstrates the interest of senior medical students in an advanced cardiovascular resuscitation course and their willingness to see such a course integrated as a part of their regular curriculum.

(Interact J Med Res 2023;12:e46075) doi:10.2196/46075

KEYWORDS

advanced cardiovascular life support; undergraduate medical education; cardiopulmonary resuscitation; CPR; medical education; resuscitation; web-based questionnaire; collaborative design; implementation; medical course; curriculum; student; life support; training; cardiac arrest; medical student

Introduction

Background

Senior medical students about to graduate are sometimes expected to have skills close to those of certified physicians even though their training is still ongoing [1]. During the COVID-19 pandemic, these expectations were particularly emphasized, and many senior undergraduate students were given responsibilities akin to those of residents [2]. Regardless of the context of this recent crisis, any medical student can be exposed to critical emergencies whether inside or outside the hospital [1,3,4]. Indeed, they are expected to adequately manage resuscitations even though they have not yet graduated [5]. Moreover, in Switzerland as in many other countries, newly graduated physicians spend their first year of residency in peripheral hospitals where senior practitioners are not constantly present. These physicians should therefore be able to manage the first 10 minutes of resuscitation without external help. However, many studies show that senior medical students who are about to graduate are not proficient in resuscitation skills despite significant improvements throughout their curriculum [5-10]. Furthermore, most senior medical students feel unconfident putting these skills into action [1,11]. This feeling could be explained by the paucity of refresher sessions during their undergraduate curriculum and by the impression that their resuscitation training is not comprehensive or advanced enough [12].

Local Setting

During their 6-year curriculum, medical students at the University of Geneva Faculty of Medicine (UGFM), Switzerland, follow 4 basic life support (BLS) training sessions. Even though the last of these sessions, which takes place during their fifth year, confronts them with more difficult cases requiring a better understanding of the pathophysiology of cardiac arrest (CA), they are not expected to practice skills beyond standard BLS procedures.

In 2021, a small group of fifth-year medical students was interested in taking an Advanced Cardiovascular Life Support (ACLS) course since they were concerned about their resuscitation skills. Since this course is both expensive and time-consuming, they contacted the prehospital emergency medical service (Mobile Emergency and Resuscitation Service [SMUR for Service Mobile d'Urgence et de Réanimation, in French]) of the Geneva University Hospitals to determine if an introductory advanced cardiovascular resuscitation course could be organized free of charge. A total of 60 training slots were provided to the 157 students of the fifth-year promotion, all of which were filled in less than 8 hours. This unexpected enthusiasm prompted the elaboration of a questionnaire to explore fifth-year medical students' support of the inclusion of an introductory advanced cardiovascular resuscitation course during the undergraduate emergency training program as part of the standard curriculum.

Objectives

The goal of this study was to describe the development, pilot implementation, and assessment of an introductory advanced cardiovascular resuscitation course designed to enable senior medical students to manage the first 10 minutes of resuscitation in case of CA.

Methods

Ethics Approval

In accordance with the Swiss federal law on human research, the regional ethics committee (Commission Cantonale d'Ethique de la Recherche sur l'être humain—CCER—Geneva, Switzerland) issued a "declaration of no objection" (Req-2021-00628) regarding this study [13].

Study Design

This was a retrospective analysis of data collected prospectively through fully automated web-based questionnaires. Participants were informed that anonymized data would be collected and presented to the UGFM committee for undergraduate education. The students were also informed by email that a publication was considered and had the opportunity to ask questions and express their potential opposition. Methods and results are reported according to the CHERRIES (Checklist for Reporting Results of Internet E-Surveys) guidelines [14].

Enrollment and Precourse Questionnaire

Information about the course was dispatched to all fifth-year medical students by email on February 18, 2021, and was simultaneously posted on the social network used for the promotion. The fifth-year UGFM promotion included 157 students representing our convenience sample. Students could register for the course by responding to the invitation email on a first come, first served basis. No financial incentive was given to promote participation, and there were no exclusion criteria.

A web-based platform was created using the Joomla 3.9 content management system (Open Source Matters) to host the web-based questionnaires, which were created using the Community Surveys 5.5 component (Shondalai). The platform and questionnaires were thoroughly tested by 4 of the authors. The first questionnaire was sent to the participants along with practical information regarding the training sessions on April 1, 2021, and a reminder was sent 3 days later. This questionnaire was divided into 2 sections (Table 1). The aim of the first section was to determine the potential number of students interested in such a training and was intended for the all fifth-year UGFM students, regardless of their participation in or interest for the course. Registered participants accessed the second part, which was specifically designed to assess their comfort with the BLS skills they had acquired through the standard training curriculum and their desire to follow an official ACLS course prior to registering to the introductory advanced cardiovascular resuscitation course. After completing this part of the questionnaire, participants were prompted to create an account on the web-based platform.

To make sure that participants would reap the highest possible benefit from the course and to facilitate the flow of the practice sessions, they were asked to read a summary of the ACLS guidelines [15] before attending the course.

Table 1. Precourse questionnaire.

Survey page, field, and questions	Type of question
Page 1	
Demographics	
Age	Open[a]
Gender	MCQ[b]
Knowledge and interest about ACLS[c]	
Ever heard of ACLS	Yes or no
Benefits expected from attending an ACLS course	MAQ[d]
Regarding the class organized this year	
Registration	Yes or no
If yes: go to questions on page 2	
If no: wished they could have participated	Yes or no
Page 2	
BLS[e] awareness	
Number of BLS-AED[f] courses attended	MCQ
Ability to use the already acquired BLS knowledge	1-5 Likert scale
Ever performed chest compression in a real situation	Yes or no
Interest in taking an ACLS class before graduation	
Wish to obtain an ACLS certification	Yes or no
Impeding factors	MAQ
Advanced resuscitation course as part of the curriculum	1-5 Likert scale
Interest in emergency medicine	
Willingness to specialize in emergency medicine	MCQ

[a]A Regex (regular expression) validation rule was used to avoid invalid entries.
[b]MCQ: multiple-choice question (only 1 answer accepted).
[c]ACLS: advanced cardiovascular life support.
[d]MAQ: multiple-answer question (more than 1 answer accepted).
[e]BLS: basic life support.
[f]AED: automatic external defibrillator.

Design and Sequence of the Course

There were 5 training sessions between April 10 and June 12, 2021, and students were divided into groups accordingly. The course was designed by the SMUR team in collaboration with senior medical students whose pre-existing knowledge was taken into account. Course sessions lasted 5.5 hours and were scheduled during the weekends to reduce the impact on the standard teaching curriculum. Their structure, mostly based on current ACLS guidelines [11], is detailed in Table 2.

Three main themes were covered during the simulation sessions: rhythm recognition and management, drug use and timing of drug administration, and specialized care after the return of spontaneous circulation. Nontechnical skills such as leadership and communication were also practiced during these simulations. Regarding airway management procedures, students were taught to prepare and insert an i-gel supraglottic airway device since such devices enhance oxygenation and ventilation and do not require the level of expertise needed to perform endotracheal intubation [16].

To ensure a personalized and efficient training experience, there was a ratio of 1 instructor for 4 medical students. This also helped adhere to the COVID-19 infection prevention guidelines, which were in effect at the time of this study.

All instructors were SMUR paramedics certified in BLS training and used to teach advanced resuscitation skills. In Switzerland, paramedics follow a 3-year curriculum and are able to take care autonomously of a wide range of injured or ill patients of all ages. They are allowed to insert intravenous lines, use supraglottic airway devices, and administer a wide variety of drugs including epinephrine, antiarrhythmic agents such as amiodarone, and opiates such as fentanyl [17]. Paramedics staffing SMUR units possess additional skills and have access to advanced medications such as hypnotics and neuromuscular blocking agents. These advanced paramedics always work in pairs with physicians who are either senior residents, registrars, or even senior specialists in prehospital emergency medicine [18].

An unofficial course completion certificate was given to the students at the end of the course.

Table 2. Contents of the introductory advanced life support training course.

Topic and method	Duration (minutes)
Primary survey	
Theoretical explanation and demonstration	30
BLS-AED[a] refresher	
Workshop	45
Airway management procedures	
Theoretical explanation, demonstration, and practice	30
Team dynamics during cardiac arrest situations	
Demonstration	25
Simulation and debriefing	90
Immediate care after ROSC[b]	
Demonstration	30
Simulation and debriefing	90

[a]BLS-AED: basic life support and automatic external defibrillator.

[b]ROSC: return of spontaneous circulation.

Postcourse Questionnaire

A few hours after the end of the training session, participants received an email containing a link to the postcourse questionnaire (Table 3). The goal of this questionnaire, which was administered using the same platform, was to determine whether participants thought that this course should be integrated into the regular curriculum.

Since participants had to be registered on the platform to answer this last questionnaire, we were able to send regular reminders to enhance the participation rate.

Table 3. Postcourse questionnaire.

Survey page, field, and questions	Type of question
Page 1	
General opinion	
Liked the course?	Yes or no
If yes: why	MAQ[a]
If no: why	MAQ
Was the course format adequate?	Yes or no
If no: why	MCQ[b]
Suggestions for improvement	Open
Usefulness of the content	1-5 Likert scale
Met expectations	1-5 Likert scale
Link with faculty program	
Should be integrated in the standard curriculum	1-5 Likert scale
Confidence	
Confidence in their ability to apply the knowledge learnt	1-5 Likert scale

[a]MAQ: multiple-answer question (more than 1 answer accepted).
[b]MCQ: multiple-choice question (only 1 answer accepted).

Outcomes

The primary outcome was the proportion of fifth-year students wishing to follow an introductory advanced cardiovascular resuscitation course, regardless of whether they had been able to attend one of the sessions organized for their promotion. Secondary outcomes were the assessment of the motivations and impeding factors to follow such a course. Comments about the usefulness of the pilot course and potential modification, their confidence regarding the skills they learned, and their opinion on its integration into the standard curriculum were also analyzed.

Data Curation and Statistical Analysis

All data were stored on an encrypted MySQL database (MariaDB 5.5.5) and extracted to a CSV file. Data curation and analysis were performed under Stata (version 16.1; StataCorp LLC). Incomplete questionnaires were not analyzed. The results are presented using descriptive statistics (n [%] and median [IQR]). Normality was assessed graphically, and between-group comparisons were carried out using the chi-square test, except for age for which the Mann-Whitney Wilcoxon test was used. A *P* value lower than .05 was considered significant.

Results

Precourse Questionnaire

Out of 157 fifth-year medical students, 73 completed the first questionnaire (73/157, 46%). There was no opposition from students regarding the use of their answers for publication. Participation was significantly higher (*P*<.001) in the subgroup of students who had registered to attend a training session (48/60, 80%). The characteristics of the respondents who had been able to register were not different from those who had not been able to register (Table 4).

The proportion of participants who wished to attend an introductory advanced cardiovascular resuscitation course was 85% (62/73). Their main motivation was that they thought it would help them to prepare better for residency (61/62, 98%). Most students were also interested in improving their resuscitation skills (53/62, 85%) and in increasing their knowledge (46/62, 74%). One student (1/62, 2%) reported an interest in the interprofessional aspect of this course that is, working alongside paramedics. Few of those who wished to attend thought that the current BLS-automatic external defibrillator (AED) curriculum provided them with enough knowledge and skills regarding resuscitation (2/62, 3%). This proportion was significantly higher (*P*=.001) in those who did not wish to attend such a course (5/11, 45%).

Students who had registered for a practice session had previously attended a median number of 4 (IQR 3-4) BLS-AED courses. Most felt confident or very confident in their BLS-AED abilities (35/48, 73%), but only a few of them had already performed cardiopulmonary resuscitation in an actual CA (4/48, 9%). Of the 10 (21%) participants motivated to follow the full ACLS course before graduating, all were set back by the cost of this course (10/10, 100%).

Of the 48 participants who completed the precourse questionnaire, 19 (40%) considered specializing in emergency medicine, intensive care medicine, or anesthesiology.

Table 4. Comparison of subgroups who answered the questionnaire.

Characteristics	Registered (n=48)	Unregistered (n=25)	*P* value
Gender, n (%)			.64
Men	16 (33)	11 (44)	
Women	31 (64)	14 (56)	
Refused to answer	1 (2)	0 (0)	
Age (years), median (IQR)	24 (21-27)	24 (23-24)	.90
Knew what ACLS[a] was before course proposition, n (%)	32 (67)	20 (80)	.28

[a]ACLS: advanced cardiovascular life support.

Postcourse Questionnaire

Forty-two of the 48 students who had completed the precourse questionnaire filled the postcourse questionnaire (42/48, 87%). There was a 93% (56/60) attendance to the practice sessions.

All the participants who answered the postcourse questionnaire reported that they liked the course (42/42, 100%) and that they found its content and structure adequate. All participants thought that the course was either very useful (39/42, 93%) or useful (3/42, 7%). They linked their appreciation to the acquisition of new knowledge (40/42, 95%), to the perceived usefulness of the course (36/42, 86%), and to the opportunity of practicing resuscitation skills (39/42, 93%). Nine students wrote free comments. Four of these comments were directly linked to the importance of practicing resuscitation skills. One student commented "we tend to take for granted too easily the gestures in theory when we notice that it is not the case in practice." Three students complimented the pedagogy of the instructors: "Very friendly instructors, and very interactive course." One student acknowledged the interest of including leadership skills training in the course: "Very interesting leadership training." One student answered that they would feel less helpless if confronted with a CA.

Most students answered that the content of the course had completely fulfilled their expectations (39/42, 93%). The most common suggestion for improvement was to provide an electronic learning (e-learning) module rather than text documents as references prior to the course.

After the course, the majority of participants (39/42, 93%) felt either confident (26/42, 62%) or very confident (13/42, 31%) in their ability to apply the skills they had been taught during the course.

Finally, all participants thought that such a course should be part of the regular curriculum, with 93% of them answering that it should "absolutely" be a part of it.

Discussion

Main Considerations

This study shows that many fifth-year medical students are highly supportive of the integration of an introductory advanced cardiovascular resuscitation course in their curriculum to feel better prepared for their first professional experience and the responsibilities it implies. The process which led to the creation of this advanced cardiovascular resuscitation course and the results of this study indicate that many students feel unprepared to manage time-critical emergencies such as CA, and ACLS courses are often offered too late during residency. Consequently, while BLS-AED refresher courses need to be held on a regular basis, advanced resuscitation skills should also be taught at least to senior medical students.

Conflicting with the results of many previous studies reporting that BLS skills are often lacking among health care students [5-7], most participants answered that they felt comfortable using such skills. This is of particular importance since the lack of confidence is often reported as a reason for inaction, especially in CA [19-21]. Promoting the confidence of health care students in their abilities to adequately provide resuscitation maneuvers is of paramount importance for several reasons. First, the expectations of the population toward medical students increase as they progress through their studies, and their resuscitation skills should therefore be developed enough to enable them to respond adequately in case of out-of-hospital cardiac arrest (OHCA). Second, the proportion of OHCA victims receiving BLS can and should be improved. In Geneva, Switzerland, the proportion of OHCA victims receiving BLS was lower than 40% between 2009 and 2012 [22]. Off-duty medical students could help improve this proportion, either on their own or as part of a first-responder network [21].

The disappointing results regarding resuscitation skills reported by many studies are probably linked to skill and information retention. Indeed, it has been shown that BLS-AED skills and knowledge decrease significantly within months after the last training session [23,24]. The same issue affects advanced life support skills among anesthetists [25]. The 2021 European Resuscitation Council Guidelines for education do not rule on the optimal frequency and method to prevent skill decay, which varies greatly according to the population studied [26].

Since the undergraduate medical curriculum is already very dense and demanding, there is significant tension between the importance of teaching more advanced resuscitation skills to senior medical students and the need to continue practicing basic resuscitation maneuvers. Alternative teaching methods could therefore be considered to enhance flexibility and efficiency. In line with one of the comments recorded by a student, including an e-learning module could prove worthwhile. Indeed, interactive e-learning modules and serious games have shown many advantages compared to traditional lectures and their use has been tremendously developed during the

COVID-19 pandemic [27-29]. These modules tend to decrease costs since asynchronous distance learning does not require the presence of an instructor or even the availability of a classroom. However, despite the aforementioned advantages and their ability to significantly enhance knowledge acquisition [30], e-learning modules also present incontrovertible limitations since skill acquisition can hardly be achieved through theoretical interventions [31]. Blended courses, which ally the best of both worlds (e-learning and hands-on practice sessions), could therefore represent an effective solution [32,33]. This is even more important in the context of an advanced cardiovascular resuscitation course, which must include elements linked to nontechnical skills such as leadership, decision-making, and team working [34]. Advanced simulations have proven to be particularly effective in helping to develop such skills while honing technical ones [35].

Limitations

A selection bias cannot be ruled out since the questionnaires were mostly completed by students who had already registered as participants for this course and were therefore probably more interested than some of their colleagues in this particular domain. Therefore, the postcourse confidence may be overestimated, and the lack of a similar question in the precourse questionnaire prevented the assessment of the participants' confidence. In addition, the design of the first questionnaire, the second part of which could not be filled by the students who had not been able to register for the course may have prevented the acquisition of potentially useful data. Moreover, even though there is considerable overlap between the pregraduate medical curriculum of many universities, we must acknowledge that our convenience sample only consisted of UGFM medical students and that our results might not apply to other universities. Finally, the methods used in this study were hardly ideal. Indeed, given the unforeseen enthusiasm of senior medical students for an advanced cardiovascular resuscitation course, this was not a preplanned study, and the timing of some interventions was far from perfect (1 practice session was held during the end-of-year examinations, while another took place during the holidays). This last limitation might have dampened the participation rate.

Perspectives

Following our initial results, the UGFM committee for undergraduate education decided to integrate a mandatory blended learning advanced resuscitation course, including an interactive e-learning module, into the pregraduate medical curriculum. The uptake of this course should now be assessed, and its potential shortcomings addressed before an assessment of its actual impact on the knowledge, skills, and confidence of senior medical students can be carried out. Assessing the impact of such a course on confidence through a thorough validated questionnaire would be most relevant according to the theory of planned behavior [36,37]. Finally, particular attention should be paid to the development of nontechnical skills (leadership, task management, decision-making, team working, and situational awareness), which deserve to be carefully assessed and improved.

Conclusions

This study demonstrates the interest of senior medical students in an advanced cardiovascular resuscitation course and their willingness to see such a course integrated as part of their regular curriculum. Regardless of the course format, enabling senior medical students to acquire advanced life support knowledge and skills should help them manage more efficiently the time-critical emergencies they will encounter either prior to graduation or during their first years of residency.

Data Availability

The data sets used and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

TH, ED, LF, EG, and EB conceptualized the study design. TH, ED, LF, and LS performed methodology; LS and MS performed software analysis; Validation was done by LS. Formal analysis was done by TH and LS. TH, ED, LF, and LS investigated the study. EB, EG, ES, and PN gathered the resources. Data curation was done by TH and LS; Writing of the original draft was done by TH. Writing, review, and editing were done by LF, ED, EG, EB, ES, PN, OG, BAG, MS, and LS. LS Supervised the study. Project administration was done by TH, LF, ED, EG, EB, ES, and PN.

Conflicts of Interest

None declared.

References

1. Xie JYX, Frost R, Meakin R. Not quite a doctor, but should I help? A qualitative exploration of medical students' attitudes towards responding to medical emergencies that occur in the public domain. BMJ Open 2019;9(4):e028035 [FREE Full text] [doi: 10.1136/bmjopen-2018-028035] [Medline: 30962243]
2. Klasen JM, Bingisser R, Meienberg A, Bogie B. Harnessing unique experiences to build competence: medical student engagement in frontline care during the COVID-19 pandemic. Swiss Med Wkly 2021;151:w20480 [FREE Full text] [doi: 10.4414/smw.2021.20480] [Medline: 33641110]
3. Seligman WH, Ganatra S, England D, Black JJM. Initial experience in setting up a medical student first responder scheme in South Central England. Emerg Med J 2016;33(2):155-158. [doi: 10.1136/emermed-2015-204638] [Medline: 26253147]

4. Freund Y, Duchateau FX, Baker EC, Goulet H, Carreira S, Schmidt M, et al. Self-perception of knowledge and confidence in performing basic life support among medical students. Eur J Emerg Med 2013;20(3):193-196. [doi: 10.1097/MEJ.0b013e328355fd59] [Medline: 22735507]
5. Willmore RD, Veljanoski D, Ozdes F, Stephens B, Mooney J, Crumley SG, et al. Do medical students studying in the United Kingdom have an adequate factual knowledge of basic life support? World J Emerg Med 2019;10(2):75-80 [FREE Full text] [doi: 10.5847/wjem.j.1920-8642.2019.02.002] [Medline: 30687442]
6. Almesned A, Almeman A, Alakhtar AM, AlAboudi AA, Alotaibi AZ, Al-Ghasham YA, et al. Basic life support knowledge of healthcare students and professionals in the Qassim University. Int J Health Sci (Qassim) 2014;8(2):141-150 [FREE Full text] [doi: 10.12816/0006080] [Medline: 25246881]
7. Baldi E, Contri E, Bailoni A, Rendic K, Turcan V, Donchev N, et al. Final-year medical students' knowledge of cardiac arrest and CPR: we must do more!. Int J Cardiol 2019;296:76-80. [doi: 10.1016/j.ijcard.2019.07.016] [Medline: 31375334]
8. Grześkowiak M. The effects of teaching basic cardiopulmonary resuscitation: a comparison between first and sixth year medical students. Resuscitation 2006;68(3):391-397. [doi: 10.1016/j.resuscitation.2005.07.017] [Medline: 16457932]
9. Kiyan S, Yanturali S, Musal B, Gursel Y, Aksay E, Turkcuer I. Determination of advanced life support knowledge level of residents in a Turkish university hospital. J Emerg Med 2008;35(2):213-222. [doi: 10.1016/j.jemermed.2007.02.049] [Medline: 17976765]
10. Boonmak P, Boonmak S, Chongarunngamsang W, Maharungruengrat K. Advanced cardiac life support knowledge among medical staff and residents in university hospital. Srinagarind Med J 2009;24(4):296-301 [FREE Full text]
11. Niemi-Murola L, Mäkinen M, Castren M, ECCE Study Group. Medical and nursing students' attitudes toward cardiopulmonary resuscitation and current practice guidelines. Resuscitation 2007;72(2):257-263. [doi: 10.1016/j.resuscitation.2006.07.006] [Medline: 17126983]
12. Dias DL, Kamat AS, Gomes SS, Gomes EJ, Bhounsule SA. Quantitative analysis of competency levels in medical interns of a tertiary care hospital in India: a questionnaire based cross sectional study. J Assoc Physicians India 2019;67(9):34-38 [FREE Full text] [Medline: 31561686]
13. Health - Employment - Social security: Health: CC 810.30 Federal Act of 30 September 2011 on research involving human beings (Human Research Act, HRA). The Federal Assembly of the Swiss Confederation. 2020. URL: https://www.fedlex.admin.ch/eli/cc/2013/617/en#cha [accessed 2023-05-30]
14. Eysenbach G. Improving the quality of web surveys: the checklist for reporting results of internet e-surveys (CHERRIES). J Med Internet Res 2004;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
15. Perkins GD, Handley AJ, Koster RW, Castrén M, Smyth MA, Olasveengen T, Adult Basic Life Support and Automated External Defibrillation Section Collaborators. European Resuscitation Council Guidelines for Resuscitation 2015: Section 2. Adult basic life support and automated external defibrillation. Resuscitation 2015 Oct;95:81-99. [doi: 10.1016/j.resuscitation.2015.07.015] [Medline: 26477420]
16. Suppan L, Fehlmann CA, Stuby L, Suppan M. The importance of acknowledging an intermediate category of airway management devices in the prehospital setting. Healthcare (Basel) 2022;10(5):961 [FREE Full text] [doi: 10.3390/healthcare10050961] [Medline: 35628096]
17. Stuby L, Jampen L, Sierro J, Bergeron M, Paus E, Spichiger T, et al. Effect of early supraglottic airway device insertion on chest compression fraction during simulated out-of-hospital cardiac arrest: randomised controlled trial. J Clin Med 2021;11(1):217 [FREE Full text] [doi: 10.3390/jcm11010217] [Medline: 35011958]
18. Chan M, Fehlmann CA, Pasquier M, Suppan L, Savoldelli GL. Endotracheal intubation success rate in an urban, supervised, resident-staffed emergency mobile system: an 11-year retrospective cohort study. J Clin Med 2020;9(1):238 [FREE Full text] [doi: 10.3390/jcm9010238] [Medline: 31963162]
19. Vincent A, Semmer NK, Becker C, Beck K, Tschan F, Bobst C, et al. Does stress influence the performance of cardiopulmonary resuscitation? A narrative review of the literature. J Crit Care 2021;63:223-230 [FREE Full text] [doi: 10.1016/j.jcrc.2020.09.020] [Medline: 33046274]
20. Anderson KL, Niknam K, Laufman L, Sebok-Syer SS, Andrabi S. Multi-community cardiopulmonary resuscitation education by medical students. Cureus 2020;12(6):e8647 [FREE Full text] [doi: 10.7759/cureus.8647] [Medline: 32685315]
21. Taramarcaz V, Herren T, Golay E, Regard S, Martin-Achard S, Mach F, et al. A short intervention and an interactive e-learning module to motivate medical and dental students to enlist as first responders: implementation study. J Med Internet Res 2022;24(5):e38508 [FREE Full text] [doi: 10.2196/38508] [Medline: 35583927]
22. Larribau R, Deham H, Niquille M, Sarasin FP. Improvement of out-of-hospital cardiac arrest survival rate after implementation of the 2010 resuscitation guidelines. PLoS One 2018;13(9):e0204169 [FREE Full text] [doi: 10.1371/journal.pone.0204169] [Medline: 30248116]
23. Moretti MA, de Oliveira Camboim A, Ferrandez CA, Ramos IC, Costa IB, Canonaco JS, et al. Retention of cardiopulmonary resuscitation skills in medical students. Arq Bras Cardiol 2021;117(5):1030-1035 [FREE Full text] [doi: 10.36660/abc.20200546] [Medline: 34406321]
24. Andersen LW, Holmberg MJ, Berg KM, Donnino MW, Granfeldt A. In-hospital cardiac arrest: a review. JAMA 2019;321(12):1200-1210 [FREE Full text] [doi: 10.1001/jama.2019.1696] [Medline: 30912843]

25. Semeraro F, Signore L, Cerchiari EL. Retention of CPR performance in anaesthetists. Resuscitation 2006;68(1):101-108. [doi: 10.1016/j.resuscitation.2005.06.011] [Medline: 16325986]
26. Greif R, Lockey A, Breckwoldt J, Carmona F, Conaghan P, Kuzovlev A, et al. European resuscitation council guidelines 2021: education for resuscitation. Resuscitation 2021;161:388-407. [doi: 10.1016/j.resuscitation.2021.02.016] [Medline: 33773831]
27. van Gaalen AEJ, Brouwer J, Schönrock-Adema J, Bouwkamp-Timmer T, Jaarsma ADC, Georgiadis JR. Gamification of health professions education: a systematic review. Adv Health Sci Educ Theory Pract 2021;26(2):683-711 [FREE Full text] [doi: 10.1007/s10459-020-10000-3] [Medline: 33128662]
28. Sardi L, Idri A, Fernández-Alemán JL. A systematic review of gamification in e-Health. J Biomed Inform 2017 Jul;71:31-48 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.011] [Medline: 28536062]
29. Wheeler S. E-learning and digital learning. In: Seel NM, editor. Encyclopedia of the Sciences of Learning. New York, NY: Springer; 2011:1109-1111.
30. Suppan M, Stuby L, Carrera E, Cottet P, Koka A, Assal F, et al. Asynchronous distance learning of the national institutes of health stroke scale during the COVID-19 pandemic (e-learning vs video): randomized controlled trial. J Med Internet Res 2021;23(1):e23594 [FREE Full text] [doi: 10.2196/23594] [Medline: 33428581]
31. Lehmann R, Thiessen C, Frick B, Bosse HM, Nikendei C, Hoffmann GF, et al. Improving pediatric basic life support performance through blended learning with web-based virtual patients: randomized controlled trial. J Med Internet Res 2015;17(7):e162 [FREE Full text] [doi: 10.2196/jmir.4141] [Medline: 26139388]
32. Liu Q, Peng W, Zhang F, Hu R, Li Y, Yan W. The effectiveness of blended learning in health professions: systematic review and meta-analysis. J Med Internet Res 2016;18(1):e2 [FREE Full text] [doi: 10.2196/jmir.4807] [Medline: 26729058]
33. Vallée A, Blacher J, Cariou A, Sorbets E. Blended learning compared to traditional learning in medical education: systematic review and meta-analysis. J Med Internet Res 2020;22(8):e16504 [FREE Full text] [doi: 10.2196/16504] [Medline: 32773378]
34. Yeung J, Perkins G, Davies R, Bullock I, Lockey A, Gwinnutt C, et al. Introducing non-technical skills teaching to the resuscitation council (UK) advanced life support course. Resuscitation 2014;85:S71. [doi: 10.1016/j.resuscitation.2014.03.178]
35. Ricci S, Calandrino A, Borgonovo G, Chirico M, Casadio M. Viewpoint: virtual and augmented reality in basic and advanced life support training. JMIR Serious Games 2022;10(1):e28595 [FREE Full text] [doi: 10.2196/28595] [Medline: 35319477]
36. Ajzen I. From intentions to actions: a theory of planned behavior. In: Kuhl J, Beckmann J, editors. Action Control: From Cognition to Behavior. Berlin, Heidelberg: Springer; 1985:11-39.
37. Regard S, Rosa D, Suppan M, Giangaspero C, Larribau R, Niquille M, et al. Evolution of bystander intention to perform resuscitation since last training: web-based survey. JMIR Form Res 2020;4(11):e24798 [FREE Full text] [doi: 10.2196/24798] [Medline: 33252342]

Abbreviations

ACLS: advanced cardiovascular life support
AED: automatic external defibrillator
BLS: basic life support
CA: cardiac arrest
CHERRIES: Checklist for Reporting Results of Internet e-Surveys
e-learning: electronic learning
OHCA: out-of-hospital cardiac arrest
SMUR: Service Mobile d'Urgence et de Réanimation
UGFM: University of Geneva Faculty of Medicine

Edited by T de Azevedo Cardoso; submitted 29.01.23; peer-reviewed by A Calandrino, A Missel; comments to author 03.03.23; revised version received 11.03.23; accepted 25.05.23; published 27.06.23.

Please cite as:
Herren T, Fivaz L, Dufeil E, Golay E, Braun E, Straub E, Nidegger P, Grosgurin O, Gartner BA, Suppan M, Suppan L
Development of a Pilot Introductory Advanced Cardiovascular Resuscitation Course for Senior Medical Students in Switzerland: Student-Driven Implementation Study
Interact J Med Res 2023;12:e46075
URL: https://www.i-jmr.org/2023/1/e46075
doi:10.2196/46075
PMID:37231610

Viewpoint

Bringing the Pediatric Endocrine Spanish Speaking Community Together: First Virtual Pediatric Endocrine Meeting in Low- and Middle-Income Countries in Central and South America

Roberto Bogarin[1*], MD; Luis Elizondo[2*], MD; Evangelia Kalaitzoglou[3*], MD; Jadranka Popovic[4*], MD; Alan Rogol[5*], MD; Erick Richmond[1*], MD; Jean-Pierre Chanoine[6*], MD; Jose M Lopez-Pedrosa[7], PhD; Francis Ruiz Salazar[8*], MD; Patricia Vuguin[9*], MD, MSc

[1]Costa Rica National Children´s Hospital, San Jose, Costa Rica
[2]Hospital Calderon Guardia, San Jose, Costa Rica
[3]University of Kentucky, Lexington, KY, United States
[4]Pediatric Alliance Allegheny Health Network, Pittsburgh, PA, United States
[5]University of Virginia, Charlottesville, VA, United States
[6]University of British Columbia, Vancouver, BC, Canada
[7]Abbott Nutrition Research and Development, Abbott Laboratories, Granada, Spain
[8]Hospital Mexico, San Jose, Costa Rica
[9]Columbia University, New York, NY, United States
[*]these authors contributed equally

Corresponding Author:
Patricia Vuguin, MD, MSc
Columbia University
3959 Broadway
New York, NY, 10032
United States
Phone: 1 212 305 8999
Email: pv2267@cumc.columbia.edu

Abstract

Background: Pediatric endocrinology is a specialty that is struggling worldwide to maintain adequately trained professionals. Pediatric endocrine care in Central America and Caribbean countries is often performed by pediatricians or adult endocrinologists due to the limited number of pediatric endocrinologists. These health care providers are seldom members of endocrine societies and frequently lack formal training in the field.

Objective: In this study, we describe the scope of a virtual conference in pediatric endocrinology and diabetes targeted to low- and middle-income countries to provide equal opportunities for access to medical education for health care professionals.

Methods: The virtual conference was sponsored by the Pediatric Endocrine Society (North America), Asociación Costarricense de Endocrinología (previously, Asociación Nacional Pro Estudio de la Diabetes, Endocrinología y Metabolismo), and Asociacion Centroamericana y del Caribe de Endocrinologia Pediátrica. The conference was free to participants and comprised 23 sessions that were either synchronous with ability for real-time interactive sessions or asynchronous sessions, where content was available online to access at their convenience. Topics included idiopathic short stature, polycystic ovarian syndrome, diabetes mellitus, telemedicine, Turner syndrome, congenital adrenal hyperplasia, obesity, central precocious puberty, and subclinical hypothyroidism. The participants were asked to evaluate the conference after its completion with a questionnaire.

Results: A total of 8 speakers from Spain, Canada, Costa Rica, and the United States delivered the virtual event to 668 health care professionals from Guatemala, Venezuela, Dominican Republic, Costa Rica, Ecuador, Peru, Uruguay, Mexico, Honduras, Argentina, the United States, Bolivia, Chile, Panama, El Salvador, Nicaragua, Paraguay, Belize, Spain, and Colombia. Name, profession, and country were fully disclosed by 410 (61.4%) of the 668 health care professionals. The profession or level of training of participants were as follows: pediatric endocrinologists (n=129, 19.3%), pediatricians (n=116, 17.4%), general practitioners (n=77, 11.5%), adult endocrinologists (n=34, 5.1%), medical students (n=23, 3.4%), residents in various specialties (n=14, 2.1%), and others (n=17, 2.6%). A total of 23 sessions were offered, most of which were bilingual (Spanish and English).

Feedback from the evaluation questionnaire indicated that the content of the conference was very relevant to the participants' professional practice. Additionally, the participants reported that they were very satisfied with the organization, the web-based platform, and the sessions of the conference.

Conclusions: Lack of accessibility to the latest and cutting-edge medical education in pediatric endocrinology and diabetes for medical professionals from low- and middle-income countries can be overcome with a virtual conference. Online availability, low cost, and easy-to-use technology were well received from the participants, who were overall very satisfied by the quality and the relevance of the sessions to their professional practice.

(Interact J Med Res 2023;12:e41353) doi:10.2196/41353

KEYWORDS

continuing medical education; continuing education; medical education; professional development; pediatric; child; endocrinology; endocrine; pediatric endocrinology; diabetes; low- and middle-income countries; Latin America; Spanish; virtual; resources; digital

Introduction

Pediatric endocrinology is a field that is faced with increased demand, partly due to an increase in the burden of obesity and diabetes [1]. Interestingly, one-fifth of individuals with type 1 diabetes are in low- and lower-middle–income countries [2]. Although there is a high demand for pediatric endocrinologists to meet clinical care needs, even in developed countries, there is concern about the future of the pediatric endocrinology workforce due to a diminished interest in the specialty [3]. Furthermore, global inequality in pediatric endocrine care has been well recognized and may be in part due to the lack of access to a formal clinical training in pediatric endocrinology in developing countries [4].

An initiative to address these disparities included a free and globally accessible multilingual e-learning website for pediatric endocrinology and diabetes [5] supported by the European Society for Paediatric Endocrinology [6]. The website provides health care professionals access to topics in pediatric endocrinology and diabetes aimed to improve their clinical skills and competencies through interactive learning. In addition to self-directed learning, it can serve as a valuable resource that can facilitate classroom teaching and promote interaction with experts in the field [7]. A section of the website, targeted toward health care professionals from resource-limited settings, is available in 5 languages, including Spanish [6]. Another initiative to overcome the global inequality to access to medical knowledge and training has been the creation of a Paediatric Endocrinology Training Center for Africa, which resulted in 54 fellows from 12 countries being trained in pediatric endocrinology and diabetes [8]. Additional initiatives have been implemented in Sudan to improve pediatric endocrinology services with the establishment of a local pediatric endocrinology training program [9].

Pediatric endocrine care in Central America and the Caribbean countries is often provided by general pediatricians or adult endocrinologists due to lack of trained pediatric endocrinologists. General pediatricians are rarely Pediatric Endocrine Society (PES) members or members of other local pediatric endocrine organizations. Additionally, many of them are not able to participate in such conferences due to financial limitations. Specifically, as outlined by Pulido et al [10], medical education in Latin America has been characterized by marked differences due to variable socioeconomic and cultural factors. Moreover, the countries with the lowest density of human resources in health in the Americas are all in Central America (Guatemala and Honduras), South America (Bolivia and Guyana), and the Caribbean (Haiti) [11]. Previous conferences with primary focus on pediatric endocrine care were organized by the International Relations Committee from the PES and were held in person, one in Costa Rica in 2014 and one in the Dominican Republic in 2019. The conference in Costa Rica was partly supported by PES and attracted more than 50 attendees from Central America with focus on the educational needs of the local pediatricians and pediatric endocrinologists. In 2019, a second conference was organized by PES and Sociedad Domincana de Pediatria Endocrinologica, which was well attended (150 attendees from the Dominican Republic and other Central American countries).

Due to challenges associated with the COVID-19 pandemic, holding another in-person conference was not feasible. Instead, a virtual conference was held from August 31 to September 21, 2020 (Multimedia Appendices 1 and 2). The goals of this virtual meeting were to generate a new approach to encourage interactions between clinicians interested in pediatric endocrinology and diabetes and to improve medical knowledge in this field. This initiative was developed by the PES and local organizations to support a group of underrepresented and diverse health care professionals (clinicians, researchers, and educators) from Spanish and non–Spanish speaking countries with the goal to educate and promote health and eventually, reduce pediatric endocrine health inequalities in Central America and in the Caribbean. This outreach program has provided education to local providers focused on endocrine-related conditions in underserved areas.

Methods

Overview

A 3-week conference (1er Congreso Virtual Asociación Centroamericana de Endocrino Pediátrica), sponsored by the Pediatric Endocrine Society (North America), Asociación Costarricense de Endocrinología (previously Asociación Nacional Pro Estudio de la Diabetes, Endocrinología y Metabolismo) and Asociación Centroamericana y del Caribe de Endocrinología Pediátrica—the Central American and

Caribbean Association of Endocrinology, Diabetes and Metabolism—took place from August 31 to September 21, 2020. The speakers were pediatric endocrinologists from the United States, Canada, Spain, and Costa Rica, and most were members of the PES. During the conference, which consisted of a hybrid model of synchronous and asynchronous sessions, a total of 23 sessions were delivered to a diverse audience. Topics included idiopathic short stature, polycystic ovarian syndrome, diabetes, telemedicine, Turner syndrome, congenital adrenal hyperplasia, obesity, precocious puberty, and subclinical hypothyroidism (Table 1).

Table 1. Agenda of the conference, as well as synchropnous (S) or asynchronous (A) format.

Number	Session title	S or A
1	Catch up growth in stunted children	A
2	Idiopathic short stature	A
3	Practical Implementation of telemedicine	S
4	Subclinical hypothyroidism	A
5	Catch up growth in stunted children: Meet the expert	S
6	What should I do with an obese patient?	S
7	Idiopathis short stature: Meet the expert	S
8	Subclinical Hypothyroidism: Meet the expert	S
9	Turner Syndrome	A
10	Precocious puberty: Controversies in management	A
11	Growth Hormone indications	S
12	Congenital Adrenal Hyperplasia	A
13	What would I do with a patient with PCOS?	S
14	Turner Syndrome: Meet the expert	S
15	Congenital Adrenal Hyperplasia: Meet the expert	S
16	Challenges managing the adolescent with diabetes	A
17	Insulin pumps: Update/What's new	A
18	Liraglutide in Children and Adolescents with Type 2 diabetes	S
19	Pubertal delay: Approach and treatment	A
20	What would I do with a patient with gynecomastia?	S
21	How do I approach a patient with secondary amenorrhea?	S
22	Diabetes: Meet the expert	S
23	Pubertal delay: Meet the expert	S

Registration was free for participants and attendees registered online [12]. The participants had access to live sessions using GoTo Webinar, a landing page at endopediatrica's website [12], or a Moodle website to access the lectures either synchronously or asynchronously. The asynchronous sessions (Vimeo; Vimeo, Inc.) included content that was available online and could be accessed by the participants at their convenience. The designated speaker made a presentation in the asynchronous prerecorded sessions with a defined theme. The audience had the opportunity to interact with this speaker for a live or synchronous session 3-5 days later. These sessions were in "question & answer," "clinical cases," or "ask the expert" format. Live sessions were broadcasted using Facebook live (Meta Platforms, Inc.) and Vimeo with opportunities to participate in interactive sessions with Q&A format. Attendees obtained information and technological tools to access the meeting from multiple sources (Facebook, WhatsApp [Meta Platforms, Inc.], GoTo Webinar, Zoom [Zoom Video Communications], and Vimeo), which allowed for the participation of individuals beyond the Central America and Dominican Republic regions. At the end of the meeting, the participants were invited to complete an evaluation of the conference. Conference evaluations included questions about the relevance of the topics covered, registration process, satisfaction with technological platforms, access to the event, and the invited speakers' knowledge of the subjects and presentations.

Ethical Considerations

No ethics approval was applied for because the conference-reported data are anonymous or deidentified and therefore exempt from oversight. There was no process of data linkage and recording, and dissemination did not generate identifiable information. No monetary compensation was offered to the speakers or participants. The meeting received generous support from Novo Nordisk, Pfizer, and Merck Foundation to cover the expenses of the English-to-Spanish translation, the production of the platform, and the promotion of the event. Topics were chosen by the organizers based on the needs and interest of the audience. Novo Nordisk, Pfizer, or Merck did not provide topics for discussion.

Results

A total of 8 speakers from Spain, Canada, Costa Rica, and the United States delivered the virtual event to 668 health care professionals. The audience consisted mainly of native Spanish speakers, and Spanish was the preferred language at the conference. Full registration information was available for most of the participants, including their country of origin and their profession or level of training. The country with the most attendees was Guatemala (n=69, 10.3%), followed by Venezuela (n=66, 9.9%), and Dominican Republic (n=63, 9.4%; Table 2). Participants were from 20 countries (Guatemala, Venezuela, Dominican Republic, Costa Rica, Ecuador, Peru, Uruguay, Mexico, Honduras, Argentina, the United States, Bolivia, Chile, Panama, El Salvador, Nicaragua, Paraguay, Belize, Spain, and Colombia). The majority of the participants were pediatric endocrinologists, pediatricians, or general practitioners (Figure 1), with residents and medical students also participating.

Of the 23 sessions, 14 (61%) sessions were live (Table 1), and if presented by an English speaker, they had simultaneous Spanish translation. Using the GoTo Webinar platform, 360 records of access were documented for the live sessions. Meet-the-expert and live sessions, which consisted of questions and answers as well as interaction with the audience showed an average interest by the audience of 92%, and a total number of 168 questions were asked throughout the event. Access to the lectures post meeting was evaluated via the Vimeo platform during September 2020. Of the 2937 times a video was played, 1285 (43.8%) times the platform reported 100% completion of the video. Vimeo reported 1331 total hours of content playback.

The number of people who viewed a prerecorded session (Vimeo) on a particular day ranged from 41 to 253.

At the end of the meeting, 408 attendees evaluated the conference and provided feedback in the form of a questionnaire. Questions with participant responses are presented in Figures 2 and 3. All participants reported that the conference was highly relevant to their professional practice. Additionally, using a scale of 1 to 5 (5 being extremely satisfied and 1 not satisfied at all), the majority of the participants were extremely satisfied (n=320, 78%) or very satisfied (n=83, 20%) with the event. The majority of the participants were extremely satisfied (n=236, 58%) or very satisfied (n=134, 33%) with the platforms used for the live sessions. Approximately 30% (n=122) of the attendees suggested improvements to the technological platform and onboarding of the event. Suggested improvements included schedule improvement, onboarding of the platform, times, adding more topics, increasing time for questions, and adding clinical cases. The attendees reported feeling that the event provided a novel way to disseminate medical knowledge within clinical endocrinology. Below are some comments provided by the attendees:

> *Very good initiative.*
>
> *Thank you for continuing distance learning.*
>
> *The truth is that everything turned out very well, at first it was very difficult with the use of technology, but it improved a lot over the days.*
>
> *I loved the topics and the didactics of the professionals. They were very practical and of great relevance for my daily practice in the clinic. Thanks a lot!*
>
> *I am very grateful to the entire organization because it gave me the opportunity to listen to true references in pediatric endocrinology and in turn in such a way that it will be very useful for my professional practice, which at the moment is in general medicine.*

Table 2. Origin country of attendees. The number of attendees per country and the percent per country is presented. Information available for 668 attendees.

Country	Attendees, n (%)
Guatemala	69 (10.3)
Venezuela	66 (9.9)
Dominican Republic	63 (9.4)
Costa Rica	57 (8.5)
Ecuador	54 (8.1)
Peru	50 (7.5)
Uruguay	40 (6.0)
Mexico	36 (5.4)
Honduras	35 (5.2)
Argentina	33 (4.9)
USA	31 (4.6)
Bolivia	25 (3.7)
Chile	21 (3.1)
Countries with 17 or less attendees	88 (13.2)
Total	668 (100)

Figure 1. Profession and/or level of training of attendees. Information available for 410 attendees.

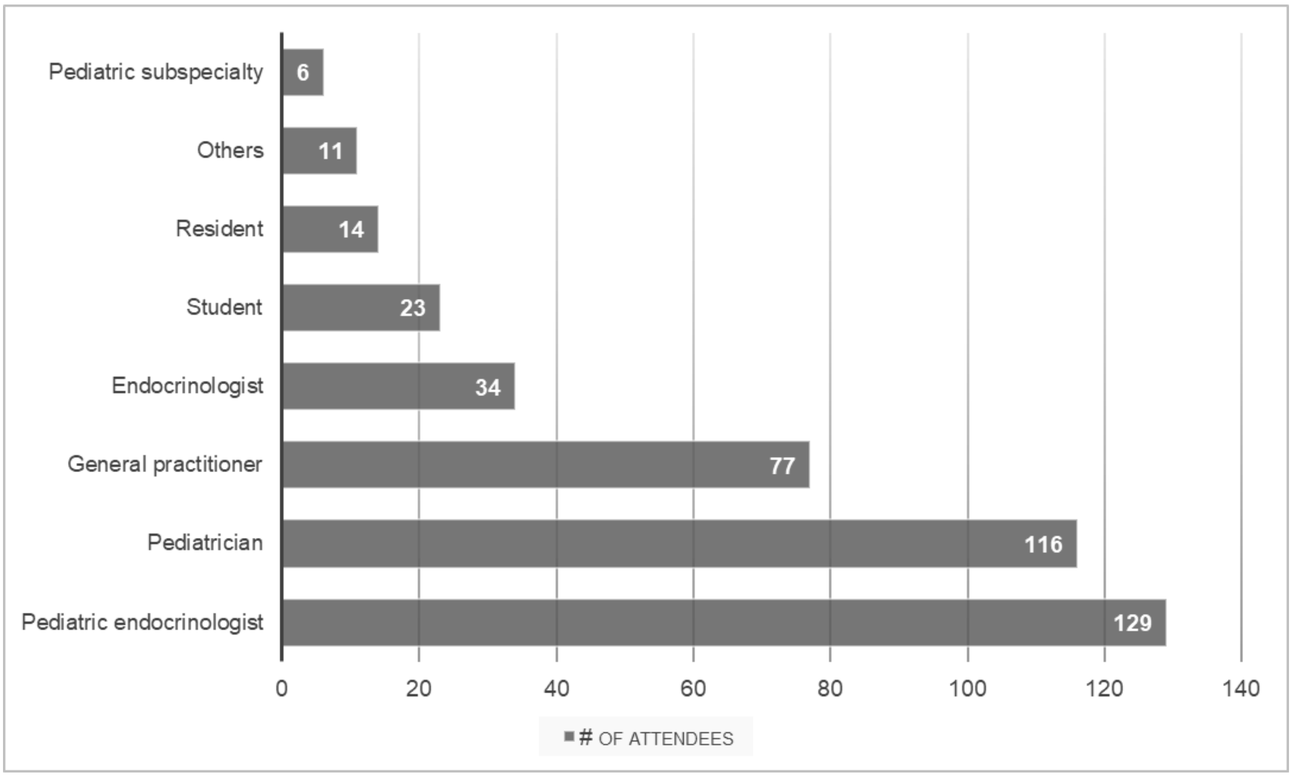


Figure 2. "How satisfied were you with the event?" Evaluation of the conference by the attendees. Information available for 408 attendees (1=not satisfied to 5=very satisfied).

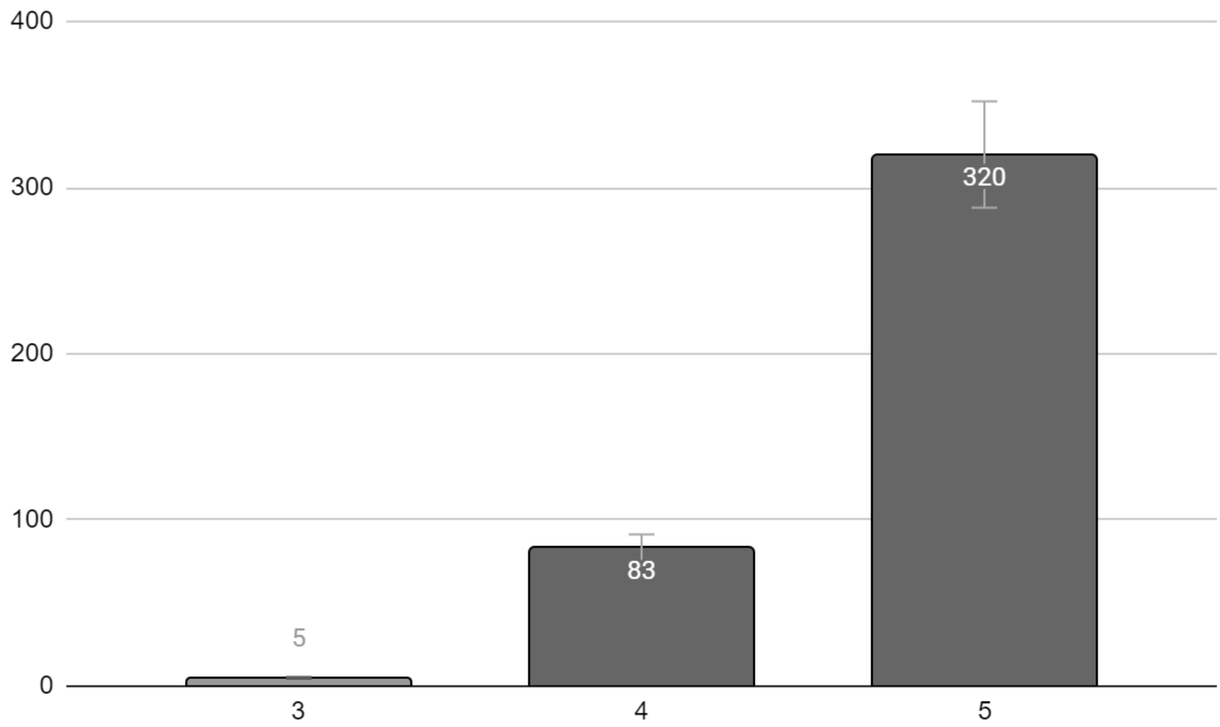
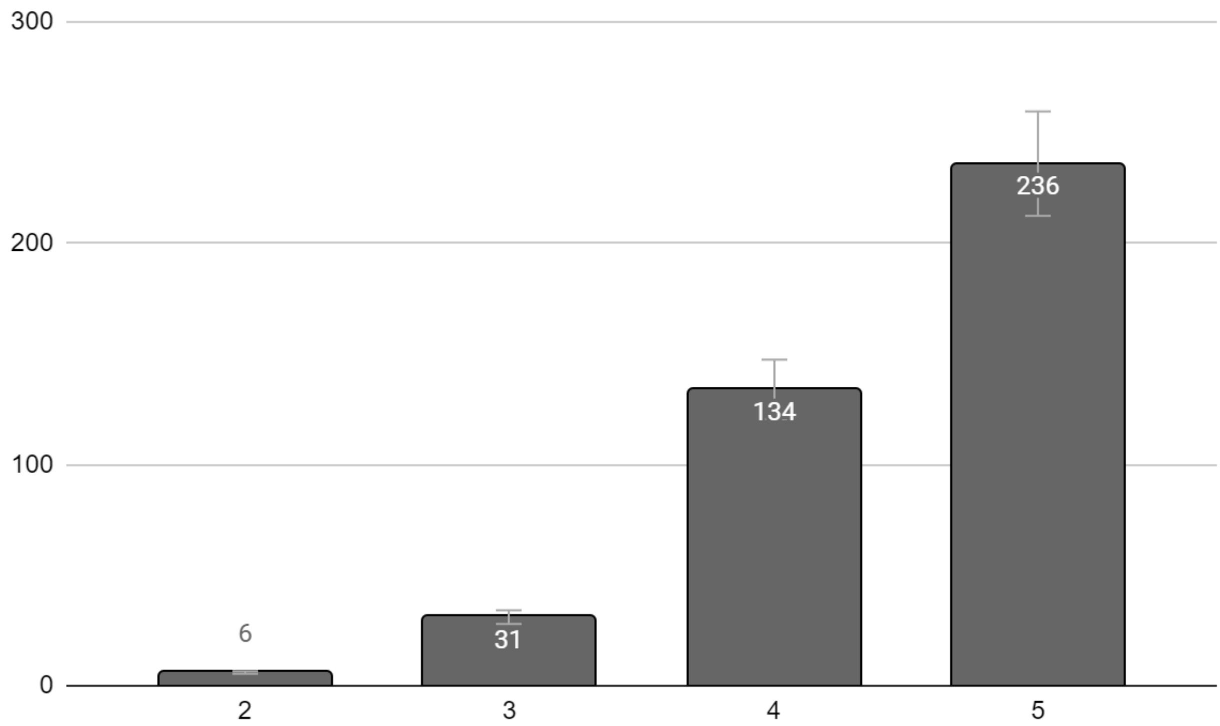


Figure 3. "How satisfied were you with the platforms used for the sessions (ie, Vimeo, GotoWebinar, and Facebook Live)?" Evaluation of the conference by the attendees. Information available for 408 attendees (1=not satisfied to 5=very satisfied).



Discussion

Principal Findings

Latin America has on average 2 doctors per 1000 population, and a low number of doctor consultations per capita [13]. Additionally, there is a lack of formally trained medical professionals in the field of pediatric endocrinology in several countries in Central and South America [14]. Some countries, such as Argentina, Brazil, and Mexico, have several trained medical professionals in pediatric endocrinology; however, other countries such as Nicaragua and Bolivia have a limited number of pediatric endocrinologists, available only in urban centers [15,16]. Previous conferences focused on pediatric endocrinology and diabetes were successfully held in Central America with the support of PES, as mentioned; however, participation was limited. As mentioned by Zacharin et al [14], access to up-to-date medical education for physicians and participation in medical conferences can be challenging for many medical professionals, trainees, and students from low-income countries where expertise and formal training is lacking [14]. Central and South America include several countries where there is a need for more formal training in pediatric endocrinology. Previous initiatives by the PES, International Society for Pediatric and Adolescent Diabetes, and the European Society for Paediatric Endocrinology have been welcomed and have provided opportunities for formal training with success [14]. However, most of these initiatives have required in-person participation, which can be particularly challenging for medical professionals from low-income countries.

Here, we present the first virtual pediatric endocrine conference in low- and middle-resource countries in Central and South America and the Caribbean with the support of the PES. There were several factors that contributed to the success of the conference. The virtual platform of the meeting resulted in a larger number of participants (668 participants) compared to the previous conferences in Costa Rica (close to 50 participants) and the Dominican Republic (close to 150 participants) and offered an opportunity for attendance to health care professionals from 20 countries. Over half of the participants were pediatricians and general practitioners who do not specialize in pediatric endocrinology, and who might have not had the opportunity to attend the conference if it was only offered in person. Additionally, most of the sessions were available in Spanish, the native language for most participants. Lastly, the conference covered a wide variety of topics related to pediatric endocrinology and diabetes (Table 1) and, according to the participants, these topics were highly relevant to their professional practice. The participants were very satisfied with the event (Figure 2) and the platforms used for the live sessions (Figure 3).

The recent challenges due to the COVID-19 pandemic have resulted in most large conferences being held virtually. Virtual conferences are more accessible, more inclusive, and more affordable, and offer opportunity to the attendees to access material at their convenience. They are also less time-consuming as they do not include travel [17]. Registration for this conference was free, and there were many sessions that allowed for direct interaction between the presenters and the audience in the form of Q&A session, which is not always available in virtual conferences. Additionally, the participants were mostly satisfied with the technological and organizational aspects of the event, providing useful feedback for future virtual conferences.

A limitation to reporting our results was that attendees used different platforms to access the academic content. The objective was to facilitate access, thus generating fewer barriers to accessing knowledge at the cost of making the analysis of access to data from multiple web-based platforms more complex.

Conclusions

Under this virtual conference, with the use of different virtual platforms, ease of access with free registration, and up-to-date technology, professionals who otherwise would not have attended international conferences were able to attend high-level scientific lectures on pediatric endocrinology and diabetes. The participants were satisfied with the conference content, which was relevant to their practice and well organized. The availability of low-cost and low–technical complexity access to digital platforms could promote the dissemination of medical education and is a complementary initiative to improve training opportunities for health care professionals from low-income and low- to middle-income countries in pediatric endocrinology and diabetes.

Acknowledgments

We would like to thank the sponsors for this conference: Pediatric Endocrine Society (North America), Asociación Costarricense de Endocrinología (ASCEND; previously, Asociación Nacional Pro Estudio de la Diabetes, Endocrinología y Metabolismo), and Asociación Centroamericana y del Caribe de Endocrinología Pediátrica (ACCEP—the Central American and Caribbean Association of Endocrinology, Diabetes and Metabolism)

Conflicts of Interest

None declared.

Multimedia Appendix 1
Agenda.
[PNG File , 1268 KB - ijmr_v12i1e41353_app1.png]

Multimedia Appendix 2
First Virtual Meeting.
[PNG File , 2472 KB - ijmr_v12i1e41353_app2.png]

References

1. Dabelea D, Mayer-Davis EJ, Saydah S, Imperatore G, Linder B, Divers J, SEARCH for Diabetes in Youth Study. Prevalence of type 1 and type 2 diabetes among children and adolescents from 2001 to 2009. JAMA 2014 May 07;311(17):1778-1786 [FREE Full text] [doi: 10.1001/jama.2014.3201] [Medline: 24794371]
2. Gregory GA, Robinson TIG, Linklater SE, Wang F, Colagiuri S, de Beaufort C, et al. Global incidence, prevalence, and mortality of type 1 diabetes in 2021 with projection to 2040: a modelling study. The Lancet Diabetes & Endocrinology 2022 Oct;10(10):741-760. [doi: 10.1016/s2213-8587(22)00218-2]
3. Allen DB, Aye T, Boney CM, Eugster EA, Misra M, Singer K, et al. Sustaining the Pediatric Endocrinology Workforce: Recommendations from the Pediatric Endocrine Society Workforce Task Force. J Pediatr 2021 Jun;233:4-7. [doi: 10.1016/j.jpeds.2020.10.063] [Medline: 33137317]
4. Savage MO, Cassorla FG, Gluckman PD, Grueters-Kieslich A, Raghupathy P, Silink M, International Societies for Paediatric Endocrinology. Global inequalities in paediatric endocrine practice: statement of minimal acceptable care. Statement from the international societies for paediatric endocrinology. Horm Res 2006 Mar 29;65(3):111-113. [doi: 10.1159/000091278] [Medline: 16462146]
5. ESPE e-learning. URL: https://www.espe-elearning.org [accessed 2023-04-11]
6. Kalaitzoglou E, Majaliwa E, Zacharin M, de Beaufort C, Chanoine J, van Wijngaard-DeVugt C, et al. Multilingual Global E-Learning Pediatric Endocrinology and Diabetes Curriculum for Front Line Health Care Providers in Resource-Limited Countries: Development Study. JMIR Form Res 2020 Nov 05;4(11):e18555 [FREE Full text] [doi: 10.2196/18555] [Medline: 33151156]
7. Ng SM, Kalaitzoglou E, Utari A, van Wijngaard-deVugt C, Donaldson M, Wolfsdorf JI, et al. 10 years' experience of a global and freely accessible e-Learning website for paediatric endocrinology and diabetes. Horm Res Paediatr 2022 Nov 08:1-10 [FREE Full text] [doi: 10.1159/000527984] [Medline: 36349753]
8. Odundo P, Ngwiri T, Otuoma O, Laigong P, Mukhwana R, Limbe MS, et al. The Impact and Successes of a Paediatric Endocrinology Fellowship Program in Africa. Int J Endocrinol 2016;2016:1560248 [FREE Full text] [doi: 10.1155/2016/1560248] [Medline: 26904118]
9. Abdullah M, Elhassan G. Establishing paediatric endocrinology services in a limited resource country: experience from Sudan. Sudan J Paed 2021:116-122. [doi: 10.24911/sjp.106-1610648000]
10. Pulido MA, Cravioto A, Pereda A, Rondón R, Pereira G. Changes, trends and challenges of medical education in Latin America. Med Teach 2006 Feb 03;28(1):24-29. [doi: 10.1080/01421590500441869] [Medline: 16627317]
11. Most countries in the Americas have sufficient health personnel but face challenges in distribution, migration and training. Pan American Health Organization. 2013. URL: https://tinyurl.com/2a3ewwty [accessed 2023-04-11]
12. Endopediatrica.com. URL: https://endopediatrica.com [accessed 2023-04-11]
13. OECD/The World Bank (2020). Health at a Glance: Latin America and the Caribbean 2020. Paris, France: OECD Publishing; 2020.
14. Zacharin M, Chanoine JP, Cassorla F, Brink S, Hanas R, Fideleff HL, l 70 Global Pediatric EndocrinologyDiabetes Active Members. Promoting excellence in the care of pediatric endocrine diseases in the developing world. Pediatrics 2013 Feb;131(2):e573-e578. [doi: 10.1542/peds.2012-0848] [Medline: 23339226]
15. Snouffer, E.W., Barbero, R. Type 1 diabetes paediatric care challenges in Bolivia. International Diabetes Federation. 2016. URL: https://www.idf.org/component/attachments/attachments.html?id=592&task=download [accessed 2023-04-11]
16. Executive Summary of Nicaragua Report. International Insulin Foundation. 2007. URL: http://www.access2insulin.org/executive-summary-of-nicaragua-report.html [accessed 2023-04-11]
17. Sarabipour S. Virtual conferences raise standards for accessibility and interactions. Elife 2020 Nov 04;9:9 [FREE Full text] [doi: 10.7554/eLife.62668] [Medline: 33143847]

Abbreviations

PES: Pediatric Endocrine Society

Edited by A Mavragani; submitted 22.07.22; peer-reviewed by V Podder, A Sikder, A Dedeilia; comments to author 27.01.23; revised version received 30.03.23; accepted 03.04.23; published 08.05.23.

Please cite as:
Bogarin R, Elizondo L, Kalaitzoglou E, Popovic J, Rogol A, Richmond E, Chanoine JP, Lopez-Pedrosa JM, Ruiz Salazar F, Vuguin P
Bringing the Pediatric Endocrine Spanish Speaking Community Together: First Virtual Pediatric Endocrine Meeting in Low- and Middle-Income Countries in Central and South America
Interact J Med Res 2023;12:e41353
URL: https://www.i-jmr.org/2023/1/e41353
doi:10.2196/41353
PMID:37155229

Original Paper

Electronic Phenotype for Advanced Chronic Kidney Disease in a Veteran Health Care System Clinical Database: Systems-Based Strategy for Model Development and Evaluation

Gajapathiraju Chamarthi[1], MD; Tatiana Orozco[2], PhD; Popy Shell[2], MPH; Devin Fu[2], MPH; Jennifer Hale-Gallardo[2], PhD; Huanguang Jia[2], PhD; Ashutosh M Shukla[1,2], MD

[1]Division of Nephrology, Hypertension and Transplantation, University of Florida, Gainesville, FL, United States

[2]Advanced Chronic Kidney Disease and Home Dialysis Program, North Florida/South Georgia Veteran Healthcare System, Gainesville, FL, United States

Corresponding Author:
Ashutosh M Shukla, MD
Advanced Chronic Kidney Disease and Home Dialysis Program
North Florida/South Georgia Veteran Healthcare System
1600 Archer Road
Gainesville, FL, 32610
United States
Phone: 1 352 273 8821
Fax: 1 352 392 5465
Email: Ashutosh.Shukla@medicine.ufl.edu

Abstract

Background: Identifying advanced (stages 4 and 5) chronic kidney disease (CKD) cohorts in clinical databases is complicated and often unreliable. Accurately identifying these patients can allow targeting this population for their specialized clinical and research needs.

Objective: This study was conducted as a system-based strategy to identify all prevalent Veterans with advanced CKD for subsequent enrollment in a clinical trial. We aimed to examine the prevalence and accuracy of conventionally used diagnosis codes and estimated glomerular filtration rate (eGFR)-based phenotypes for advanced CKD in an electronic health record (EHR) database. We sought to develop a pragmatic EHR phenotype capable of improving the real-time identification of advanced CKD cohorts in a regional Veterans health care system.

Methods: Using the Veterans Affairs Informatics and Computing Infrastructure services, we extracted the source cohort of Veterans with advanced CKD based on a combination of the latest eGFR value ≤30 $ml \cdot min^{-1} \cdot 1.73\ m^{-2}$ or existing International Classification of Diseases (ICD)-10 diagnosis codes for advanced CKD (N18.4 and N18.5) in the last 12 months. We estimated the prevalence of advanced CKD using various prior published EHR phenotypes (ie, advanced CKD diagnosis codes, using the latest single eGFR <30 $ml \cdot min^{-1} \cdot 1.73\ m^{-2}$, utilizing two eGFR values) and our operational EHR phenotypes of a high-, intermediate-, and low-risk advanced CKD cohort. We evaluated the accuracy of these phenotypes by examining the likelihood of a sustained reduction of eGFR <30 $ml \cdot min^{-1} \cdot 1.73\ m^{-2}$ over a 6-month follow-up period.

Results: Of the 133,756 active Veteran enrollees at North Florida/South Georgia Veterans Health System (NF/SG VHS), we identified a source cohort of 1759 Veterans with advanced nondialysis CKD. Among these, 1102 (62.9%) Veterans had diagnosis codes for advanced CKD; 1391(79.1%) had the index eGFR <30 $ml \cdot min^{-1} \cdot 1.73\ m^{-2}$; and 928 (52.7%), 480 (27.2%), and 315 (17.9%) Veterans had high-, intermediate-, and low-risk advanced CKD, respectively. The prevalence of advanced CKD among Veterans at NF/SG VHS varied between 1% and 1.5% depending on the EHR phenotype. At the 6-month follow-up, the probability of Veterans remaining in the advanced CKD stage was 65.3% in the group defined by the ICD-10 codes and 90% in the groups defined by eGFR values. Based on our phenotype, 94.2% of high-risk, 71% of intermediate-risk, and 16.1% of low-risk groups remained in the advanced CKD category.

Conclusions: While the prevalence of advanced CKD has limited variation between different EHR phenotypes, the accuracy can be improved by utilizing two eGFR values in a stratified manner. We report the development of a pragmatic EHR-based model to identify advanced CKD within a regional Veterans health care system in real time with a tiered approach that allows targeting the needs of the groups at risk of progression to end-stage kidney disease.

(Interact J Med Res 2023;12:e43384) doi:10.2196/43384

KEYWORDS

advanced chronic kidney disease; EHR phenotype; Veteran Health System; CKD cohort; kidney disease; chronic; clinical; database; data; diagnosis; risk; disease

Introduction

Advanced chronic kidney disease (CKD) progressing to end-stage kidney disease (ESKD) is a huge burden for the US health care system [1]. Patients with advanced CKD are at increased risk for adverse outcomes, including progression to ESKD and death. Prior studies show that providing pre-ESKD nephrology care and comprehensive pre-ESKD education improves clinical outcomes; reduces health care costs; and increases home dialysis, transplantation utilization, and patient survival [2-6]. Despite these positive outcomes, approximately 40% of patients with incident ESKD in the United States have either limited (less than 6 months) or no access to nephrology care before initiating dialysis and even fewer (<1%) receive kidney disease education services [7,8]. Accurately identifying the advanced (stages 4 and 5) CKD population at risk for ESKD can facilitate targeted needs assessment studies to improve pre-ESKD nephrology care and provide comprehensive pre-ESKD education for this high-risk population [9].

Clinically, CKD is diagnosed by sustained alterations in the structure or function of the kidney for more than 3 months with implications for health. The Kidney Disease: Improving Global Outcomes (KDIGO) Work Group recommends staging CKD based on cause, estimated glomerular filtration rate (eGFR), and albuminuria [10]. Unfortunately, the asymptomatic nature of CKD creates a lack of awareness for patients and providers alike [1,11]. Investigators conventionally use the International Classification of Diseases (ICD)-based diagnosis codes or electronic health record (EHR)-based phenotypes according to the eGFR to identify patients with CKD in clinical databases [12]. These phenotypes recommend using two eGFR values below 60 $ml \cdot min^{-1} \cdot 1.73\ m^{-2}$, obtained more than 90 days apart, to identify a population with CKD of stage 3 or higher in the databases [12]. However, similar guidance is not available to identify an advanced CKD population within clinical databases, and epidemiological investigations frequently use a single latest eGFR value while ascertaining the advanced CKD burden within the database [3,13,14]. Considering the variability in the frequency of measurement, pragmatic fluctuations in the serum creatinine value and concerns for intervening acute kidney injury (AKI) episodes can cause errors in classifying one's CKD stage [15]. Thus, there is a need to establish an optimal EHR-based method capable of identifying patients with advanced CKD within clinical databases in real time to improve kidney disease care and research.

Using the clinical database of the North Florida/ South Georgia (NF/SG) Veterans Health System (VHS), we sought to assess the burden of advanced CKD prevalence in real time using various EHR-recorded advanced CKD phenotypes within the Veterans Health Administration (VHA) [14,16]. We further examined the accuracy of different EHR phenotypes for advanced CKD by prospectively following the cohorts for 6 months and assessed the number of Veterans remaining in the advanced CKD stage after the initial classification. Furthermore, considering the lack of consensus on EHR phenotyping for identifying an advanced CKD cohort within clinical databases, we also sought to explore a new tiered pragmatic method for estimating the Veteran cohort with advanced CKD in real time.

Methods

Data Source and Cohort Selection

This study was conducted as a system-based strategy to identify all prevalent Veterans with advanced (stages 4 and 5) nondialysis CKD. The identified participants were then approached for enrollment in the Trial to Evaluate and Assess the effects of Comprehensive pre-ESKD education on Home dialysis among Veterans (TEACH-VET), which aims to assess the impact of a universal approach for comprehensive pre-ESKD education for all patients with advanced CKD on various clinical, patient-reported, and health services outcomes [17]. We used the Veterans Affairs (VA) Corporate Data Warehouse (CDW) and VA Informatics and Computing Infrastructure (VINCI) to identify the advanced CKD cohort. In brief, the VINCI services initially queried the VA CDW in April 2021 to identify all Veterans registered for service at NF/SG VHS during the 12 months prior to the data extraction (source cohort). The Veterans with an active laboratory value of creatinine were identified and their eGFR was calculated by applying the Modification of Diet in Renal Disease (MDRD) equation [18]. The use of the MDRD equation was determined by the then-prevalent method of eGFR estimation for the VINCI services. We then created a source cohort of Veterans with advanced CKD who either had the latest eGFR value ≤30 $ml \cdot min^{-1} \cdot 1.73m^{-2}$ (index eGFR) or an existing ICD-10 diagnosis code for advanced CKD (ICD-10 codes: N18.4 and N18.5) within the last 12 months (Figure 1). Patients on dialysis were excluded using the ICD-10 and Current Procedural Terminology (CPT) codes for dialysis (see Table S1 in Multimedia Appendix 1). The prevalence of advanced CKD was estimated in real time using various methods, including advanced CKD diagnosis codes or by eGFR phenotypes described in the literature (ie, by ICD-10 advanced CKD diagnosis codes, by using single [index] eGFR < 30 $ml \cdot min^{-1} \cdot 1.73\ m^{-2}$, and by using the two eGFR values 90 days apart with the index eGFR <30 $ml \cdot min^{-1} \cdot 1.73\ m^{-2}$ and 90-day prior eGFR < 60 $ml \cdot min^{-1} \cdot 1.73\ m^{-2}$) [14,16,19]. The cumulative prevalence of CKD was calculated by combining the data extracted over 6 months. Patient-level data included age, sex, race, ethnicity, religion, marital status, Veteran era, and residential zip codes used for defining the rurality by applying Rural-Urban Commuting Area codes. Statistical analyses were performed using R software version 4.0.4 (R Core Team, 2021) [20].

The source cohort (ie, April 2021 cohort) was divided into a high-, intermediate-, and low-risk advanced CKD cohort utilizing the latest (index) eGFR and 90-day prior eGFR and diagnostic codes (Table 1). Patients with both eGFR values below 30 $ml \cdot min^{-1} \cdot 1.73\ m^{-2}$ were considered to have a high risk of advanced CKD, whereas those with one of the two eGFR values less than 30 $ml \cdot min^{-1} \cdot 1.73\ m^{-2}$ but with the other value ≥30 but <60 $ml \cdot min^{-1} \cdot 1.73\ m^{-2}$ were considered to have an intermediate risk of having advanced CKD. The intermediate-risk cohort with an index eGFR below 30 $ml \cdot min^{-1} \cdot 1.73\ m^{-2}$ was further refined by excluding patients diagnosed with AKI within the 90 days prior to their latest eGFR values using ICD-10 codes. Veterans with both eGFR values ≥30 $ml \cdot min^{-1} \cdot 1.73\ m^{-2}$ but with diagnosis codes for advanced CKD were regarded as having a low risk of advanced CKD (Table 1). The source cohort was followed prospectively for 6 consecutive months until September 2021 using similar queries to examine the eGFR laboratory behavior of the patients with advanced CKD.

Figure 1. Selection of an advanced nondialysis chronic kidney disease (CKD) cohort at North Florida/South Georgia (NF/SG) Veterans Health System. CPT: Current Procedural Terminology; eGFR: estimated glomerular filtration rate; ESKD: end-stage kidney disease; ICD-10: International Classification of Diseases, Tenth Revision.

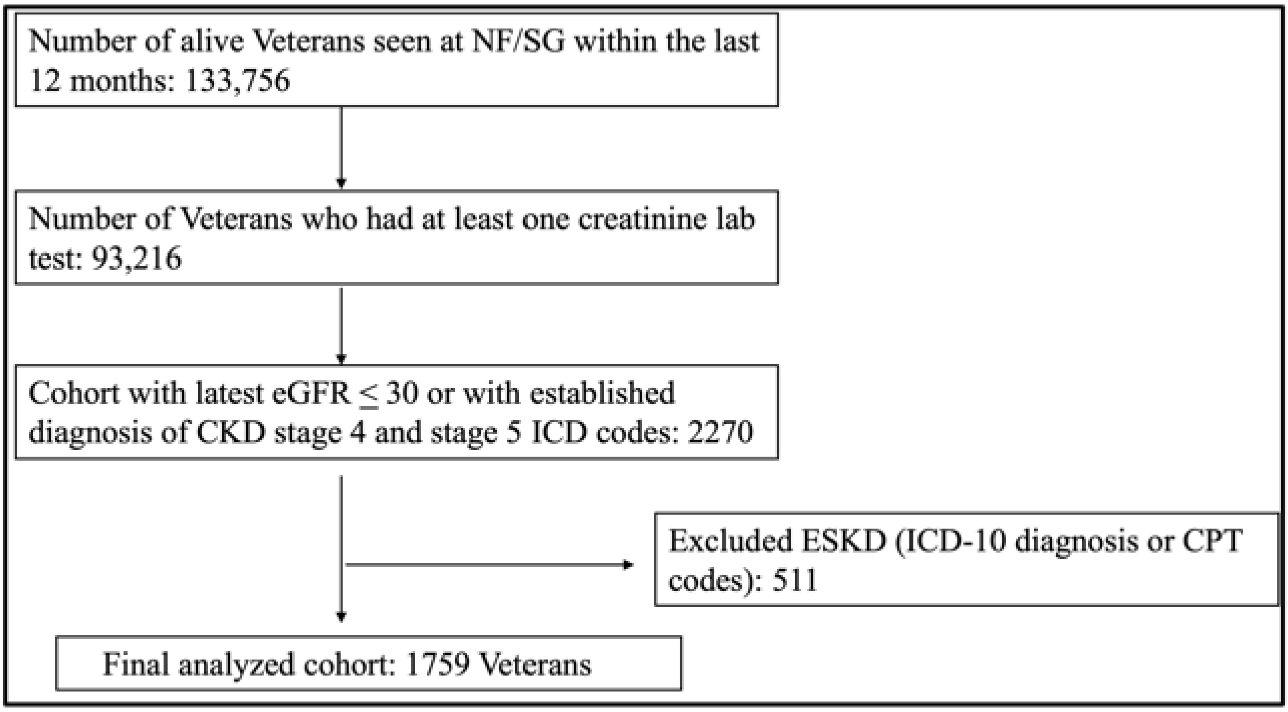


Table 1. Defining parameters for identifying cohorts at high, intermediate, and low risk of advanced chronic kidney disease (CKD).

Cohort	Index eGFR[a] ($ml \cdot min^{-1} \cdot m^{-2}$)	≥90 days prior eGFR ($ml \cdot min^{-1} \cdot m^{-2}$)	Additional criteria
High-risk advanced CKD	<30	< 30	None
Intermediate-risk advanced CKD			
Subgroup 1	<30	≥30 and <60	Excluding AKI[b] using ICD-10[c] codes (N17)
Subgroup 2	≥30 and <60	<30	Patients have ICD-10 codes for stage 4 and 5 CKD (N18.4 and N18.5)
Low-risk advanced CKD	≥30	≥30	Patients have ICD-10 codes for stage 4 and 5 CKD (N18.4 and N18.5)

[a]eGFR: estimated glomerular filtration rate; index eGFR refers to the latest eGFR at the time of extraction of the cohort.

[b]AKI: acute kidney injury.

[c]ICD-10: International Classification of Diseases, Tenth Revision.

Outcomes

The primary goal of this study was to assess the prevalence and accuracy of various EHR phenotypes for extraction of an advanced CKD cohort in a clinical database utilizing diagnosis codes and eGFR models (ie, by ICD-10 advanced CKD diagnosis codes, by using single latest [index] eGFR <30 $ml·min^{-1}·1.73\ m^{-2}$, and by using the two eGFR values 90 days apart, with the index eGFR <30 and 90 days prior eGFR <60) and our tiered EHR phenotype (high, intermediate, and low risk). Considering that nearly one-third of Veterans do not regularly obtain laboratory testing from within the VA, the denominator population for estimating the prevalence of advanced CKD was judged by only including the Veterans with a valid creatinine value measured over the prior 12 months. Considering EHR phenotypes as a standard for identification of patients with advanced CKD, cross-sectional accuracy for identifying patients with advanced CKD using only ICD-10 codes was assessed by comparison with laboratory-based eGFR EHR phenotypes, analyzed by calculating the sensitivity and positive predictive value (PPV). A manual chart review was conducted in a small randomly selected sample to identify errors related to automated advanced nondialysis CKD identification. Prospective accuracy of all EHR phenotypes, including our pragmatic tiered approach of high-, intermediate-, and low-risk advanced CKD cohorts, was assessed by ascertaining the longitudinal follow-up of laboratory values and identifying the likelihood of remaining in the advanced CKD stage at the end of the 6-month follow-up.

Ethical Approval

The regulatory approvals for the study were obtained from the institutional review board of the University of Florida (201900870). The study data are stored in secured systems at NF/SG VHS as per the institutional guidelines.

Results

We identified 133,756 active enrollees with 93,216 enrollees having at least one value of measured creatinine during an outpatient or inpatient visit at NF/SG VHS in the prior 12 months. After excluding the Veterans with ESKD by additional ICD and CPT codes, a source cohort of 1759 Veterans was identified as either having the latest eGFR ≤30 $ml·min^{-1}·1.73\ m^{-2}$ or an existing ICD-10 diagnosis code for advanced CKD (ICD-10 codes N18.4 and N18.5) within the last 12 months (Figure 1). The overall cohort had a mean age of 75 (SD 11.1) years and consisted of a predominantly male (95.8%) and white (67.8%) population. These Veterans lived approximately 126.3 (SD 229.5) miles from the nephrology service–providing VA center, with rural Veterans constituting a significant proportion (751/1759, 42.7%) of the cohort (Table 2). A manual chart review was performed on 116 records and 13 Veterans with ESKD were identified, yielding an 11.2% error rate for advanced nondialysis CKD identification.

Of the total cohort of 1759 Veterans, only 1102 (62.9%) had diagnosis codes for advanced CKD, whereas 1391 (79.1%) had the latest (index) eGFR <30 $ml·min^{-1}·1.73\ m^{-2}$. Incorporating two eGFR values where the latest eGFR was <30 $ml·min^{-1}·1.73\ m^{-2}$ and the 90-day prior eGFR was <60 $ml·min^{-1}·1.73\ m^{-2}$, we found 1346 Veterans to have advanced CKD. We then categorized 928 (52.7%) as high risk, 480 (27.2%) as intermediate risk, and 315 (17.9%) as low risk of advanced CKD based on the definitions described above (Tables 1 and 2). The mean eGFR for the initial advanced CKD cohort was 26.2 (SD 12.1) $ml·min^{-1}·1.73\ m^{-2}$. The mean eGFR was 27.7 $ml·min^{-1}·1.73\ m^{-2}$ in the ICD codes group, while the mean eGFR in the latest (index) eGFR <30 $ml·min^{-1}·1.73\ m^{-2}$ group was 22 $ml·min^{-1}·1.73\ m^{-2}$. The mean eGFR was 20.3 (SD 6.6), 27.4 (SD 5.6), and 42.1 (SD 16.6) $ml·min·1.73\ m^{-2}$ for the high-, intermediate-, and low-risk advanced CKD groups in the initial source cohort (Table 2). The prevalence of advanced CKD among Veterans at NF/SG VHS varied between 1% and 1.5% based on the phenotype for advanced CKD. Based on our definitions, the prevalence of advanced (high- and intermediate-risk) CKD at NF/SG VHS was approximately 1.5% (Table 3). The cumulative cohort over the 6 months yielded 1840 Veterans with high and intermediate risk (2% cumulative prevalence). The sensitivity of diagnosis codes was only 55%-65% compared to the eGFR phenotypes, and the PPV of ICD-10 diagnosis codes for advanced CKD varied between 55% and 74% (Table 4).

The source cohort was followed prospectively for 6 months to examine the variations and likelihood of a sustained reduced eGFR <30 $ml·min^{-1}·1.73\ m^{-2}$ across various EHR phenotypes. A total of 981 (55.8%) of the 1759 Veterans had at least one subsequent eGFR measurement in the initial April cohort (Table 5). The probability of any subsequent eGFR measurement above 30 $ml·min^{-1}·1.73\ m^{-2}$ after the index eGFR in the cohort defined by ICD codes was 38.3%, and was approximately 12.7% and 12.8 % in cohorts defined by index eGFR <30 $ml·min^{-1}·1.73\ m^{-2}$ and two eGFR phenotypes with index eGFR < 30 $ml·min^{-1}·1.73\ m^{-2}$ and 90-day prior eGFR < 60 $ml·min^{-1}·1.73\ m^{-2}$, respectively. Similarly, the probability of having any subsequent eGFR value above 30 $ml·min^{-1}·1.73\ m^{-2}$ after the index eGFR measurement was 7.1%, 35.7%, and 90% in the high-, intermediate-, and low-risk group, respectively. The probability of Veterans remaining in an advanced CKD stage (stages 4 and 5) noted by the recent eGFR <30 $ml·min^{-1}·1.73\ m^{-2}$ at the end of follow-up was 65.3% in the group identified by the ICD codes, whereas the probability improved to 90% in the group defined by single (index) eGFR <30 $ml·min^{-1}·1.73\ m^{-2}$ and the group defined by the index eGFR and 90-day prior eGFR method. Similarly, the probability of Veterans remaining in an advanced CKD stage at the end of the follow-up period was 94.2%, 71.0%, and 16.1% for high-, intermediate-, and low-risk groups, respectively (Figure 2, Table 5, and Table S2 in Multimedia Appendix 1).

Table 2. Demographic data for the source cohort.

Characteristics	Total cohort (N=1759)	ICD-10[a] code N18.4 or N18.5 (n=1102)	Index eGFR[b] <30 (n=1391)	Index eGFR <30 and 90 days prior eGFR <60	High-risk advanced CKD[c] (n=928)	Intermediate-risk advanced CKD (n=480)	Low-risk advanced CKD (n=315)
eGFR, mean (SD)	26.2 (12.1)	27.7 (13.7)	22.0 (6.5)	22.0 (6.5)	20.3 (6.6)	27.4 (5.6)	42.1 (16.6)
Age (years), mean (SD)	75.3 (11.1)	75.5 (10.9)	75.0 (11.0)	75.2 (10.8)	75.3 (11.0)	75.2 (10.3)	75.5 (12.2)
Sex (male), n (%)	1686 (95.8)	1057 (95.9)	1334 (95.9)	1290 (95.8)	889 (95.8)	461 (96.0)	301 (95.6)
Race, n (%)							
Black	386 (21.9)	232 (21.1)	311 (22.4)	299 (22.2)	216 (23.3)	94 (19.6)	69 (21.9)
White	1192 (67.8)	758 (68.8)	936 (67.3)	911 (67.7)	616 (66.4)	339 (70.6)	212 (67.3)
Other or unknown	181 (10.3)	112 (10.2)	144 (10.4)	136 (10.1)	96 (10.3)	47 (9.8)	34 (10.8)
Hispanic ethnicity, n (%)	35 (2.0)	23 (2.1)	31 (2.2)	30 (2.2)	23 (2.5)	9 (1.9)	2 (0.6)
Rural, n (%)	751 (42.7)	475 (43.1)	574 (41.3)	558 (41.5)	378 (40.7)	200 (41.7)	154 (48.9)
Married, n (%)	1087 (61.9)	661 (60.1)	875 (63.0)	846 (63.0)	574 (62.0)	311 (64.9)	179 (57.0)
Service era, n (%)							
Pre-Vietnam	372 (21.1)	244 (22.1)	287 (20.6)	282 (21.0)	207 (22.3)	88 (18.3)	72 (22.9)
Vietnam	1012 (57.5)	635 (57.6)	803 (57.7)	785 (58.3)	528 (56.9)	292 (60.8)	172 (54.6)
Post-Vietnam and other	375 (21.3)	223 (20.2)	301 (21.6)	279 (20.7)	193 (20.8)	100 (20.8)	71 (22.5)
Distance to VA[d] (station 573), mean (SD)	126.3 (229.5)	130.3 (249.0)	127.9 (231.0)	126.6 (225.0)	127.7 (221.2)	124.3 (231.2)	129.7 (258.0)

[a]eGFR: estimated glomerular filtration rate (ml·min^{-1}·m^{-2}).
[b]ICD-10: International Classification of Diseases, Tenth Revision.
[c]CKD: chronic kidney disease.
[d]VA: Veterans Affairs.

Table 3. Prevalence of advanced chronic kidney disease (CKD) based on different criteria.

Prevalence subpopulation definition	Users, n	VA[a] users with creatinine lab measurement within last 12 months (n=93,216), % (95% CI)	Total VA users (N=133,756), % (95% CI)
Total VA users with at least one creatinine measurement within the last 12 months	93,216	100.0 (100-100)	69.7 (69.4-69.9)
Veterans with ICD-10[b] code N18.4 or N18.5 within last 12 months	1102	1.2 (1.1-1.3)	0.8 (0.8-0.9)
Veterans with index eGFR[c] <30	1391	1.5 (1.4-1.6)	1.0 (1.0-1.1)
Veterans with index eGFR <30 and 90 days prior eGFR <60	1346	1.4 (1.4-1.5)	1.0 (1.0-1.1)
Veterans with high risk of advanced CKD	928	1.0 (0.9-1.1)	0.7 (0.6-0.7)
Veterans with high and intermediate risk of advanced CKD	1408	1.5 (1.4-1.6)	1.1 (1.0-1.1)
Cumulative prevalence of advanced CKD (6 months) based on high- and intermediate-risk groups	1840	2.0 (1.9-2.1)	1.4 (1.3-1.4)

[a]VA: Veterans Affairs.
[b]ICD-10: International Classification of Diseases, Tenth Revision.
[c]eGFR: estimated glomerular filtration rate (ml·min^{-1}·m^{-2}); index eGFR refers to the latest eGFR measurement at the time of extraction of the cohort.

Table 4. Diagnostic accuracy of International Classification of Diseases, Tenth Revision codes for advanced chronic kidney disease (CKD) compared to estimated glomerular filtration rate (eGFR)-based defining criteria.

Accuracy metric	Index eGFR[a] <30 (n=1371), point estimate (95% CI)[b]	Index eGFR <30 and 90 days prior eGFR <60 (n=1346), point estimate (95% CI)	High-risk advanced CKD (n=928), point estimate (95% CI)	High- and intermediate-risk advanced CKD (n=1408), point estimate (95% CI)
Sensitivity	0.55 (0.52-0.57)	0.55 (0.53-0.58)	0.65 (0.62-0.68)	0.57 (0.55-0.60)
Specificity	1.00 (1.00-1.00)	1.00 (1.00-1.00)	0.99 (0.99-1.00)	1.00 (1.00-1.00)
Positive predictive value	0.68 (0.66-0.71)	0.68 (0.65-0.71)	0.55 (0.52-0.58)	0.74 (0.71-0.76)
Negative predictive value	0.99 (0.99-0.99)	0.99 (0.99-0.99)	1.00 (1.00-1.00)	0.99 (0.99-0.99)

[a]Index eGFR refers to the latest eGFR measure ($ml \cdot min^{-1} \cdot m^{-2}$) at the time of extraction of the cohort.

[b]20 patients were excluded from this column because they were missing a prior eGFR value; the subsequent column criteria/definitions required two eGFR values, and thus patients without two eGFR values were excluded for consistency between column criteria/definitions.

Table 5. Probability of remaining in advanced chronic kidney disease stages (4 and 5) based on various electronic health record phenotypes at the 6-month follow-up.

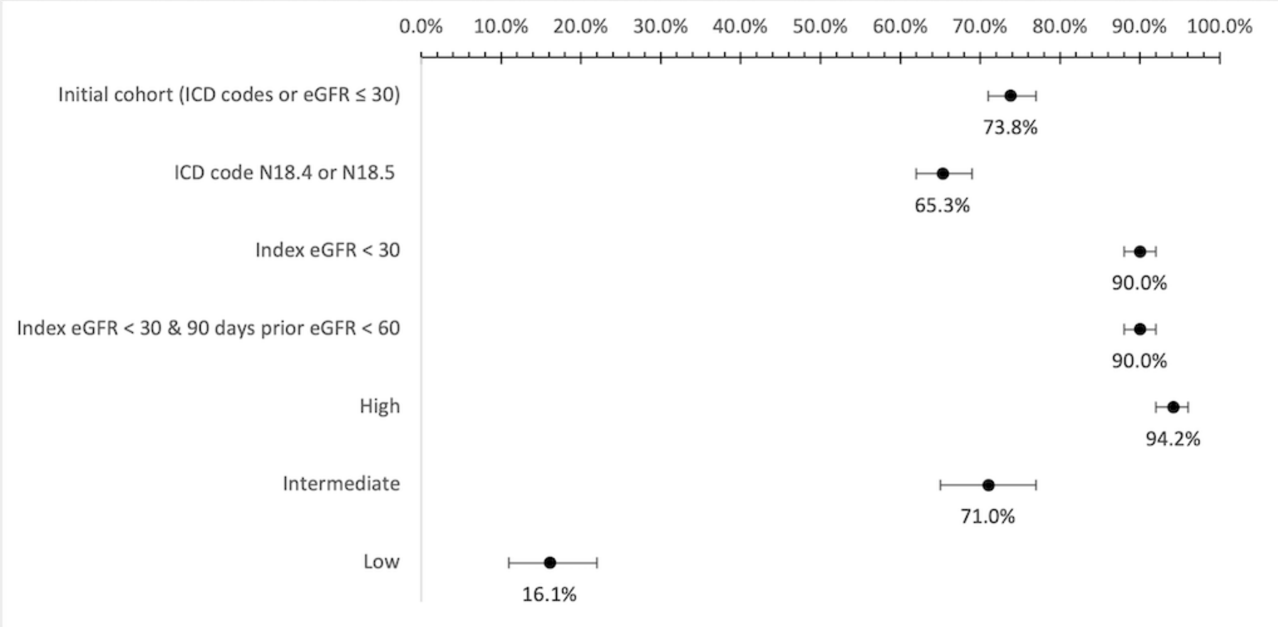
Characteristic	Initial cohort (ICD=10[a] codes or eGFR[b] ≤30) (N=1759)		ICD-10 code N18.4 or N18.5 (n=1102)		Index eGFR[c] <30 (n=1391)		Index eGFR <30 and 90 days prior eGFR <60 (n=1346)		High risk (n=928)		Intermediate risk (n=480)		Low risk (n=315)	
	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI	Value	95% CI
eGFR, mean (SD)	26.2 (12.1)	26-27	27.7 (13.7)	27-29	22.0 (6.5)	22-22	22.0 (6.5)	22-22	20.3 (6.6)	20-21	27.4 (5.6)	27-28	42.1 (16.6)	40- 44
Days between stage-defining eGFR values, mean (SD)	263.8 (394.3)	245-282	229.0 (304.9)	211-247	274.3 (399.8)	253-296	260.7 (341.7)	242-279	226.7 (159.3)	216-237	316.0 (526.0)	269-363	296.8 (594.2)	231-363
Subsequent eGFR measurement, n (%)	981 (55.8)	53%-58%	706 (64.1)	61%-67%	753 (54.1)	51%-57%	751 (55.8)	53%-58%	549 (59.2)	56%-62%	238 (49.6)	45%-54%	180 (57.1)	51%-63%
Any subsequent eGFR ≥30 after index eGFR, n (%)	293 (16.7)	15%-19%	271 (24.6)	22%-27%	96 (6.9)	5.7%-8.4%	96 (7.1)	5.8%-8.7%	39 (4.2)	3.0%-5.8%	85 (17.7)	14%-21%	162 (51.4)	46%-57%
Current eGFR ≥30 at 6-month follow-up, n (%)	257 (26.2)	23%-29%	245 (34.7)	31%-38%	75 (10.0)	8.0%-12.0%	75 (10.0)	8.0%-12.0%	32 (5.8)	4.1%-8.2%	69 (29.0)	23%-35%	151 (83.9)	78%-89%

[a]ICD-10: International Classification of Diseases, Tenth Revision.

[b]eGFR: estimated glomerular filtration rate ($ml \cdot min^{-1} \cdot m^{-2}$).

[c]Index eGFR: latest eGFR measure at the time of extraction of the cohort.

Figure 2. Probability of remaining in advanced CKD stages (4 and 5) based on various EHR phenotypes at 6-month follow-up. CKD: chronic kidney disease; EHR: electronic health record; eGFR: estimated glomerular filtration rate; ICD-10: International Classification of Diseases, Tenth Revision.



Discussion

Principal Findings

Accurate identification of an advanced CKD cohort within a clinical database can allow large health care organizations to provide targeted evidence-based clinical care, conduct system-wide needs assessment studies, and facilitate clinical and epidemiological outcome studies. Several EHR-based models to identify CKD using ICD codes and laboratory values have been published [12,21,22]. While there is a reasonable consensus regarding the EHR-based strategies to define CKD within a clinical database, no targeted study has examined the feasibility of extracting an advanced CKD cohort in such databases. Exploring the clinical database of one of the largest regional Veterans health care systems in the country, we identified several coding, identification, and accuracy-related concerns in extracting an advanced CKD cohort.

Researchers have conventionally used the provider diagnosis codes to identify and stage patients with CKD in clinical databases. Using the more accurate eGFR-based definitions, several investigators have shown that identifying CKD cohorts purely by diagnostic codes underestimates its true prevalence [23]. For example, Diamantidis et al [24] showed that the clinical recognition of CKD utilizing diagnostic codes was only 11.8% among Medicare beneficiaries. In a systemic review of studies primarily conducted on non-VHA health care databases, Grams et al [23] found that the coding accuracy for CKD varies widely between 8% and 83%, depending on providers' awareness, and rises with the comorbidity burden and severity of CKD.

Few investigators have evaluated the use and accuracy of CKD diagnosis codes in the VHA clinical database. In a recent analysis of the national VHA database, Saran et al [16] estimated the burden and cost of CKD care on VHA among over 6 million VHA-registered Veterans. While the investigators did not examine the coding accuracy, they found its overall use to be very low (3.2%) compared to much higher estimates (8.02%-27%) obtained using laboratory values [16]. Similar results were recently obtained by Bansal et al [19] in a selective cohort of Veterans with diabetes/hypertension at Veteran Integrated Service Network 17. They found that the laboratory-based prevalence of CKD was approximately 36%, but only 44% of them had diagnosis codes for CKD [19]. Similarly, Norton et al [25] found that 63% of entries lacked CKD codes in a military health system. In conjunction with these reports, our analysis showed that the sensitivity and PPV of diagnosis codes, when compared to the eGFR-based phenotypes, to identify advanced CKD is low, in the range of 55%-65% and 55%-74 %, respectively. Our study further shows that when prospectively followed, nearly one-third of the cohort defined by diagnosis codes had an eGFR value over 30 $ml{\cdot}min^{-1}{\cdot}1.73\ m^{-2}$ at the end of 6-month study. Overall, our findings confirm that the utility and accuracy of diagnosis codes for identifying advanced CKD cohorts in the VHA clinical database is poor.

There are also concerns about using an eGFR-based staging system in clinical databases. EHR-based phenotypes require laboratory measurements of creatinine; however, the regular and periodic availability of creatinine may be inconsistent in the clinical databases. For example, Norton et al [14] showed that only 55% of the study sample had eGFR measurements while validating their CKD EHR phenotype. Similarly, a study examining the VA database showed that only 65% of the VA users had any measurements of eGFR during the study period [16]. This lack of availability of eGFR measures can generate errors in the measurement of disease burden. Further, while the definition of CKD requires the demonstration of a persistent reduction of renal function, many studies report CKD staging statistics using a single eGFR value, with a significant fraction of the cohort lacking the second reported eGFR value. For example, in an analysis performed by the National Kidney

Disease Education Program Workgroup, 31% of patients with stage-4 CKD and 36% of patients with stage-5 CKD did not have a prior eGFR <60 $ml \cdot min^{-1} \cdot 1.73\ m^{-2}$ value available [14]. Similarly, in the analysis by Saran et al [16] examining the burden of CKD in the VA database, only approximately 27% of Veterans had two eGFR measurements more than 90 days apart, raising concerns about the accuracy of the disease burden. However, in our analysis, focusing on the advanced stages of CKD, we found that over 1723 (98%) of Veterans had two eGFR values reported for the initial source cohort, substantially increasing the reliability of screening for advanced CKD. Additionally, we noticed that over 55% (n=981) of the source cohort had subsequent measurements of eGFR over the prospective 6 months (Table 5), further providing a more robust overall reliability of our advanced CKD estimates.

While using eGFR-based phenotypes improves the identification of CKD, staging CKD into stages 3, 4, and 5 can be complex in a clinical database due to physiologic variability in creatinine levels, performance of biochemical tests, frequency of measurements, and intercurrent illness and volume status [13]. Examining such variations in repeat estimations over 3-6 months in the VHA database, Shahinian et al [26] reported that nearly 30% of patients with stage-4 CKD and 6% of patients with stage-5 CKD had eGFR values ≥30 $ml \cdot min^{-1} \cdot 1.73\ m^{-2}$ in the repeat measurements, thus misclassifying as advanced CKD instead of CKD stage 3 [26]. These inaccuracies can lead to the misidentification of patients with advanced CKD, creating misappropriations of clinical resource allocation or errors in research outcomes for studies that target a specific advanced CKD population.

Considering these inherent limitations of eGFR and diagnostic codes, we sought to refine the predictive accuracy of isolating an advanced CKD cohort for TEACH-VET by categorizing our EHR-derived source cohort into high-, intermediate-, and low-risk advanced CKD cohorts using the two latest eGFR values obtained 90 days apart. Assessing the cohort prospectively for 6 months, we found a very high and graded level of stability with our tiered approach, with 94% and 71% of Veterans in the high-risk and intermediate-risk groups having a eGFR less than 30 $ml \cdot min^{-1} \cdot 1.73\ m^{-2}$ at the study end point, thus remaining in an advanced CKD stage. These findings suggest that such an operational definition can significantly improve clinical and research decision-making and optimize resource allocations, which is currently used to prioritize and enroll Veterans in a clinical study targeting advanced CKD [17]. At the same time, we show that approximately 16% of those with a low risk for advanced CKD had an eGFR below 30 $ml \cdot min^{-1} \cdot 1.73\ m^{-2}$ at the 6-month follow-up, highlighting the high-risk individuals even among those with apparent inaccuracies in diagnosis codes.

Our study explored various available methods to provide a more optimal method to obtain the population statistics for an advanced CKD burden and stratified this cohort based on their longitudinal probability of requiring stage-specific care. Examining real-time data and accurately determining the denominator to only those with an available eGFR estimation within the prespecified 12-month period, we found that the prevalence of advanced CKD (high and intermediate risk) was 1.5%, which is 2-3 times higher compared to the US general population estimates (0.5%) derived from National Health and Nutrition Examination Survey (NHANES) enrollees [1,27], but is less than VHA estimates (1.62%) provided by Saran et al [16]. Even based on the conservative estimates and accounting for all the VA users as the denominator, the prevalence of advanced CKD seems to be higher than that of the general population (Table 3). Recently, VHA has implemented a clinical tool for identifying a CKD cohort based on a single eGFR measurement [28]. Further refinements in the tool by implementing the proposed tiered risk approach to identify an advanced CKD population can allow the VHA to implement judicious allocation of care and resources to those in the highest need. A manual chart review showed an error rate of 11%, mainly attributed to the Veterans being on dialysis. Although the VA database can be linked to the United States Renal Data System (USRDS) database and help exclude dialysis patients, there is a lag in the USRDS data and hence this might not be helpful when the need for identification of advanced CKD in real time arises, as intended in our study for enrollment into a clinical trial [8,17]. In the VHS system, using the community care dialysis list can further increase the sensitivity of the screened list and reduce the error rate by excluding the Veterans who are currently receiving dialysis.

Limitations

Our study has a few limitations. In recent times, investigators have described advanced EHR algorithms to identify patients with CKD [29]. However, such phenotypes require complex machine-learning algorithms and validation for the target population, and their application in staging CKD is even further away. This study aimed to explore a pragmatic model for identifying Veterans with high, intermediate, and low risk of advanced CKD in real time that can be easily implemented in routine practice and across a large health care system. Second, we did not incorporate the presence or severity of albuminuria within our parsimonious risk model. However, we believe that it is unlikely to improve upon the model for several reasons. Measurement of albuminuria or even proteinuria is uncommon in clinical databases, including the VHA database, and frequently requires the use of proteinuria categorization on routine urinalysis. The risk for complications and adverse outcomes is significantly high for advanced CKD, as highlighted in the KDIGO classification, irrespective of the degree of albuminuria. Considering the unreliable availability of urine protein measurement, it is likely to be of limited additional value, if any [10]. We acknowledge that the true significance of our parsimonious approach will require studies examining longitudinal clinical outcomes. Third, our eGFR values are based on the creatinine values and utilizing the MDRD equation, according to the then-prevalent practices of the VA CDW at the time of the study. Since the overall intention of the study was to evaluate the methodologies for identifying advanced CKD cohorts within a health care system such as VHA, this is unlikely to change the outcome of the study. Future analyses will need to consider the updated CKD-Epidemiology Collaboration equations incorporating creatinine and cystine values for more accurate staging of CKD. Finally, it needs to be mentioned that our results are applicable only among the active VHA users rather than all VHA-registered Veterans, and thus may misrepresent the true burden of advanced CKD among the entire Veteran population. EHR phenotypes, in general, may exclude people with reduced access to care.

Conclusion

We found that the prevalence of advanced CKD at NF/SG VHS is higher than that in the general population as per various EHR phenotypes, including our EHR model. There is significant discordance between coding and laboratory parameters for the identification of advanced CKD, consistent with other studies. EHR phenotypes based on CKD diagnosis codes alone are insufficient for identification of an advanced CKD cohort in a clinical database. We report a simplified and pragmatic EHR-based model to identify advanced CKD within a regional VHS in real time with a tiered approach that allows allocation of resources to the groups requiring immediate attention and are at risk of progression to ESKD. Further testing of this model is needed to determine its broader applicability across the VHA. If validated, similar models can be tested across the non-VHA databases to identify the true burden of advanced CKD and target clinical care in real time.

Acknowledgments

This study was supported by funding from the Department of Veterans Affairs, Office of Rural Health FY21 Awards, and Health Service Research and Development Awards (I01HX002639). AMS additionally receives ongoing grant support from the Department of Veterans Affairs, Clinical Science Research and Development Merit Grant (I01CX001661).

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Conflicts of Interest

None declared.

Multimedia Appendix 1

ICD codes used for deriving the source cohort and stratifying the advanced CKD cohort (Tables S1) and source data for Figure 2 (Table S2).

[DOCX File , 16 KB - ijmr_v12i1e43384_app1.docx]

References

1. Chronic Kidney Disease Surveillance System-United States. Centers for Disease Control and Prevention. URL: http://www.cdc.gov/ckd [accessed 2021-12-06]
2. Smart NA, Titus TT. Outcomes of early versus late nephrology referral in chronic kidney disease: a systematic review. Am J Med 2011 Nov;124(11):1073-1080. [doi: 10.1016/j.amjmed.2011.04.026] [Medline: 22017785]
3. Fung E, Chang TI, Chertow GM, Thomas I, Asch SM, Kurella Tamura M. Receipt of nephrology care and clinical outcomes among veterans with advanced CKD. Am J Kidney Dis 2017 Nov;70(5):705-714 [FREE Full text] [doi: 10.1053/j.ajkd.2017.06.025] [Medline: 28811048]
4. Shukla AM, Easom A, Singh M, Pandey R, Rotaru D, Wen X, et al. Effects of a comprehensive predialysis education program on the home dialysis therapies: a retrospective cohort study. Perit Dial Int 2017;37(5):542-547. [doi: 10.3747/pdi.2016.00270] [Medline: 28546368]
5. Shukla AM, Hinkamp C, Segal E, Ozrazgat Baslanti T, Martinez T, Thomas M, et al. What do the US advanced kidney disease patients want? Comprehensive pre-ESRD Patient Education (CPE) and choice of dialysis modality. PLoS One 2019 Apr 9;14(4):e0215091 [FREE Full text] [doi: 10.1371/journal.pone.0215091] [Medline: 30964936]
6. Easom AM, Shukla AM, Rotaru D, Ounpraseuth S, Shah SV, Arthur JM, et al. Home run-results of a chronic kidney disease telemedicine patient education study. Clin Kidney J 2020 Oct;13(5):867-872 [FREE Full text] [doi: 10.1093/ckj/sfz096] [Medline: 33123362]
7. Shukla AM, Bozorgmehri S, Ruchi R, Mohandas R, Hale-Gallardo JL, Ozrazgat-Baslanti T, et al. Utilization of CMS pre-ESRD Kidney Disease Education services and its associations with the home dialysis therapies. Perit Dial Int 2021 Sep 01;41(5):453-462 [FREE Full text] [doi: 10.1177/0896860820975586] [Medline: 33258420]
8. Saran R, Li Y, Robinson B, Abbott KC, Agodoa LYC, Ayanian J, et al. US Renal Data System 2015 Annual Data Report: epidemiology of kidney disease in the United States. Am J Kidney Dis 2016 Mar;67(3 Suppl 1):Svii, S1-Svii,305 [FREE Full text] [doi: 10.1053/j.ajkd.2015.12.014] [Medline: 26925525]
9. Shukla AM, Cavanaugh KL, Wadhwa A, Crowley ST, Fried L. Basic requirements for improving home dialysis utilization: universal access to specialty nephrology care and comprehensive pre-ESKD education. J Am Soc Nephrol 2023 Jan 01;34(1):21-25. [doi: 10.1681/ASN.2022060685] [Medline: 36283810]

10. Kidney Disease: Improving Global Outcomes (KDIGO) CKD Working Group. KDIGO 2012 Clinical practice Guideline for the Evaluation and Mangement of Chroinc Kidney disease. Kidney Int Suppl 2012:19-62.
11. VHA directive 1053. Chronic kidney disease prevention, early recognition, and management. Department of Veterans Affairs. 2020. URL: https://www.va.gov/vhapublications/publications.cfm?pub=1 [accessed 2021-12-06]
12. Drawz P, Archdeacon P, McDonald CJ, Powe N, Smith K, Norton J, et al. CKD as a model for improving chronic disease care through electronic health records. Clin J Am Soc Nephrol 2015 Aug 07;10(8):1488-1499 [FREE Full text] [doi: 10.2215/CJN.00940115] [Medline: 26111857]
13. Shang N, Khan A, Polubriaginof F, Zanoni F, Mehl K, Fasel D, et al. Medical records-based chronic kidney disease phenotype for clinical care and "big data" observational and genetic studies. NPJ Digit Med 2021 Apr 13;4(1):70. [doi: 10.1038/s41746-021-00428-1] [Medline: 33850243]
14. Norton JM, Ali K, Jurkovitz CT, Kiryluk K, Park M, Kawamoto K, et al. Development and validation of a pragmatic electronic phenotype for CKD. Clin J Am Soc Nephrol 2019 Sep 06;14(9):1306-1314 [FREE Full text] [doi: 10.2215/CJN.00360119] [Medline: 31405830]
15. Tummalapalli SL, Peralta CA. An electronic CKD phenotype: a step forward in improving kidney care. Clin J Am Soc Nephrol 2019 Sep 06;14(9):1277-1279 [FREE Full text] [doi: 10.2215/CJN.08180719] [Medline: 31492804]
16. Saran R, Pearson A, Tilea A, Shahinian V, Bragg-Gresham J, Heung M, VA-REINS Steering Committee, VA Advisory Board. Burden and cost of caring for US Veterans with CKD: initial findings from the VA Renal Information System (VA-REINS). Am J Kidney Dis 2021 Mar;77(3):397-405. [doi: 10.1053/j.ajkd.2020.07.013] [Medline: 32890592]
17. Shukla AM, Hale-Gallardo J, Orozco T, Freytes I, Purvis Z, Romero S, et al. A randomized controlled trial to evaluate and assess the effect of comprehensive pre-end stage kidney disease education on home dialysis use in veterans, rationale and design. BMC Nephrol 2022 Mar 30;23(1):121 [FREE Full text] [doi: 10.1186/s12882-022-02740-8] [Medline: 35354430]
18. Levey AS, Bosch JP, Lewis JB, Greene T, Rogers N, Roth D. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. Ann Intern Med 1999 Mar 16;130(6):461-470. [doi: 10.7326/0003-4819-130-6-199903160-00002] [Medline: 10075613]
19. Bansal S, Mader M, Pugh JA. Screening and recognition of chronic kidney disease in VA Health Care System primary care clinics. Kidney360 2020 Sep 24;1(9):904-915 [FREE Full text] [doi: 10.34067/KID.0000532020] [Medline: 35369564]
20. R: A language and environment for statistical computing. URL: https://www.R-project.org/ [accessed 2021-12-06]
21. Mendu ML, Ahmed S, Maron JK, Rao SK, Chaguturu SK, May MF, et al. Development of an electronic health record-based chronic kidney disease registry to promote population health management. BMC Nephrol 2019 Mar 01;20(1):72 [FREE Full text] [doi: 10.1186/s12882-019-1260-y] [Medline: 30823871]
22. Mendu M, Schneider L, Aizer A, Singh K, Leaf D, Lee T, et al. Implementation of a CKD checklist for primary care providers. Clin J Am Soc Nephrol 2014 Sep 05;9(9):1526-1535 [FREE Full text] [doi: 10.2215/CJN.01660214] [Medline: 25135764]
23. Grams ME, Plantinga LC, Hedgeman E, Saran R, Myers GL, Williams DE, CDC CKD Surveillance Team. Validation of CKD and related conditions in existing data sets: a systematic review. Am J Kidney Dis 2011 Jan;57(1):44-54 [FREE Full text] [doi: 10.1053/j.ajkd.2010.05.013] [Medline: 20692079]
24. Diamantidis CJ, Hale SL, Wang V, Smith VA, Scholle SH, Maciejewski ML. Lab-based and diagnosis-based chronic kidney disease recognition and staging concordance. BMC Nephrol 2019 Sep 14;20(1):357 [FREE Full text] [doi: 10.1186/s12882-019-1551-3] [Medline: 31521124]
25. Norton JM, Grunwald L, Banaag A, Olsen C, Narva AS, Marks E, et al. CKD prevalence in the Military Health System: coded versus uncoded CKD. Kidney Med 2021 Jul;3(4):586-595 [FREE Full text] [doi: 10.1016/j.xkme.2021.03.015] [Medline: 34401726]
26. Shahinian VB, Hedgeman E, Gillespie BW, Young EW, Robinson B, Hsu C, CDC CKD Surveillance System. Estimating prevalence of CKD stages 3-5 using health system data. Am J Kidney Dis 2013 Jun;61(6):930-938 [FREE Full text] [doi: 10.1053/j.ajkd.2013.01.018] [Medline: 23489675]
27. Myers O, Pankratz V, Norris K, Vassalotti J, Unruh M, Argyropoulos C. Surveillance of CKD epidemiology in the US - a joint analysis of NHANES and KEEP. Sci Rep 2018 Oct 26;8(1):15900. [doi: 10.1038/s41598-018-34233-w] [Medline: 30367154]
28. Chronic Kidney Disease Dashboards. powerbigov.us. US Department of Veterans Affairs. URL: https://app.powerbigov.us/groups/me/reports/1033dde8-98b0-4d13-9117-0553a6213576/ReportSection3c3f482e02a663a111c7 [accessed 2022-09-23]
29. Ilyas H, Ali S, Ponum M, Hasan O, Mahmood MT, Iftikhar M, et al. Chronic kidney disease diagnosis using decision tree algorithms. BMC Nephrol 2021 Aug 09;22(1):273 [FREE Full text] [doi: 10.1186/s12882-021-02474-z] [Medline: 34372817]

Abbreviations

AKI: acute kidney injury
CDW: Corporate Data Warehouse
CKD: chronic kidney disease
CPT: Current Procedural Terminology

eGFR: estimated glomerular filtration rate
EHR: electronic health record
ESKD: end-stage kidney disease
ICD: International Classification of Diseases
KDIGO: Kidney Disease: Improving Global Outcomes
MDRD: Modification of Diet in Renal Disease
NF/SG: North Florida/South Georgia
NHANES: National Health and Nutrition Examination Survey
PPV: positive predictive value
TEACH-VET: Trial to Evaluate and Assess the effects of Comprehensive pre-ESKD education on Home dialysis among Veterans
USRDS: United States Renal Data System
VA: Veterans Affairs
VHA: Veterans Health Administration
VHS: Veterans Health System
VINCI: Veterans Affairs Informatics and Computing Infrastructure

Edited by A Mavragani; submitted 10.10.22; peer-reviewed by C Yang, J Norton, R Kotagiri; comments to author 21.02.23; revised version received 14.04.23; accepted 17.04.23; published 24.07.23.

Please cite as:
Chamarthi G, Orozco T, Shell P, Fu D, Hale-Gallardo J, Jia H, Shukla AM
Electronic Phenotype for Advanced Chronic Kidney Disease in a Veteran Health Care System Clinical Database: Systems-Based Strategy for Model Development and Evaluation
Interact J Med Res 2023;12:e43384
URL: https://www.i-jmr.org/2023/1/e43384
doi:10.2196/43384
PMID:37486757

Original Paper

Left Head Rotation as an Alternative to Difficult Tracheal Intubation: Randomized Open Label Clinical Trial

Danya P Chan[1], MD; George Carlos Rosendo M Jularbal III[1], MD, DPBA; Ismael Julius R Mapili[1], MD, DPBA, DPBPM

Department of Anesthesiology, Baguio General Hospital and Medical Center, Baguio City, Philippines

Corresponding Author:
Danya P Chan, MD
Department of Anesthesiology
Baguio General Hospital and Medical Center
Governor Pack Road
Baguio City, 2600
Philippines
Phone: 63 9958515061
Email: danyachan@yahoo.com

Abstract

Background: Tracheal intubation is a life-saving intervention, and optimizing the patient's head and neck position for the best glottic view is a crucial step that accelerates the procedure. The left head rotation maneuver has been recently described as an innovative alternative to the traditional sniffing position used for tracheal intubation with marked improvement in glottic visualization.

Objective: This study compared the glottic view and intubating conditions in the sniffing position versus left head rotation during direct laryngoscopy.

Methods: This randomized, open-label clinical trial enrolled 52 adult patients admitted to Baguio General Hospital and Medical Center from September 2020 to January 2021 for an elective surgical procedure requiring tracheal intubation under general anesthesia. Intubation was done using a 45° left head rotation in the experimental group (n=26), while the control group (n=26) was intubated using the conventional sniffing position. Glottic visualization and intubation difficulty with the two procedures were assessed using the Cormack-Lehane grade and Intubation Difficulty Scale, respectively. Successful intubation is measured by observing a capnographic waveform in the end-tidal CO_2 monitor after placement of the endotracheal tube.

Results: There was no statistically significant difference in the Cormack-Lehane grade, with 85% (n=44) of patients classified under grades 1 (n=11 and n=15) and 2 (n=11 and n=7) in the left head rotation and sniffing position groups, respectively. In addition, there were no statistically significant differences in the Intubation Difficulty Scale scores of patients intubated with left head rotation or sniffing position; 30.7% (n=8) of patients in both groups were easily intubated, while 53.8% (n=14) in left head rotation and 57.6% (n=15) in sniffing position groups were intubated with slight difficulty. Similarly, there were no significant differences between the 2 techniques in any of the 7 parameters of the Intubation Difficulty Scale, although numerically fewer patients required the application of additional lifting force (n=7, 26.9% vs n=11, 42.3%) or laryngeal pressure (n=3, 11.5% vs n=7, 26.9%) when intubated with left head rotation. The intubation success rate with left head rotation was 92.3% versus 100% in the sniffing position, but this difference was not statistically significant.

Conclusions: Left head rotation produces comparable laryngeal exposure and intubation ease to the conventional sniffing position. Therefore, left head rotation may be an alternative for patients who cannot be intubated in the sniffing position, especially in hospitals where advanced techniques such as video laryngoscopes and flexible bronchoscopes are unavailable, as is the case in this study. However, since our sample size was small, studies with a larger study population are warranted to establish the generalizability of our findings. In addition, we observed inadequate familiarity among anesthesiologists with the left head rotation technique, and the intubation success rate may improve as practitioners attain greater technical familiarization.

Trial Registration: International Standard Randomised Controlled Trial Number (ISRCTN)ISRCTN23442026; https://www.isrctn.com/ISRCTN23442026

(Interact J Med Res 2023;12:e42500) doi:10.2196/42500

KEYWORDS

tracheal; endotracheal; intubation; airway management; sniffing position; LeHeR

Introduction

Tracheal intubation is an essential life-saving intervention. However, patient intubation in a difficult airway requires specialized technical skills, availability of appropriate equipment, and proper assessment of the clinical situation and priorities [1]. Consequently, experienced and inexperienced physicians or allied health professionals routinely encounter difficult intubation situations in the hospital and prehospital settings [2]. Moreover, predicting airway management-related difficulties remains a challenge and cause of frustration among anesthesiologists [3]. Although some studies have attempted to predict difficult intubation using a simple bedside physical examination [4], others have noted the limited and inconsistent capacity of bedside physical examination to identify patients with difficult airways [5]. Furthermore, assessing the risk of difficult airway intubation beforehand may be impossible during emergencies [6].

A study by Cheong et al [7] on airway practices suggested that standard airway examinations could predict only about half of the difficult intubations. Poor visualization of the larynx often leads to difficult intubation, which may result in complications such as aspiration, esophageal intubation, and prolonged hypoxia. Subsequently, these complications may increase patient morbidity and mortality [8]. Therefore, optimizing the patient's head and neck position for the best glottic view is crucial for successful tracheal intubation [9]. Achieving optimal head and neck position is also included in the Difficult Airway Society guidelines for managing adult patients with unanticipated difficult tracheal intubation [10].

Several head and body positions are used to facilitate tracheal intubation. The sniffing position, which is achieved by the flexion of the lower cervical spine, the extension of the upper cervical spine, and the extension of the atlanto-occipital joint [9], is the preferred position among anesthesiologists [11] and is the current gold standard in the intubation process [12]. Several studies have reported attaining an optimal head position for direct laryngoscopy and intubation with the normal airway in the sniffing position [9,12,13]. However, in some studies, the sniffing position did not improve glottic visualization, the success rate on first intubation, or intubation time [14,15]. These inconsistent findings with sniffing position pose a challenge for tracheal intubation in cases where alternate intubation techniques and devices, such as video laryngoscopes and flexible bronchoscopes, especially in low- and middle-income settings where advanced techniques may not be readily available in all hospitals. Therefore, anesthesiologists continuously explore other modalities to optimize the glottic view during direct laryngoscopy [16]. Consequently, various maneuvers have emerged as an alternative to the sniffing position, such as cricoid pressure application [17]; backward, upward, and rightward pressure [18]; head extension [19]; and external laryngeal manipulation [20].

Intubation in the lateral position has been especially well studied [21-24]. A systematic review of different intubation positions in trauma patients suggests reduced airway patency in the supine position compared to the lateral position [25]. In a supine position, the mechanisms of upper airway obstruction include reduction of pharyngeal dilator muscle activity and gravitational effects on anterior upper airway structures [26]. In contrast, lateral position widens the upper airway [27]; hence, upper airway obstruction can be significantly reduced to improve laryngeal visualization. Although some studies suggest that the lateral position may be more difficult than the supine position [28], a reduction in intubation time has been noted after the third attempt in the lateral position [29]. In a more recent study by Goh et al [30], patients were successfully intubated in the lateral position by anesthesiology trainees on the first attempt, with a mean duration of intubation of 57.3 (SD 36.4) seconds. The successful use of a video laryngoscope in the lateral position has also been previously reported [31]. Furthermore, some studies suggest that the head-elevated laryngeal position may be superior to the sniffing position [14,32], although the degree of head elevation necessary to facilitate the external auditory meatus and sternal notch alignment may vary among patients. Thus, Myatra [16] proposed abandoning the conventional “one size fits all” approach with headrests at a fixed height and considering an individualized intervention when positioning patients for laryngoscopy.

Adding to the range of available head and body positions to facilitate tracheal intubation, in 2019, Yezid et al [8] reported using the left head rotation maneuver to optimize head and neck position during tracheal intubation in nontrauma patients. Like the lateral position, left head rotation increases the upper airway's cross-sectional area due to the lateral displacement of the esophagus to the left of the cricoid cartilage. However, this lateral displacement of the esophagus has only been reported in awake nontrauma patients [33], while studies in sleeping subjects did not observe a decreased pharyngeal pressure with left head rotation [34].

Thus, whether head rotation improves airway patency and glottic visualization in anesthetized individuals remains uncertain. Therefore, in this randomized open-label clinical trial, we aimed to compare the glottic view and ease of intubation with left head rotation versus the conventional sniffing position during direct laryngoscopy of patients undergoing elective surgery and evaluate if the left head rotation maneuver is a viable alternative for difficult endotracheal intubation.

Methods

Research Design

This randomized open-label clinical trial enrolled patients admitted to Baguio General Hospital and Medical Center, Baguio City, Cordillera Administrative Region, the Philippines, from September 2020 to January 2021 for an elective surgical procedure requiring tracheal intubation under general anesthesia.

Study Outcomes

The primary study outcome was intubation success rate with direct laryngoscopy using 45-degree left head rotation. Intubation was deemed successful if a capnographic waveform in the end-tidal CO_2 monitor was observed after the endotracheal tube placement, and the intubation attempt was no longer than 10 minutes. Alternative techniques were used to facilitate that intubation in case intubation was unsuccessful with left head rotation or the sniffing position alone (Table 1). The order in which alternative techniques like cricoid pressure, stylet, and change in operator were used was left to the clinician's discretion. If the intubation was deemed unsuccessful after 2 attempts despite the use of alternative techniques, an alternative position was used (change to sniffing position if difficulty intubating with left head rotation, and vice versa).

Table 1. Description of the 7 parameters and scoring scheme of the Intubation Difficulty Scale.

Parameter	Score
N1: number of attempts >1	One point for every additional attempt if unsuccessful in the first attempt.
N2: number of operators >1	One point for each additional operator.
N3: number of alternative techniques[a]	One point for each alternative technique used.
N4: Cormack-Lehane grade	Zero points for successful intubation; otherwise, add Cormack-Lehane grade for the first attempt.
N5: lifting force required	Zero points for normal lifting force and 1 point for increased force.
N6: external laryngeal manipulation used	Zero points if not used and 1 point if used.
N7: vocal cord mobility	Zero points for abduction and 1 point for adduction.
Total IDS[b] score	Sum of N1 to N7.

[a]Alternative techniques included the change of blade or tube, adding a stylet, changing to nasotracheal intubation, applying pressure on the cricoid cartilage, and using fiberoptic intubation or intubating laryngeal mask airway.

[b]IDS: Intubation Difficulty Scale.

Sample Size

Due to limited studies with intubation using left head rotation, the sample size computation was based on the study by Khan et al [28], where the authors reported a 68% success rate in intubation with direct laryngoscopy using the left lateral position. Therefore, a sample size of 52, with 26 participants in each group, was computed using OPEN-EPI (version 3.1) with a 95% CI and 80% power, assuming a success rate of 68% with left head rotation and 100% with conventional intubation in the sniffing position.

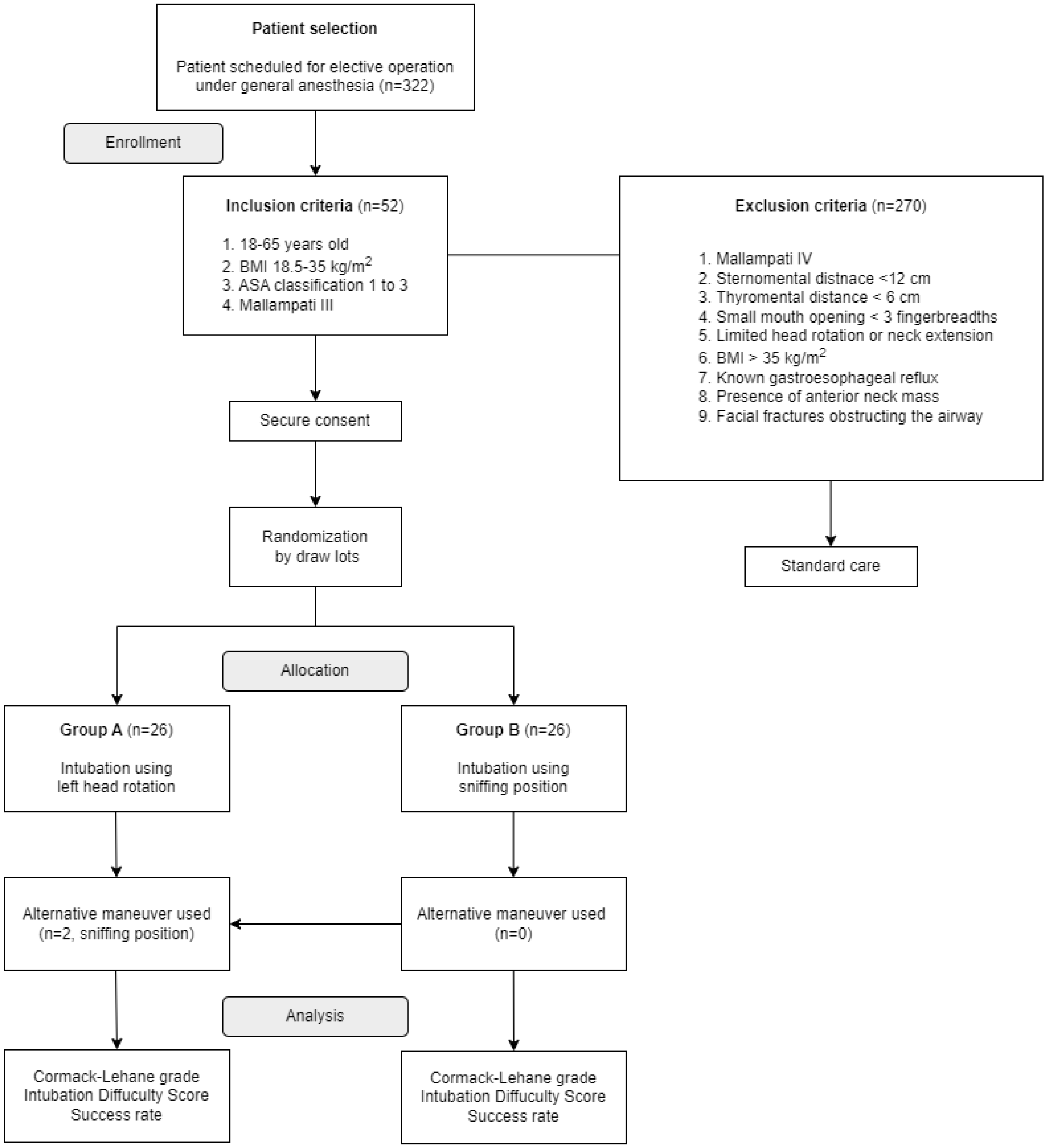
Inclusion or Exclusion Criteria

The criteria for inclusion in this study were patients aged 18-65 years old, BMI range of 18.5-35.0 kg/m^2, American Society of Anesthesiology Physical Status I to III (see [35] for details of American Society of Anesthesiology Physical Status staging), and Mallampati grade III. Mallampati grade measures the visibility of pharyngeal structures (tonsillar pillars, soft palate, and base of uvula), which is noted by instructing the patient to open his or her mouth and protrude the tongue maximally in the sitting posture (see [36] for details of Mallampati grade classification). Patients with sternomental distance <12 cm, thyromental distance <6 cm, small mouth opening <3 fingerbreadths, limited head rotation or neck extension, BMI >35 kg/m^2, known gastroesophageal reflux, presence of anterior neck mass, or facial fractures obstructing the airway were excluded from this study.

Randomization

Enrolled participants who met the inclusion criteria were randomized by draw lots into the experimental (intubated with left head rotation; n=26) and control groups (intubated in the sniffing position; n=26). Group assignments were written on a sheet of paper, which were either "group A" (left head rotation) or "group B" (sniffing position). The papers were shuffled for randomization and numbered for equal participant allocation to each group. The consultant or senior anesthesiology resident opened the papers drawn prior to the induction of anesthesia to determine group assignment. Thus, the consultant or senior anesthesiology resident served as the observer, and the researcher (DPC) was blinded during data collection to avoid bias. In addition, senior anesthesiologists who participated in the data collection were in year 2 or year 3 of their clinical residency. The flow of patient selection and randomization is described in Figure 1.

Figure 1. The flow of patient selection and randomization procedure. ASA: American Society of Anesthesiology.



Ethics Approval

The protocol and informed consent forms were reviewed and approved by the institutional ethics board of Baguio General Hospital and Medical Center (protocol BGHMC-ERC-2020-27). The researcher obtained written informed consent the day before the scheduled operation. The consent form was available in English, Filipino, and Ilocano, with identical content covering the nature of this study; study procedure; risks, benefits, and complications; data security and confidentiality; and voluntary participation and withdrawal. The contents of the consent forms were also verbally explained to the participants, and they were reminded that they were free to withdraw from this study at any point, and if they decided to withdraw prior to the surgical procedure, treatment quality would not vary, and standard care will be provided. The researcher also provided an audio-visual presentation of the intubation procedures in a manner or language that the patient, senior resident, and consultant understood. Several steps were taken to ensure the confidentiality and security of the data. Only DPC has access to the password-protected data, and upon completion of this study, all data were archived in the Hospital Information and Management System office for future reference.

Intubation Procedure

The anesthesiology resident or consultant in charge performed a physical examination and a thorough airway evaluation during the preoperative evaluation to assess the ease of intubation. Laryngoscopy was done using an EMS Fiber Optic Laryngoscope Stubby Handle (EMS) throughout this study period with Macintosh Mega Mac Blade (EMS). Laryngoscope blades were disinfected with Caviwipes (Metrex Research LLC), washed with soap and water, and sterilized to prevent cross-contamination. Before intubation, the laryngoscope's functionality and battery status was checked by a senior resident.

Standard American Society of Anesthesiology monitors (electrocardiogram, noninvasive blood pressure, and pulse oximetry) were applied upon arrival at the operating room. Preprocedural medication included intravenous (IV) injections of midazolam (0.1 mg/kg) for anesthesia and fentanyl (2 mcg/kg) for analgesia. In addition, all patients were preoxygenated with 100% oxygen for 3 minutes through a circle system and a standard face mask with a carbon dioxide or flow sensor between the mask and the breathing circuit. Standard induction included injection of propofol at 2-2.5 mg/kg IV or until the loss of eyelash reflex was achieved and injection of rocuronium 0.6 mg/kg IV for muscle relaxation to facilitate intubation.

Macintosh number 3 or 4 laryngoscope blade was used depending on the anesthesiologist's decision. Intubation was performed with a tracheal tube size of 7.0 in women and 7.5 in men. Intubation was done using a 45-degree left head rotation (estimated with the aid of a protractor) in the experimental group, while the control group was intubated using a sniffing position by placing a cushion under the head such that the external auditory meatus and sternal notch are on the same horizontal plane. Glottic visualization and intubation difficulty with left head rotation and sniffing position were assessed using Cormack-Lehane grade [37] and Intubation Difficulty Scale [38], respectively, which were evaluated by the consultant or senior anesthesiology resident in charge (the researcher was not involved in the scoring).

Cormack-Lehane grade is a conventionally used scale that measures laryngoscopic or glottic view during laryngoscopy [39]. The 4 Cormack-Lehane grades are as follows: complete visualization of the vocal cords (grade 1), visualization of the inferior portion of the glottis (grade 2), visualization of only the epiglottis (grade 3), and nonvisualized epiglottis (grade 4). No external laryngeal pressure was applied for grading the laryngoscopic view [37].

The Intubation Difficulty Scale is an objective and comprehensive assessment of the intubation difficulty based on 7 parameters [38], as described in Table 1. A score of 0 on the Intubation Difficulty Scale represents ideal intubation with minimum difficulty, scores between 1 and 5 represent slight difficulty with intubation, and a score greater than 5 represents moderate to major difficulty with intubation (Table 1).

A carbon dioxide or flow sensor measured end-tidal carbon dioxide, the gold standard for confirming successful tracheal intubation. The airway was secured, and breaths were delivered through the endotracheal tube using an anesthesia ventilator by pressure-regulated volume control mode at 12 breaths per minute, inspiratory to expiratory ratio of 1:2, positive inspiratory pressure of 15 cm H_2O, and positive end-expiratory pressure of 0 cm H_2O. This study protocol ended at this point, and the intended surgical procedure proceeded as planned.

Safety Considerations

The anesthesiologist in charge prioritized the patient's comfort and safety, and any changes in vital signs, such as hypotension and bradycardia, were actively monitored. Adequate hydration, oxygenation, and pain control were maintained throughout the procedure, and the risk of desaturation was minimized with 100% oxygen insufflation during laryngoscopy. Patient safety during apnea was ensured by continued physiological monitoring, including pulse oximetry in all cases. Although routine suction of secretions from the upper airways is not explicitly recommended, it was performed if symptoms suggestive of secretion accumulation were observed. The induction of the anesthetic, as well as the use of neuromuscular blocking agents, followed the latest anesthetic guidelines.

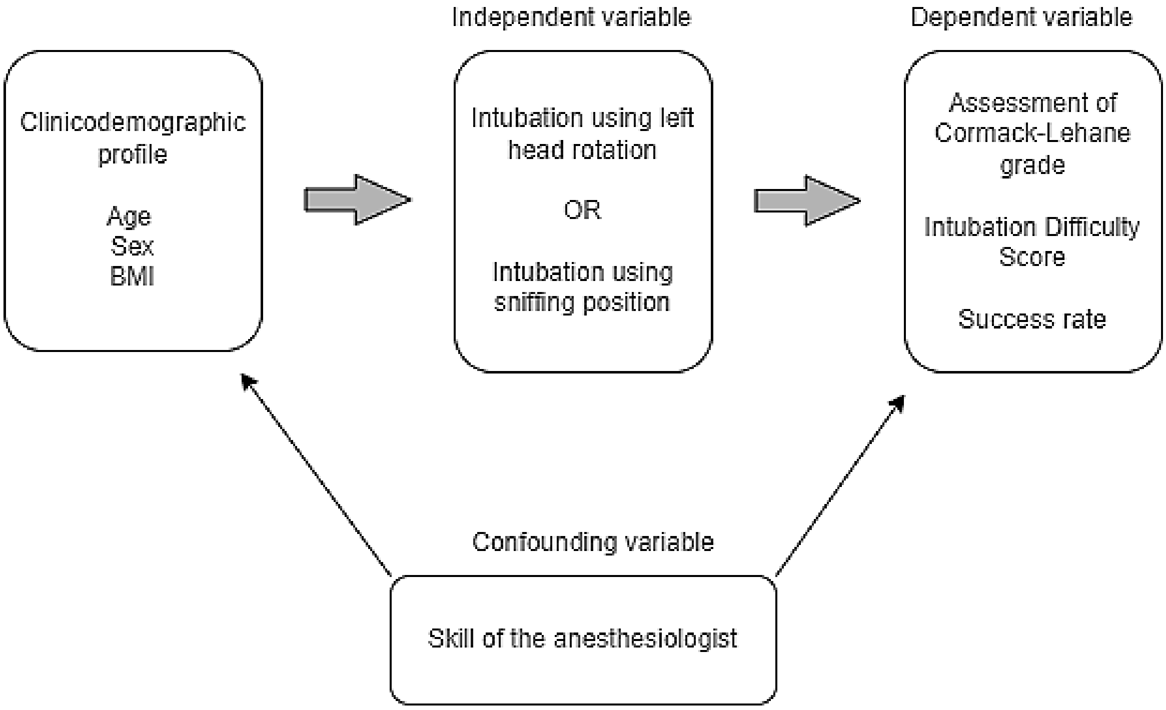
Injuries caused during difficult intubation were managed as follows: if a tooth was chipped or extracted, the patient's watchers were informed, and strict aspiration precautions were applied. Minor lacerations on the lips were allowed to heal via secondary intention, while large lacerations with persistent bleeding were sutured. Patients who failed to be intubated using left head rotation or standard sniffing position received an appropriate standard point of care based on Difficult Airway Society guidelines. An otolaryngologist or general surgeon was available if the procedure required invasive airway access, such as tracheostomy or cricothyrotomy. Untoward reactions were included in the report, and close follow-up was advised.

Statistical Analysis

Statistical analysis was conducted using SPSS (version 17.0; SPSS Inc). The conceptual framework for the analysis is described in Figure 2. Baseline characteristics, which included patient sex, age, and BMI, were presented as frequency and percentage, and the differences between experimental and control groups were compared using the chi-square test.

Noncontinuous variables, including the Cormack-Lehane grade (grade 1-4) and intubation difficulty (minimum, slight, and moderate to major) distribution of patients were presented as frequency and percentage, and the difference between the 2 study groups was assessed using the chi-square test. Additionally, the central tendency in the Cormack-Lehane grade of patients in the left head rotation and sniffing positions was presented as mean (SD) and median (IQR), and the difference between the 2 study groups was assessed using Student *t* test. Individual components (N1-N7) of the Intubation Difficulty Scale were presented as frequency and percentage (assessed using Fisher exact test) as well as mean (SD) and median (IQR) (assessed using Student *t* test). Finally, the intubation success rate was presented as frequency and percentage, and the difference between the 2 study groups was assessed using Fisher exact test. All tests were 2-sided, and P values of <.05 were considered statistically significant.

Figure 2. Conceptual framework: the relationship of variables in left head rotation.



Results

Baseline Characteristics

In total, 52 adult patients were enrolled in this study; 57.7% (n=30) were male, and 55.8% (n=29) were 45 years or older (Table 2). The BMI of 50% (n=26) of the patients in both groups was in the normal range, while the rest were overweight or obese. No between-group differences were noted in the clinicodemographic characteristics of patients intubated with left head rotation or in the sniffing position (Table 2).

Table 2. Baseline characteristics of patients undergoing tracheal intubation with left head rotation or sniffing position (N=26).

Variables	Left head rotation	Sniffing position	*P* value[a]
Sex, n (%)			.16
Male	18 (69.2)	12 (46.1)	
Female	8 (30.7)	14 (53.8)	
Age (years), n (%)			.71
18-26	5 (19.2)	2 (7.6)	
27-35	4 (15.3)	4 (15.3)	
36-44	3 (11.5)	5 (19.2)	
45-53	6 (23.0)	8 (30.7)	
54-65	8 (30.7)	7 (26.9)	
BMI			.92
Normal (18.5-24.9 kg/m^2)			
n (%)	13 (50)	13 (50)	
Mean (SD)	22.8 (1.7)	22.26 (1.8)	
Overweight (25.0-29.9 kg/m^2)			
n (%)	8 (30.7)	7 (26.9)	
Mean (SD)	27.1 (1.4)	27.05 (1.4)	
Obese I (30.0-34.9 kg/m^2)			
n (%)	5 (19.2)	6 (23.0)	
Mean (SD)	32.4 (1.87)	32.14 (1.7)	

[a]Compared using the chi-square test. All values have been truncated to 1 decimal point.

Glottic Visualization and Intubation Difficulty

Glottic visualization in nearly 85% (n=44) of the patients in the left head rotation and sniffing position was classified under grades 1 and 2 on the Cormack-Lehane grade scale. There was no significant association between Cormack-Lehane grade and the 2 intubation positions (*P*=.45; Table 3).

Further, 30.7% (n=16) of patients in both positions were intubated with minimum difficulty, 53.8% (n=14) in left head rotation and 57.6% (n=15) in sniffing position were intubated with slight difficulty, and moderate to major difficulty with intubation was noted in only a small number of patients in the 2 groups (n=4, 15.3% in left head rotation and n=3, 11.5% in sniffing position; Table 3). However, intubation difficulty was not significantly different between the 2 positions (*P*=.91; Table 3).

Although the proportion of patients with an Intubation Difficulty Scale score of 0-1 was numerically higher in the left head rotation group (n=18, 69.2% vs n=13, 50%), the difference was statistically insignificant (*P*=.26; Table 3). Similarly, the differences in median or median scores of the 7 variables of the Intubation Difficulty Scale were statistically insignificant between the 2 intubation positions (Table 3); the number of patients requiring more than one intubation attempt (N1; n=6, 23% vs n=5, 19.2% patients; *P*>.99), more than one operator (N2; n=3, 11.5% vs n=1, 3.8%; *P*=.61), or the use of alternate techniques for the successful passage of the endotracheal tube through the glottis (N3; n=7, 26.9% vs n=6, 23%; *P*>.99) was statistically not different between left head rotation and sniffing positions (Table 3). Similarly, the number of patients for whom the Cormack-Lehane grade for the unsuccessful first attempt was added to the total Intubation Difficulty Scale score (N4; n=14, 53.8% vs n=11, 42.3%; *P*=.50), required the application of additional lifting force (N5; n=7, 26.9% vs n=11, 42.3%; *P*=.38) or laryngeal pressure (N6; n=3, 11.5% vs n=7, 26.9%; *P*=.29) and displayed vocal cord mobility (N7; n=3, 11.5% vs n=2, 7.6%; *P*>.99) and was statistically not different between the 2 intubation positions (Table 3).

Table 3. Glottic visualization, intubation difficulty, and intubation success rate with left head rotation or sniffing position (N=26).

Outcomes	Left head rotation	Sniffing position	*P* value[a]
Cormack-Lehane grade			.45
Grade 1, n (%)	11 (42.3)	15 (57.6)	
Grade 2, n (%)	11 (42.3)	7 (26.9)	
Grade 3, n (%)	3 (11.5)	4 (15.3)	
Grade 4, n (%)	1 (3.8)	0 (0)	
Mean (SD)	1.8 (2.3)	1.8 (1.7)	.95
Median (IQR)	1 (0-2.5)	1.5 (0-3)	
Intubation difficulty, n (%)			.91
Minimum difficulty	8 (30.7)	8 (30.7)	
Slight difficulty	14 (53.8)	15 (57.6)	
Moderate to major difficulty	4 (15.3)	3 (11.5)	
Patients with IDS[b] score of 0-1, n (%)	18 (69.2)	13 (50)	.26
Individual IDS parameters			
Patients with N1 score of >0			
n (%)	6 (23.0)	5 (19.2)	>.99
Mean (SD)	0.2 (0.4)	0.1 (0.4)	.73
Median (IQR)	0 (0-0.2)	0 (0-0)	—[c]
Patients with N2 score of >0			
n (%)	3 (11.5)	1 (3.8)	.61
Mean (SD)	0.1 (0.3)	0.04 (0.2)	.30
Median (IQR)	0 (0-0)	0 (0-0)	—
Patients with N3 score of >0			
n (%)	7 (26.9)	6 (23.0)	>.99
Mean (SD)	0.2 (0.4)	0.2 (0.4)	.74
Median (IQR)	0 (0-1)	0 (0-0.2)	—
Patients with N4 score of >0			
n (%)	14 (53.8)	11 (42.3)	.58
Mean (SD)	0.7 (0.8)	0.5 (0.7)	.50
Median (IQR)	1 (0-1)	0 (0-1)	—
Patients with N5 score of >0			
n (%)	7 (26.9)	11 (42.3)	.38
Mean (SD)	0.2 (0.4)	0.4 (0.5)	.26
Median (IQR)	0 (0-1)	0 (0-1)	—
Patients with N6 score of >0			
n (%)	3 (11.5)	7 (26.9)	.29
Mean (SD)	0.1 (0.3)	0.27 (0.4)	.18
Median (IQR)	0 (0-0.2)	0 (0-0)	—
Patients with N7 score of >0			
n (%)	3 (11.5)	2 (7.6)	>.99
Mean (SD)	0.1 (0.3)	0.08 (0.2)	.63
Median (IQR)	0 (0-0)	0 (0-0)	—

Outcomes	Left head rotation	Sniffing position	P value[a]
Intubation success rate, n (%)	24 (92.3)	26 (100)	.49

[a]The Cormack-Lehane grade and intubation difficulty distribution were assessed using the chi-square test. The proportion of patients who scored >0 in individual IDS parameters was assessed using Fisher exact test. All mean (SD) were assessed using Student *t* test. All values have been truncated to 1 decimal point.

[b]IDS: Intubation Difficulty Scale.

[c]Not applicable.

Intubation Success Rate

The intubation success rate was 100% in the sniffing position (Table 3). Two patients in the sniffing position classified under moderate to major difficulty on the Intubation Difficulty Scale were intubated successfully after the second attempt; hence, shifting position was deemed unnecessary.

On the other hand, 92.3% (n=24) of the patients were successfully intubated using left head rotation (*P*=.49 vs intubation rate in the sniffing position; Table 3). Three patients in the left head rotation were staged under moderate to major difficulty on the Intubation Difficulty Scale. Patient 1 had an Intubation Difficulty Scale score of 6 and had successful intubation after changing the operator on the second attempt. Patient 2 had a grade 3 glottic visualization and an Intubation Difficulty Scale score of 7 in the left head rotation position. Despite using a stylet, cricoid pressure, and additional lifting force, intubation was unsuccessful in this patient after 2 attempts. However, Cormack-Lehane grade improved to grade 2 and the Intubation Difficulty Scale score to 3 upon changing to the sniffing position. Patient 3 had grade 4 glottic visualization with an Intubation Difficulty Scale score of 8. The patient's airway could not be secured using left head rotation despite 2 intubation attempts, the use of a stylet, the application of cricoid pressure and additional lifting force, or the change of operator. After changing to the sniffing position, the Cormack-Lehane grade improved from grade 4 to grade 1, the Intubation Difficulty Scale score improved from 8 to 2, and intubation was successful on the first attempt.

Discussion

Principal Findings

Considering Mallampati III as a sensitive criterion for difficult intubation, the findings of this study suggest that endotracheal intubation with left head rotation can be achieved with comparable glottic visualization and difficulty to the conventional sniffing position in anesthetized patients undergoing elective surgery. However, it is worth noting that numerically fewer patients required the application of increased lifting force and laryngeal pressure when intubated with left head rotation, even though the differences in the 7 constituent parameters of the Intubation Difficulty Scale were statistically nonsignificant between patients intubated with left head rotation and sniffing position.

To our knowledge, this study is the first to comprehensively compare the effectiveness of left head rotation with the sniffing position as the primary technique used to facilitate tracheal intubation of anesthetized nontrauma patients undergoing elective surgery. Except for the case study by Yezid et al [8], which described the intubation of 4 patients using left head rotation, the effect of axial head rotation on airway patency has not been evaluated systematically. However, from our correspondence with the author (Dr Nur Hafiza Yezid, Emergency and Trauma Department, Hospital Jitra, Kedah, Malaysia; December 2019), we are aware of 2 ongoing studies using left head rotation: one being conducted at the Department of Anesthesiology, Ampang Hospital, Malaysia and the other at the Department of Emergency Medicine, University of Malaya, Malaysia. Unfortunately, the results of these investigations are yet to be published.

Nonetheless, prior studies have used variations of left head rotation in specific circumstances. For instance, Le Bervet et al [40] showed improved Cormack-Lehane grade score and intubation efficiency with a left-handed Macintosh blade when combined with a rotation of the cervical spine to the left in about 10% of patients under general endotracheal anesthesia. Similarly, Ueda et al [41] showed that adding left head rotation to the “ramped position” improved the laryngeal view compared to the ramped position alone. Head rotation is also recommended when performing cardiopulmonary resuscitation [42] and during drug-induced sleep endoscopy in patients with obstructive sleep apnea in the supine position [43].

Furthermore, difficult mask ventilation often coexists with difficult tracheal intubation. Two crossover clinical trials [26,44] have compared the efficiency of head rotation on face mask ventilation in patients requiring general anesthesia. Head rotation of 45° in anesthetized apneic adults significantly increased the efficiency of mask ventilation compared with the neutral head position [26]. On the other hand, a 30° clockwise lateral head rotation did not significantly affect mask ventilation volume [44]. It is noteworthy that both crossover clinical trials used right head rotation. However, because airway obstruction for most individuals is symmetric, rotation in the opposite direction is unlikely to alter the findings. In all these cases, intubation with head rotation was successful after more than one intubation attempt and in conjunction with other maneuvers (ramped position, sniffing position, supine position, hyperextension, and aid of a bougie).

The clinical experience of anesthesiologists performing endotracheal intubations may have played a significant role in our assessments of the difficulty of endotracheal intubation. Senior residents and consultants who participated in this study were oriented with the research process but had limited experience with left head rotation. Some awkwardness was noted during the first intubation attempt as residents performed intubation in the left head rotation position. The senior residents also noticed the need for greater familiarization with the left head rotation technique. Since the sniffing position is almost always the default approach, simulation training of left head rotation for practitioners is warranted to provide greater familiarization. Furthermore, regular use of the left head rotation technique in the future and documentation of challenges may help improve the intubation conditions with left head rotation. In this study, most residents noted some difficulty intubating with left head rotation during the first attempt, but intubation became easier during subsequent attempts with left head rotation. Left head rotation maneuver also complies with the Difficult Airway Algorithm recommended by the Difficult Airway Society. With more technical familiarity, it may be a practical noninvasive alternative approach to improve the glottic view among anesthetized patients requiring tracheal intubation. In addition, the potential outcome of this study can benefit patients by providing quicker airway access during intubations and fewer intubation attempts, thereby improving patient safety.

It is worth noting that while this study included patients who had Mallampati III classification during preoperative evaluations, only 8 out of the 52 patients enrolled in this study had a Cormack-Lehane grade of ≥3. Modified Mallampati classification is a widely used tool for predicting difficult airways, and a Mallampati score of III or IV is considered a good predictor of difficult intubation [45,46]. For instance, previous studies by Adnet et al [47] and Oria et al [48,49] showed greater difficulty in intubating patients with Mallampati III and IV, decreased thyromental distance, reduced mouth opening or other anatomical abnormalities than patients without any predictive factors of intubation difficulty. Even though moderate to major difficulty is infrequent in earlier reports and observed in only about 8% of the patients, the rate of intubation with any problem is surprisingly low [47]. However, the Mallampati classification has exceedingly high specificity when used alone, but the sensitivity is typically low, with an increased number of false-positive results [46,50]. While multiple indicators have been identified for predicting difficult airway [4,50] and a single specific technique would be ideal for a quick and easy assessment, the observation in this study supports the findings of previous studies that Mallampati classification, when used solely, may not have adequate sensitivity in predicting difficult laryngoscopy, intubation, or bag-valve-mask ventilation [46,51].

Limitations

There are several limitations of this study. First, this study was conducted over a short timeframe and may have lacked adequate population representation. Second, although we included adequate participants assuming a success rate of 68% with left head rotation, the sample was not large enough to achieve statistical significance when the changes were minor. Therefore, more extensive trials with a larger and more diverse study population are needed to establish the effectiveness of left head rotation or lack thereof. These limitations prevented us from making firm conclusions on some study outcomes. For instance, all patients were successfully intubated in the sniffing position, while 2 patients in the left head rotation required changing to the sniffing position for successful intubation. Therefore, more than one attempt at intubation, the need for more than one operator, and using an alternative technique such as a stylet were more common in the left head rotation group. Although these results indicate that the sniffing position may provide better laryngeal exposure and intubation ease than left head rotation, the small number of patients with the outcome prevents us from drawing a firm conclusion on the superiority of the sniffing position.

Third, given the scarcity of evidence to support the use of left head rotation as a maneuver to optimize tracheal intubation, this study was limited to a patient population where a minimal delay to the intubation period would not present a significant risk to the subject, further limiting the generalizability of our findings. Fourth, this study was conducted during the COVID-19 pandemic, and level 4 personal protective equipment may have influenced the intubation techniques. Studies even before the pandemic have identified the practical problems of excessive heating and fogging while wearing a transparent face shield device during tracheal intubation of patients, although personal protective equipment had no significant effect on the intubation time [52]. Fifth, since this study is a randomized, open-label clinical trial, the observer could not be blinded due to apparent differences in head positions. Lastly, proper airway evaluation and visualization can be affected by the skill of the anesthesiologist, which was not factored in our analysis as all of them had limited experience with left head rotation. In contrast, they all had extensive experience with the sniffing position, which could have confounded our findings.

Conclusions

This study showed that left head rotation produces comparable laryngeal exposure and intubation ease to the conventional sniffing position. Therefore, left head rotation may be an alternative for patients who cannot be intubated in the sniffing position, especially in hospitals where advanced techniques such as video laryngoscopes and flexible bronchoscopes are unavailable, as is the case in this study. Since the sniffing position is used as the default, it remains plausible that better clinical outcomes may be achieved with the left head rotation technique as practitioners attain better technical familiarization. Studies with a larger study population are warranted to establish the generalizability of our findings.

Acknowledgments

The authors would like to thank the Department of Anesthesiology staff and patients from Baguio General Hospital and Medical Center for their participation. In addition, we thank Mr Rodenick Agtarap for his assistance in the data conceptualization and interpretation. Finally, we thank Dr Nur Hafiza Yezid and Dr Khadija Poh Yuen Yoong for supporting this initiative and assisting with the conceptualization of this paper, making this project possible.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author upon reasonable request.

Authors' Contributions

DPC participated in the study conception and design, the acquisition of data, the analysis and interpretation of data, and the drafting of this paper. GCRMJ and IJRM were involved in the study conception and design, data interpretation, and paper revision. All authors read and approved the final paper. The corresponding author had full access to all the data in this study and had final responsibility for the decision to submit for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2312 KB - ijmr_v12i1e42500_app1.pdf]

References

1. Evrin T, Smereka J, Gorczyca D, Bialka S, Ladny J, Katipoglu B, et al. Comparison of different intubation methods in difficult airways during simulated cardiopulmonary resuscitation with continuous chest compression: a randomized cross-over manikin trial. Emerg Med Int 2019;2019:7306204 [FREE Full text] [doi: 10.1155/2019/7306204] [Medline: 31531242]
2. Ambrosio AA, Perez C, Byrnes C, Gaconnet C, Cornelissen C, Brigger MT. Difficult pediatric airway management: a randomized trial comparing laryngeal mask airway, video-assisted, and direct laryngoscopy. Otolaryngol Head Neck Surg 2014;151(S1):P101-P101. [doi: 10.1177/0194599814541627a227]
3. Nørskov AK. Preoperative airway assessment—experience gained from a multicentre cluster randomised trial and the Danish Anaesthesia Database. Dan Med J 2016;63(5):B5241 [FREE Full text] [Medline: 27127020]
4. Shiga T, Wajima Z, Inoue T, Sakamoto A. Predicting difficult intubation in apparently normal patients: a meta-analysis of bedside screening test performance. Anesthesiology 2005;103(2):429-437 [FREE Full text] [doi: 10.1097/00000542-200508000-00027] [Medline: 16052126]
5. Vannucci A, Cavallone LF. Bedside predictors of difficult intubation: a systematic review. Minerva Anestesiol 2016;82(1):69-83 [FREE Full text] [Medline: 25990431]
6. Bohringer C, Duca J, Liu H. A synopsis of contemporary anesthesia airway management. Transl Perioper Pain Med 2019;6(1):5-16 [FREE Full text] [Medline: 31598536]
7. Cheong GPC, Kannan A, Koh KF, Venkatesan K, Seet E. Prevailing practices in airway management: a prospective single-centre observational study of endotracheal intubation. Singapore Med J 2018;59(3):144-149 [FREE Full text] [doi: 10.11622/smedj.2018028] [Medline: 29568855]
8. Yezid NH, Poh K, Noor JM, Arshad A. LeHeR, a simple novel approach for difficult airway in non-trauma patients. BMJ Case Rep 2019;12(8):e230201 [FREE Full text] [doi: 10.1136/bcr-2019-230201] [Medline: 31401573]
9. Akhtar M, Ali Z, Hassan N, Mehdi S, Wani GM, Mir AH. A randomized study comparing the sniffing position with simple head extension for glottis visualization and difficulty in intubation during direct laryngoscopy. Anesth Essays Res 2017;11(3):762-766 [FREE Full text] [doi: 10.4103/0259-1162.204206] [Medline: 28928584]
10. Frerk C, Mitchell VS, McNarry AF, Mendonca C, Bhagrath R, Patel A, Difficult Airway Society Intubation Guidelines Working Group. Difficult Airway Society 2015 guidelines for management of unanticipated difficult intubation in adults. Br J Anaesth 2015;115(6):827-848 [FREE Full text] [doi: 10.1093/bja/aev371] [Medline: 26556848]
11. Reddy RM, Adke M, Patil P, Kosheleva I, Ridley S, Anaesthetic Department at Glan Clwyd Hospital. Comparison of glottic views and intubation times in the supine and 25 degree back-up positions. BMC Anesthesiol 2016;16(1):113 [FREE Full text] [doi: 10.1186/s12871-016-0280-4] [Medline: 27852241]
12. Bhattarai B, Shrestha SK, Kandel S. Comparison of sniffing position and simple head extension for visualization of glottis during direct laryngoscopy. Kathmandu Univ Med J (KUMJ) 2011;9(33):58-63 [FREE Full text] [doi: 10.3126/kumj.v9i1.6265] [Medline: 22610812]
13. Akihisa Y, Hoshijima H, Maruyama K, Koyama Y, Andoh T. Effects of sniffing position for tracheal intubation: a meta-analysis of randomized controlled trials. Am J Emerg Med 2015;33(11):1606-1611. [doi: 10.1016/j.ajem.2015.06.049] [Medline: 26227445]
14. Schmitt HJ, Mang H. Head and neck elevation beyond the sniffing position improves laryngeal view in cases of difficult direct laryngoscopy. J Clin Anesth 2002;14(5):335-338. [doi: 10.1016/s0952-8180(02)00368-9] [Medline: 12208436]
15. Park S, Lee HG, Choi JI, Lee S, Jang EA, Bae HB, et al. Comparison of vocal cord view between neutral and sniffing position during orotracheal intubation using fiberoptic bronchoscope: a prospective, randomized cross over study. BMC Anesthesiol 2019;19(1):3 [FREE Full text] [doi: 10.1186/s12871-018-0671-9] [Medline: 30611215]

16. Myatra SN. Optimal position for laryngoscopy—time for individualization? J Anaesthesiol Clin Pharmacol 2019;35(3):289-291 [FREE Full text] [doi: 10.4103/joacp.JOACP_254_19] [Medline: 31543573]
17. Etanaa NB, Benwu KM. A survey of cricoid pressure application in a single institution in Ethiopia. BMC Res Notes 2019;12(1):546 [FREE Full text] [doi: 10.1186/s13104-019-4586-4] [Medline: 31455397]
18. Lee AR, Yang S, Shin YH, Kim JA, Chung IS, Cho HS, et al. A comparison of the BURP and conventional and modified jaw thrust manoeuvres for orotracheal intubation using the Clarus Video System. Anaesthesia 2013;68(9):931-937 [FREE Full text] [doi: 10.1111/anae.12282] [Medline: 23841798]
19. Hastings RH, Wood PR. Head extension and laryngeal view during laryngoscopy with cervical spine stabilization maneuvers. Anesthesiology 1994;80(4):825-831 [FREE Full text] [doi: 10.1097/00000542-199404000-00015] [Medline: 8024137]
20. Ali MS, Bakri MH, Mohamed HA, Shehab H, Al Taher W. External laryngeal manipulation done by the laryngoscopist makes the best laryngeal view for intubation. Saudi J Anaesth 2014;8(3):351-354 [FREE Full text] [doi: 10.4103/1658-354X.136431] [Medline: 25191185]
21. Komatsu R, Kamata K, You J, Sessler DI, Kasuya Y. Airway scope for tracheal intubation in the lateral position. Anesth Analg 2011;112(4):868-874 [FREE Full text] [doi: 10.1213/ANE.0b013e31820c7cdf] [Medline: 21385975]
22. Fevang E, Haaland K, Røislien J, Bjørshol CA. Semiprone position is superior to supine position for paediatric endotracheal intubation during massive regurgitation, a randomized crossover simulation trial. BMC Anesthesiol 2018;18(1):10 [FREE Full text] [doi: 10.1186/s12871-018-0474-z] [Medline: 29347980]
23. Arai YCP, Nakayama M, Kato N, Wakao Y, Ito H, Komatsu T. The effects of jaw thrust and the lateral position on heart rate variability in anesthetized children with obstructive sleep apnea syndrome. Anesth Analg 2007;104(6):1352-1355 [FREE Full text] [doi: 10.1213/01.ane.0000262041.46833.21] [Medline: 17513625]
24. Arai YCP, Fukunaga K, Hirota S, Fujimoto S. The effects of chin lift and jaw thrust while in the lateral position on stridor score in anesthetized children with adenotonsillar hypertrophy. Anesth Analg 2004;99(6):1638-1641 [FREE Full text] [doi: 10.1213/01.ANE.0000135637.95853.1C] [Medline: 15562046]
25. Hyldmo PK, Vist GE, Feyling AC, Rognås L, Magnusson V, Sandberg M, et al. Is the supine position associated with loss of airway patency in unconscious trauma patients? A systematic review and meta-analysis. Scand J Trauma Resusc Emerg Med 2015;23:50 [FREE Full text] [doi: 10.1186/s13049-015-0116-0] [Medline: 26129809]
26. Itagaki T, Oto J, Burns SM, Jiang Y, Kacmarek RM, Mountjoy JR. The effect of head rotation on efficiency of face mask ventilation in anaesthetised apnoeic adults: a randomised, crossover study. Eur J Anaesthesiol 2017;34(7):432-440 [FREE Full text] [doi: 10.1097/EJA.0000000000000582] [Medline: 28009638]
27. Goldmann K. Recent developments in airway management of the paediatric patient. Curr Opin Anaesthesiol 2006;19(3):278-284. [doi: 10.1097/01.aco.0000192786.93386.d5] [Medline: 16735811]
28. Khan MF, Khan FA, Minai FN. Airway management and hemodynamic response to laryngoscopy and intubation in supine and left lateral positions. Middle East J Anaesthesiol 2010;20(6):795-802. [Medline: 21526663]
29. Nathanson MH, Gajraj NM, Newson CD. Tracheal intubation in a manikin: comparison of supine and left lateral positions. Br J Anaesth 1994;73(5):690-691 [FREE Full text] [doi: 10.1093/bja/73.5.690] [Medline: 7826801]
30. Goh SY, Thong SY, Chen Y, Kong AS. Efficacy of intubation performed by trainees on patients in the lateral position. Singapore Med J 2016;57(9):503-506 [FREE Full text] [doi: 10.11622/smedj.2015165] [Medline: 26768062]
31. Bhat R, Sanickop CS, Patil MC, Umrani VS, Dhorigol MG. Comparison of Macintosh laryngoscope and C-MAC video laryngoscope for intubation in lateral position. J Anaesthesiol Clin Pharmacol 2015;31(2):226-229 [FREE Full text] [doi: 10.4103/0970-9185.155221] [Medline: 25948906]
32. El-Orbany MI, Getachew YB, Joseph NJ, Salem MR, Friedman M. Head elevation improves laryngeal exposure with direct laryngoscopy. J Clin Anesth 2015;27(2):153-158. [doi: 10.1016/j.jclinane.2014.09.012] [Medline: 25468586]
33. Dotson K, Kiger J, Carpenter C, Lewis M, Hill J, Raney L, et al. Alignment of cricoid cartilage and esophagus and its potential influence on the effectiveness of Sellick maneuver in children. Pediatr Emerg Care 2010;26(10):722-725. [doi: 10.1097/PEC.0b013e3181f39b74] [Medline: 20881908]
34. Walsh JH, Maddison KJ, Platt PR, Hillman DR, Eastwood PR. Influence of head extension, flexion, and rotation on collapsibility of the passive upper airway. Sleep 2008;31(10):1440-1447 [FREE Full text] [Medline: 18853942]
35. Hocevar LA, Fitzgerald BM. American Society of Anesthesiologists Staging. Treasure Island, FL: StatPearls; 2023.
36. Samsoon GL, Young JR. Difficult tracheal intubation: a retrospective study. Anaesthesia 1987;42(5):487-490 [FREE Full text] [doi: 10.1111/j.1365-2044.1987.tb04039.x] [Medline: 3592174]
37. Cormack RS, Lehane J. Difficult tracheal intubation in obstetrics. Anaesthesia 1984;39(11):1105-1111 [FREE Full text] [doi: 10.1111/j.1365-2044.1984.tb08932.x] [Medline: 6507827]
38. Adnet F, Borron SW, Racine SX, Clemessy JL, Fournier JL, Plaisance P, et al. The Intubation Difficulty Scale (IDS): proposal and evaluation of a new score characterizing the complexity of endotracheal intubation. Anesthesiology 1997;87(6):1290-1297 [FREE Full text] [doi: 10.1097/00000542-199712000-00005] [Medline: 9416711]
39. Krage R, van Rijn C, van Groeningen D, Loer SA, Schwarte LA, Schober P. Cormack-Lehane classification revisited. Br J Anaesth 2010;105(2):220-227 [FREE Full text] [doi: 10.1093/bja/aeq136] [Medline: 20554633]
40. Le Bervet JY, Vo Van JM, Le Roy M, Eid G, Nguyen Q. Right supine rotation of the head for intubation: A-1007. Eur J Anaesthesiol 2006;23(Suppl 37):259. [doi: 10.1097/00003643-200606001-00931]

41. Ueda W, Hatakeyama S, Arai YCP. The addition of a head rotation when the ramped position fails to provide good laryngeal visualization: a preliminary study. Anesth Pain Med 2018;8(1):e63674 [FREE Full text] [doi: 10.5812/aapm.63674] [Medline: 29868461]
42. Rottenberg EM. Two-thirds receive no bystander cardiopulmonary resuscitation: could head rotation be the solution? Am J Emerg Med 2016;34(10):2011-2013. [doi: 10.1016/j.ajem.2016.07.036] [Medline: 27496369]
43. Safiruddin F, Koutsourelakis I, de Vries N. Analysis of the influence of head rotation during drug-induced sleep endoscopy in obstructive sleep apnea. Laryngoscope 2014;124(9):2195-2199. [doi: 10.1002/lary.24598] [Medline: 24431007]
44. Matsunami S, Komasawa N, Konishi Y, Minami T. Head elevation and lateral head rotation effect on facemask ventilation efficiency: randomized crossover trials. Am J Emerg Med 2017;35(11):1709-1712. [doi: 10.1016/j.ajem.2017.05.004] [Medline: 28499786]
45. Dhanger S, Gupta SL, Vinayagam S, Bidkar PU, Elakkumanan LB, Badhe AS. Diagnostic accuracy of bedside tests for predicting difficult intubation in Indian population: an observational study. Anesth Essays Res 2016;10(1):54-58 [FREE Full text] [doi: 10.4103/0259-1162.165503] [Medline: 26957691]
46. Lundstrøm LH, Vester-Andersen M, Møller AM, Charuluxananan S, L'hermite J, Wetterslev J, Danish Anaesthesia Database. Poor prognostic value of the modified Mallampati score: a meta-analysis involving 177 088 patients. Br J Anaesth 2011;107(5):659-667 [FREE Full text] [doi: 10.1093/bja/aer292] [Medline: 21948956]
47. Adnet F, Racine SX, Borron SW, Clemessy JL, Fournier JL, Lapostolle F, et al. A survey of tracheal intubation difficulty in the operating room: a prospective observational study. Acta Anaesthesiol Scand 2001;45(3):327-332. [doi: 10.1034/j.1399-6576.2001.045003327.x] [Medline: 11207469]
48. Oria MS, Amarkhil OO, Azim H, Halimi SA. Intubation Difficulty Scale and influence of preoperative airway assessment tests in elective surgical patients. Int J Sci Res 2020;9(9):1323-1329. [doi: 10.21275/SR20923132413]
49. Oria MS, Halimi SA, Negin F, Asady A. Predisposing factors of difficult tracheal intubation among adult patients in Aliabad teaching hospital in Kabul, Afghanistan—a prospective observational study. Int J Gen Med 2022;15:1161-1169 [FREE Full text] [doi: 10.2147/IJGM.S348813] [Medline: 35153507]
50. Nurullah M, Alam MS, Hossen M, Shahnawaz M. Prediction of difficult airway by thyromental height test- a comparison with modified Mallampati test. Bangladesh J Med Sci 2018;17(3):455-461 [FREE Full text] [doi: 10.3329/bjms.v17i3.37014]
51. Green SM, Roback MG. Is the Mallampati score useful for emergency department airway management or procedural sedation? Ann Emerg Med 2019;74(2):251-259 [FREE Full text] [doi: 10.1016/j.annemergmed.2018.12.021] [Medline: 30782435]
52. Greenland KB, Tsui D, Goodyear P, Irwin MG. Personal protection equipment for biological hazards: does it affect tracheal intubation performance? Resuscitation 2007;74(1):119-126 [FREE Full text] [doi: 10.1016/j.resuscitation.2006.11.011] [Medline: 17353076]

Abbreviations

IV: intravenous

Edited by A Mavragani; submitted 08.09.22; peer-reviewed by R Mpofu, N Rehman; comments to author 02.02.23; revised version received 12.06.23; accepted 19.06.23; published 04.08.23.

Please cite as:
Chan DP, Jularbal III GCRM, Mapili IJR
Left Head Rotation as an Alternative to Difficult Tracheal Intubation: Randomized Open Label Clinical Trial
Interact J Med Res 2023;12:e42500
URL: https://www.i-jmr.org/2023/1/e42500
doi:10.2196/42500
PMID:37335071

Original Paper

Web-Based Application Based on Human-in-the-Loop Deep Learning for Deidentifying Free-Text Data in Electronic Medical Records: Development and Usability Study

Leibo Liu[1], PhD; Oscar Perez-Concha[1], PhD; Anthony Nguyen[2], PhD; Vicki Bennett[3], MHS; Victoria Blake[4], MSc; Blanca Gallego[1], PhD; Louisa Jorm[1], PhD

[1]Centre for Big Data Research in Health, University of New South Wales, Sydney, Australia

[2]Australian e-Health Research Centre (AEHRC), Commonwealth Scientific and Industrial Research Organisation (CSIRO), Brisbane, Australia

[3]Metadata, Information Management and Classifications Unit (MIMCU), Australian Institute of Health and Welfare, Canberra, Australia

[4]Eastern Heart Clinic, Prince of Wales Hospital, Randwick, Australia

Corresponding Author:
Leibo Liu, PhD
Centre for Big Data Research in Health
University of New South Wales
Level 2, AGSM Building
G27, Botany St
Sydney, 2052
Australia
Phone: 61 290657847
Fax: 61 290657847
Email: z5250377@ad.unsw.edu.au

Abstract

Background: The narrative free-text data in electronic medical records (EMRs) contain valuable clinical information for analysis and research to inform better patient care. However, the release of free text for secondary use is hindered by concerns surrounding personally identifiable information (PII), as protecting individuals' privacy is paramount. Therefore, it is necessary to deidentify free text to remove PII. Manual deidentification is a time-consuming and labor-intensive process. Numerous automated deidentification approaches and systems have been attempted to overcome this challenge over the past decade.

Objective: We sought to develop an accurate, web-based system deidentifying free text (DEFT), which can be readily and easily adopted in real-world settings for deidentification of free text in EMRs. The system has several key features including a simple and task-focused web user interface, customized PII types, use of a state-of-the-art deep learning model for tagging PII from free text, preannotation by an interactive learning loop, rapid manual annotation with autosave, support for project management and team collaboration, user access control, and central data storage.

Methods: DEFT comprises frontend and backend modules and communicates with central data storage through a filesystem path access. The frontend web user interface provides end users with a user-friendly workspace for managing and annotating free text. The backend module processes the requests from the frontend and performs relevant persistence operations. DEFT manages the deidentification workflow as a project, which can contain one or more data sets. Customized PII types and user access control can also be configured. The deep learning model is based on a Bidirectional Long Short-Term Memory-Conditional Random Field (BiLSTM-CRF) with RoBERTa as the word embedding layer. The interactive learning loop is further integrated into DEFT to speed up the deidentification process and increase its performance over time.

Results: DEFT has many advantages over existing deidentification systems in terms of its support for project management, user access control, data management, and an interactive learning process. Experimental results from DEFT on the 2014 i2b2 data set obtained the highest performance compared to 5 benchmark models in terms of microaverage strict entity–level recall and F_1-scores of 0.9563 and 0.9627, respectively. In a real-world use case of deidentifying clinical notes, extracted from 1 referral hospital in Sydney, New South Wales, Australia, DEFT achieved a high microaverage strict entity–level F_1-score of 0.9507 on a corpus of 600 annotated clinical notes. Moreover, the manual annotation process with preannotation demonstrated a 43% increase in work efficiency compared to the process without preannotation.

Conclusions: DEFT is designed for health domain researchers and data custodians to easily deidentify free text in EMRs. DEFT supports an interactive learning loop and end users with minimal technical knowledge can perform the deidentification work with only a shallow learning curve.

(Interact J Med Res 2023;12:e46322) doi:10.2196/46322

KEYWORDS

web-based system; deidentification; electronic medical records; deep learning; narrative free text; human in the loop; free text; unstructured data; electronic health records; machine learning

Introduction

Narrative free-text data in electronic medical records (EMRs) include a variety of clinical documents such as consultation notes, nursing notes, progress notes, and discharge summaries, which contain valuable information for analysis and research to inform better patient care [1-3]. The free-text data can include personally identifiable information (PII), for example, patient name, date of birth, address, phone number, and patient identifier, which can be used to identify an individual on its own or with other information. It is necessary to deidentify the free-text data by removing this PII before releasing to researchers for secondary purposes where the reuse of the data is not covered in patients' informed consent forms or when requested as part of a waiver of informed consent by an institutional review board or any other human research ethics committees, as required by legislation including the Privacy Rule of the HIPAA (Health Insurance Portability and Accountability Act) [4] in the United States and the Privacy Act, 1988 [5], in Australia. However, manual deidentification has been proven to be a time-consuming and labor-intensive task [6].

In the past decade, researchers have investigated many different automated deidentification approaches including rule-based matching [7-9] and machine learning (ML) models [10-14]. Hand-written regular expressions and ad hoc knowledge dictionaries are used in rule-based deidentification approaches for a specific free-text data set [15]. In contrast, ML-based deidentification approaches use manually annotated data to train named entity recognition (NER) models, enabling the prediction of PII entities from free-text data. Although rule-based methods do not necessitate the preparation of annotated data, the rules are challenging to generalize to other corpora without manual adjustments from experienced domain experts [16]. The Bidirectional Long Short-Term Memory-Conditional Random Field (BiLSTM-CRF) has been proven to achieve state-of-the-art or competitive results on the free text deidentification task [17-19]. With the success of transformer models in the NLP domain, some studies began to explore its use on deidentification tasks [20-22]. Johnson et al [21] fine-tuned pretrained transformer models and achieved a binary token-level F_1-score of 0.984 on the 2014 i2b2 test set. However, a benchmark study [20] showed that BiLSTM-CRF achieved better performances compared to transformer-based models. Another study conducted by Tang et al [17] demonstrated that incorporating the pretrained transformer language model as the word embedding layer in BiLSTM-CRF led to an improvement in the F_1-scores on the 2014 i2b2 deidentification data set, compared to using other word embeddings (eg, Word2Vec and ELMo). Furthermore, several ensemble approaches that combine multiple individual ML models have been proposed on deidentification tasks [14,23,24]. By leveraging the strengths of individual methods, these ensemble methods have demonstrated improved performance on deidentification tasks.

The traditional workflow of the ML-based deidentification approaches consist of three stages: (1) annotation: human annotators manually tag all the PII in the free text. The interannotator agreement is calculated to measure the quality of the annotation [13]; (2) model training: ML experts train models using the annotated free text; and (3) deidentification: the PII predicted by the models are substituted by surrogates or tags or removed completely. Although pretrained ML solutions can potentially be used "out of the box," there are significant variations between hospitals, vendors, and countries in the structure and content of EMRs and the nature of PIIs. Furthermore, data custodians may require performance metrics based on their specific data before gaining sufficient confidence to use these tools. Therefore, manual annotation remains a time-consuming process and is the main bottleneck in training ML-based deidentification models [25,26]. To overcome this, several annotation tools (eg, BRAT [27] and WAT [14]) have been used to speed up the annotation stage [6,28]. Nevertheless, the second and third stages of the workflow still require considerable input from ML experts. In recent years, some annotation tools (eg, ezTag [29], INCEpTION [30], and Prodigy [31]) have integrated interactive learning for iteratively retraining the models using the latest annotated free text to provide preannotation suggestions, which automates the second stage of the workflow. These tools can be used to handle some of the deidentification task (ie, PII tagging). Aberdeen et al [32] developed an open-source deidentification tool, MITRE Identification Scrubber Toolkit (MIST), which comprises a web-based graphical annotation tool, a training module, a tagging module, a redaction and resynthesis module, and an experiment engine. These modules work together using a "tag-a-little, learn-a-little" loop strategy to complete the deidentification task, bypassing the need for ML experts for the second and third stages of the traditional deidentification workflow. The annotation tool is used by annotators to tag the PII from the free-text files. The training module trains a conditional random field-based sequence tagger using these annotated files. The tagging module automatically tags the PII for the new files that can be manually corrected by the annotators. Furthermore, MIST provides a workspace mode to conveniently manage a corpus which needs to be deidentified. However, many operations of MIST need to be done via the command line, for example, creating a workspace, importing files into a workspace, and training models. The MIST server needs to be restarted to get the newly trained model into effect. Moreover, the end users of the MIST tool are required to have the technical knowledge to run command lines.

Off-the-shelf tools [33] such as Amazon Comprehend Medical [34], Clinacuity CliniDeID [35], and National Library of Medicine (NLM) Scrubber [36], can be used to deidentify free text directly without following the traditional deidentification workflow. These 3 tools are all HIPAA-compliant, following the HIPAA's "Safe Harbor" method [14] to remove 18 types of identifiers. Amazon Comprehend Medical is a cloud-based service which needs the data be uploaded to its service end point. This represents a significant barrier to adopting it for use with EMRs, which are stored in a secure and internet access-restricted environment. CliniDeID, originally a commercial clinical text deidentification software, has recently been made available as free open-source software since November 2022. This allows for additional retraining on specific data sets to enhance the model's performance. Although NLM Scrubber can be installed locally, its performance cannot be improved because it has no ability to learn from the end users' free-text data, which may vary considerably from the data used to develop the tools. Therefore, data custodians are responsible for reviewing and evaluating the deidentification results provided by these off-the-shelf tools to ensure it meets their benchmarks. Another main obstacle for adopting the off-the-shelf tools is that the PII types present in specific free-text data outside of a HIPAA covered entity or country (such as Australia) can be different from the 18 HIPAA PII types.

In this study, we designed and implemented a web-based system, deidentifying free text (DEFT), for tagging and substituting designated PII in free-text data in EMRs with human (annotators) in the loop. The system can be readily and easily adopted for the free text deidentification task in secure and internet access–restricted network environments. The main features of the system are listed below:

- Web-based: the web-based deidentification system can be easily accessed by multiple end users via a web browser. Compared to desktop-based applications, web-based applications are less hardware dependent and can easily be updated and upgraded.
- Powered by deep learning models: we used a deep learning NER model to recognize and tag PII entities from free-text data.
- Preannotation bootstrapped by a semiautomatic learning loop: following the learning loop we have proposed previously [14], DEFT preannotates the free text using the ML model that is automatically trained on the previous annotated free text completed by the annotators.
- Customized PII types: the PII types can vary depending on the free text corpus and different data sharing scenarios. For example, full dates may need to be kept for future research in a secure environment which can only be accessed by ethically approved users [14].
- Suitable for nontechnical end users: experienced health-domain annotators can complete the whole deidentification task using DEFT. No technical or ML knowledge is required.
- Simple and task-focused web user interface (UI): DEFT makes the annotators focus on the annotation work through a simple and well-designed web UI.
- Implements autosave: each annotation action, including annotator name, PII entity positions, PII type, and annotation time, is saved automatically.
- Fewer clicks for annotation: fewer clicks mean quick annotation and less deidentification time.
- Supports project management and team collaboration: the deidentification task can be managed as a project, which can contain one or more data sets. The team annotators can work on the same data set in the project at the same time to accelerate the annotation process.
- Implements user access control: only approved users can access specific projects.
- Uses central data storage: all the data can be stored in 1 central location which the DEFT server can access through filesystem path. This avoids importing or transferring thousands or millions of small free-text files across the network.

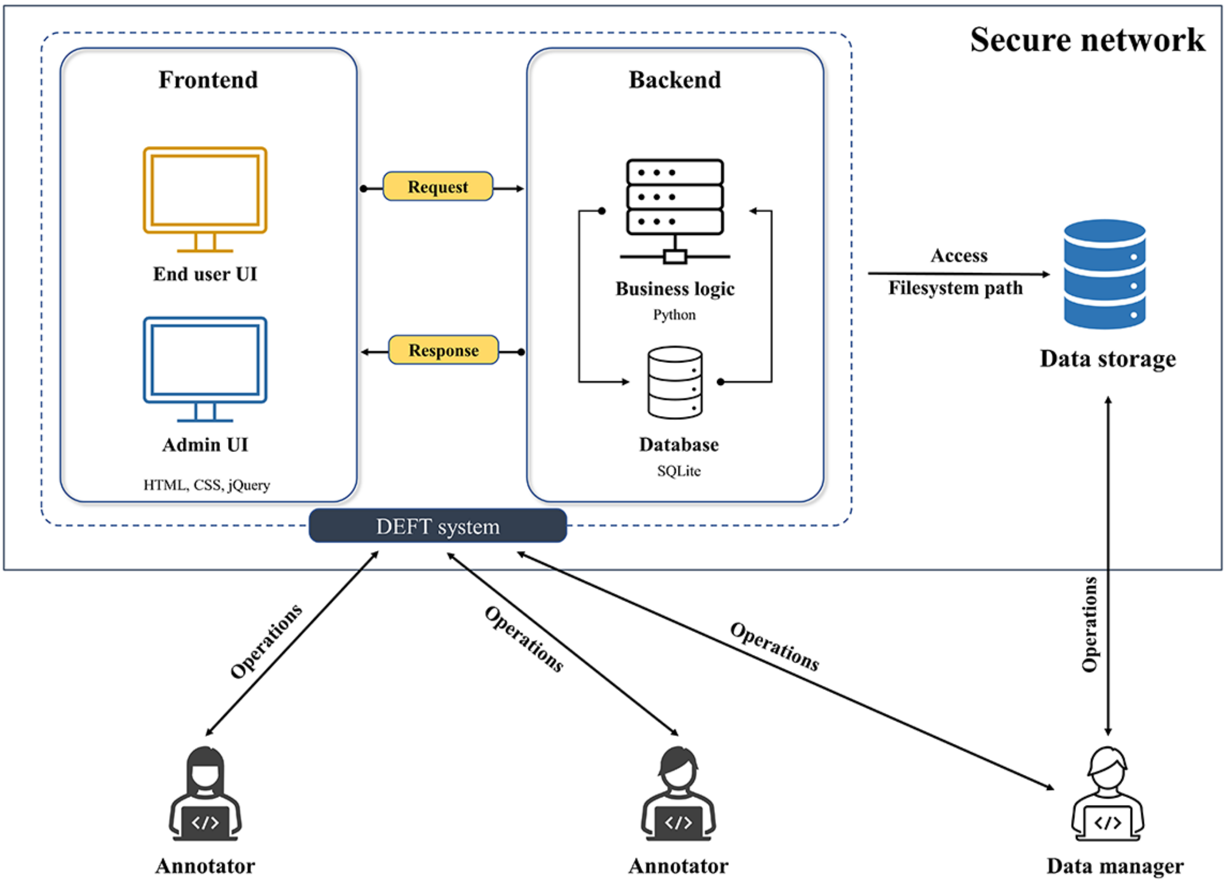
Methods

System Architecture

DEFT has been designed to be simple and manageable, so end users can easily conduct the deidentification work on their own free-text data that are stored in a secure and internet access–restricted environment. Figure 1 shows the overview of the DEFT system architecture. DEFT communicates with the central data storage by using a filesystem path to retrieve the free-text data and generate deidentified data. The annotators remotely access the DEFT web UI to annotate the free text via their own devices. DEFT is responsible for all the business logic processing. The data manager helps to manage projects and data through the DEFT web UI and direct connection to the data storage, respectively.

We implemented DEFT using Django, a high-level Python web framework. As shown in Figure 1, DEFT comprises frontend and backend modules. The frontend module was built with HTML, CSS, and jQuery and has 2 different web UIs including an end user UI and admin UI. The former is used by the annotator to tag the PII and the latter is for the data manager to manage users, models, projects, and PII types. The backend consists of 2 components, that is, business logic and database. The first one is the controller that receives the requests from the frontend and invokes the relevant business logic to produce the responses which are sent back to the frontend. The second one has the persistence component built with SQLite and interacts with the business logic component to store all the application data such as project information, and PII positions and types.

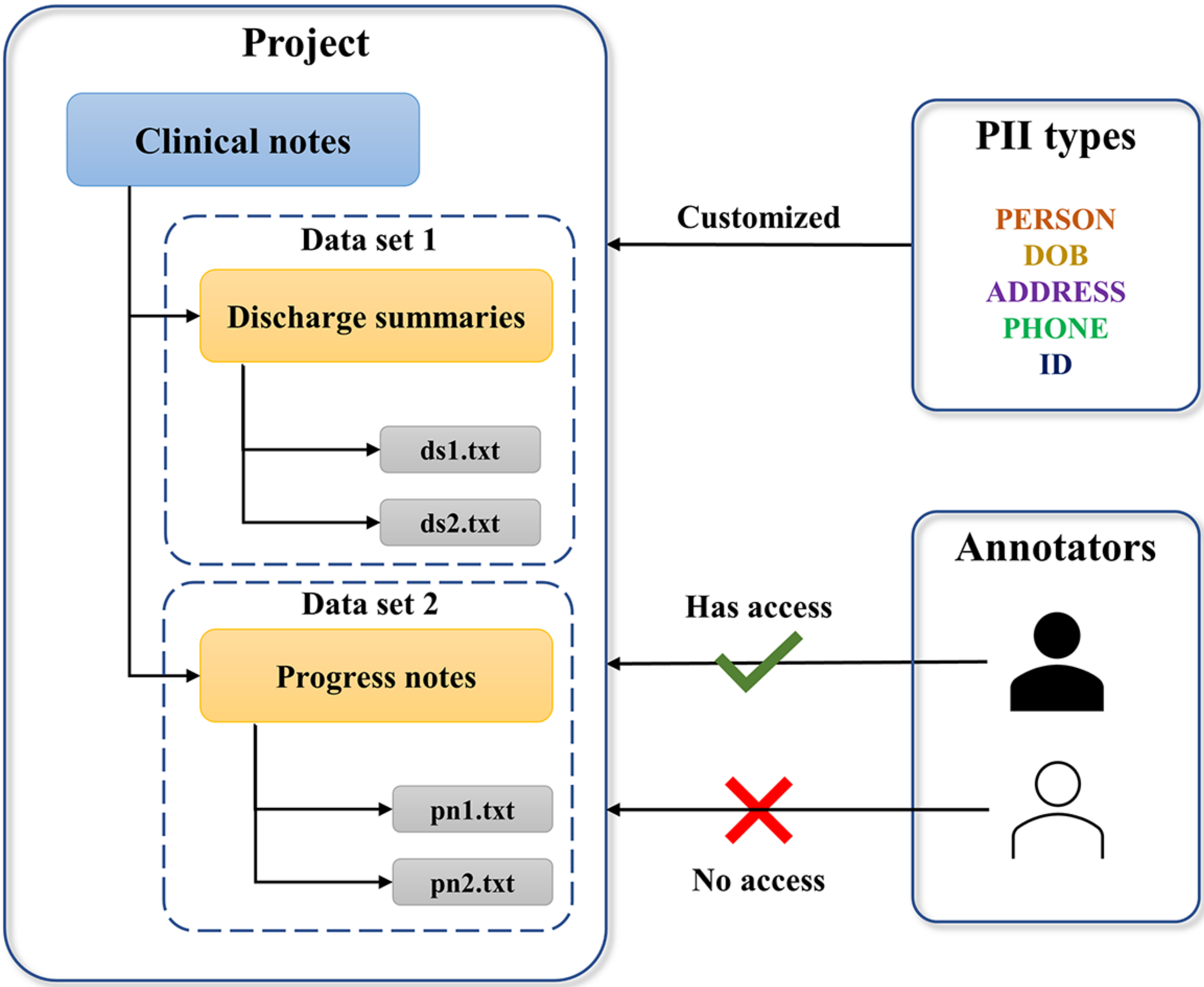
Figure 1. Overview of the DEFT system architecture. DEFT: deidentifying free text; UI: user interface.



Project-Based Deidentification Work

In the DEFT system, the deidentification work is organized as projects which must have at least 1 data set. Multiple projects can be processed at the same time. A project is created by the users for a deidentification task of a new free text corpus. Multiple data sets can be added into the project according to the users' requirements. The users can flexibly customize any number of PII types with different display colors for the project and control who has access to the project. The system does not provide default PII types, because these can vary depending on the project. The raw data files are stored in specific data storage outside the DEFT system and the access path is configured in DEFT for the relevant project. Users can import the data files (txt format) into DEFT manually or wait for the system to import them automatically. The importing operation only saves the file names into the DEFT database rather than the file contents so that all the identified data can be safely maintained in specific data storage. Figure 2 shows an example of a deidentification project structure. A project named "Clinical Notes" is created and 2 data sets are added with the names "Discharge Summaries" and "Progress Notes," respectively. A list of PII types (PERSON, date of birth [DOB], ADDRESS, PHONE, ID) are configured for the project according to the users' requirements. The user shown in solid black is assigned access to the project as the annotator. All the project configurations can be done via the admin UI by the data managers of the team.

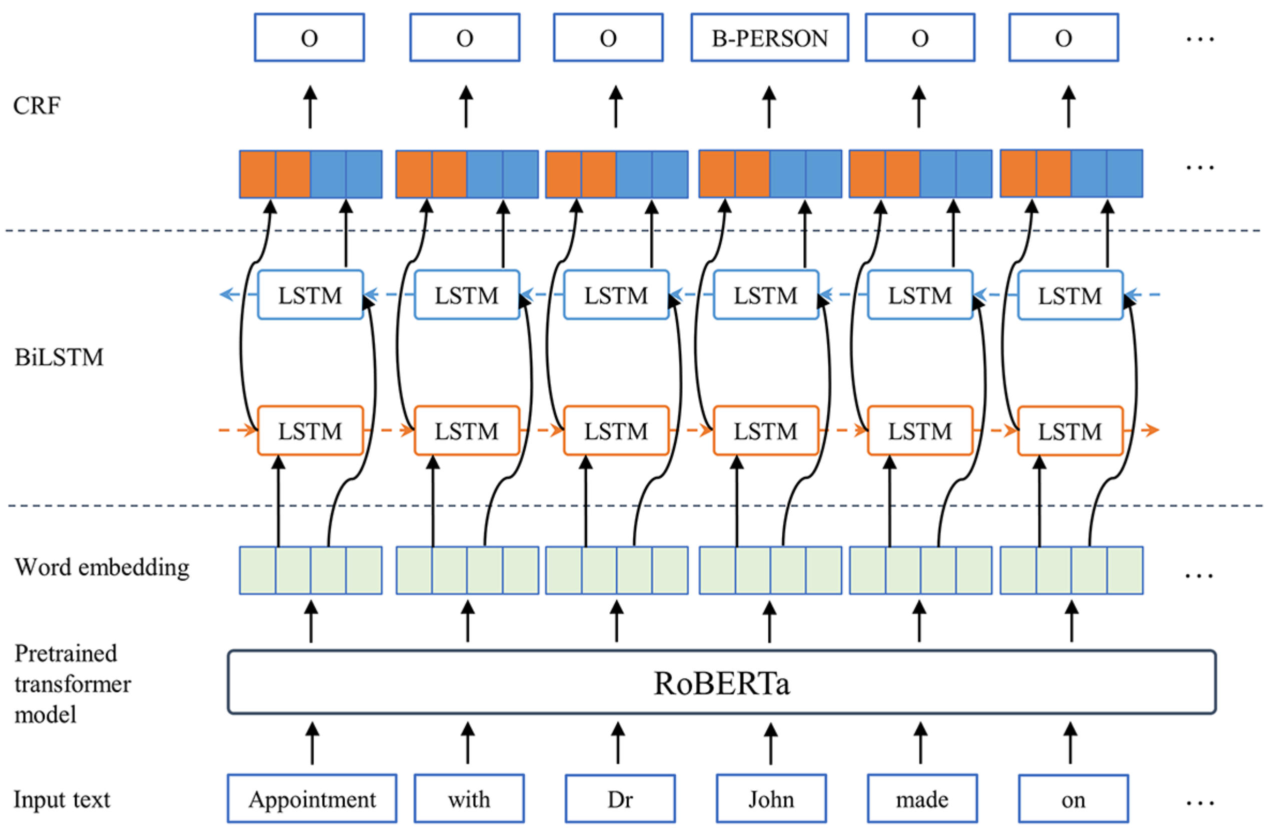
Figure 2. An example of a deidentification project structure and access control. DOB: date of birth; PII: personally identifiable information.



Deep Learning NER Model

In the DEFT system, we used the BiLSTM-CRF architecture to train a NER model that can identify word spans related to specific PII types in the working text. Although an ensemble model may perform better than a single BiLSTM-CRF model, we selected the latter for DEFT considering its competitive performance and faster training time, noting that DEFT may be deployed in settings where computing resources are constrained. The Python library FLAIR [37] was used to implement the BiLSTM-CRF NER model. The pretrained RoBERTa model [38] was selected to generate input representations in our model. Figure 3 shows the model architecture.

Figure 3. Deep learning named entity recognition model architecture. BiLSTM: Bidirectional Long Short-Term Memory; CRF: conditional random field; LSTM: Long Short-Term Memory.



Learning Loop

The core of the DEFT system's workflow is a learning loop, which comprises 3 elements: end user UI, annotated free text, and ML model, as shown in Figure 4. At the beginning of the deidentification work for a new project, there is no ML model in the system. The deidentification process follows the following six steps:

1. Feed raw free text into DEFT system. An initial set of raw free-text files are loaded into the end user UI where the annotators begin to manually tag PII in the raw free text. The end user UI provides a simple and task-focused interface to the annotators so that the annotation can be done quickly and easily.
2. Annotate raw free text or correct preannotated free text. Figure 5 shows a screenshot of the end user UI which contains 4 main areas: file list area, PII type area, annotation work area, and PII list area. First, the annotator selects 1 file in the file list area to load its free text content into the annotation work area and clicks on 1 PII type in the PII type area. Second, the annotator reviews the free text and tags the words using double left click or word spans using click drag-and-release that are related to the selected PII type. The PII entities will be surrounded by colored boxes with the PII type underneath the words. Incorrectly tagged PII entities can be removed by single left click on the entity text. Finally, the tagged PII entities are listed in the PII list area with detailed information about start index, end index, PII entity type, PII entity text, and annotator name. The annotator needs to change the "Edit" mode to "Complete" mode for marking the completion of the annotation work for the current working file. The above mentioned PII entity details, except PII entity text, are automatically saved in the DEFT database so that they can be integrated with the raw free-text file contents to generate the annotated free-text files for model training. Furthermore, not having PII entity text stored in the DEFT system protects the identified data.
3. Train ML models. Once the completed amount of the annotated free-text files reaches the preconfigured model retraining threshold which is the number of clinical notes (eg, 200 clinical notes), the system starts the ML model training process, which splits the annotated data into training, validation, and test sets. The model is automatically trained in the backend and loaded into the system after that.
4. Preannotate raw free text. Another small set of free-text files, that has been preannotated by the trained model from Step 3, are assigned to the annotators to add the PIIs which the model does not preannotate or to correct incorrect PIIs that the model has preannotated, via the end user UI. The added or corrected PII entity details are added into the system database.
5. Iterative ML model training. If the number of the new completed files reaches the retraining threshold, it will trigger the ML model to be retrained on all the annotated free-text files. Steps 2, 3, and 4 are iteratively conducted until the performance of the model meets the specified benchmark for the data set. In the DEFT system, we use

the strict entity-level microaverage F_1-score, which is the primary metric in previous deidentification challenges [16,39,40], to evaluate the performance of the model and deidentification work.

6. Deidentify raw free text. The final model is used to deidentify the remaining free text in the data set, including automatically tagging the PIIs and replacing the PIIs with the special tags. For example, the person names are replaced by “<**PERSON**>” in the deidentified free text.

Figure 4. Learning loop of DEFT system. (1) Feed raw free text into DEFT system; (2) annotate raw free text or correct preannotated free text; (3) train ML models; (4) preannotate raw free text; (5) iterative ML model training; and (6) deidentify raw free text. DEFT: deidentifying free text; ML: machine learning; UI: user interface.

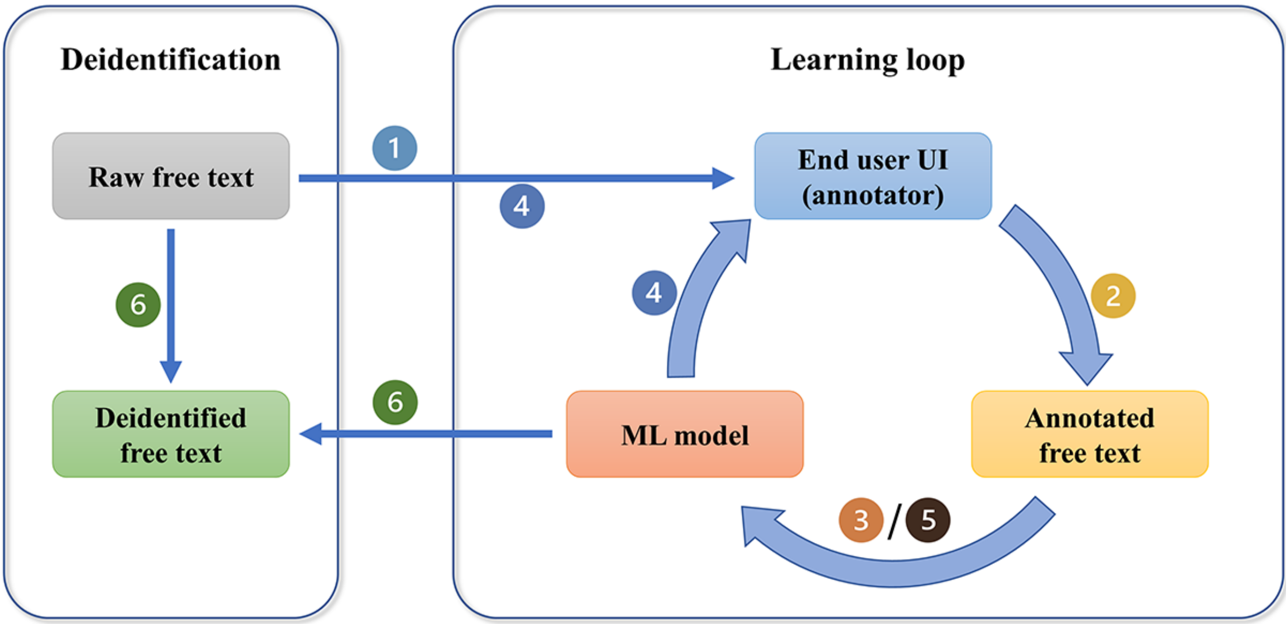
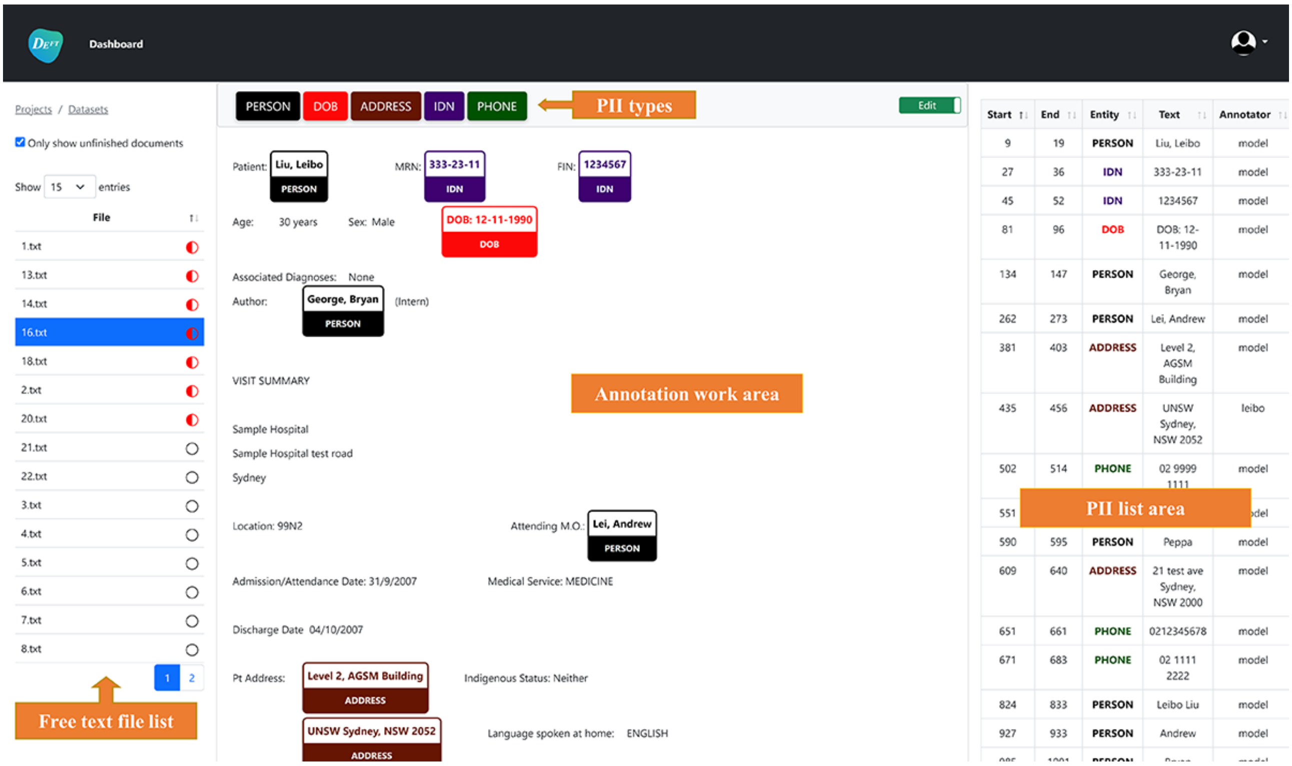


Figure 5. A demo of the end user UI for the annotators. The patient information is dummy data. PII: personally identifiable information; UI: user interface.



DEFT Functionalities

Table 1 describes the main functionalities of the DEFT system which are grouped according to the 2 DEFT web UIs. The admin UI includes the management components of the key elements such as projects, data sets, PII types, and users. The functionalities of the end user UI are mainly focused on the annotation work, for example, tagging the PII, changing the file status, and preannotation. Currently, only txt format data files are supported in DEFT. There are two export options: (1) export the annotated data files in the XML format which include the raw free text and all the PII entities (Figure S1 in Multimedia Appendix 1); the annotated XML data files could be a valuable data source for future deidentification research; and (2) export the deidentified data files in TXT format (Figure S2 in Multimedia Appendix 1). The file contents are the same as the original one except that the PII words are replaced by the special tags.

Table 1. The main functionalities of the DEFT[a] system.

UI[b] and functionality	Description
Admin	
Project management	• Create or delete or modify project information • Configure the project data path • Assign user access right
Data set management	• Create or delete or modify data set information • Import data file names • Export all the annotated data files • Export all the deidentified data files
Data file management	• Create or delete or modify data files • Export single annotated data file • Export single deidentified data file
PII[c] type management	• Create or delete or modify PII types
Model management	• Review all the trained models of the project
User management	• Create or delete or modify users
End user	
Project list	• List accessible projects
Data set list	• List the data sets of the selected projects
Annotation dashboard	• List all the files of the selected data set • Open 1 file • Tag the PII entities in the free text
Auto save	• The add or remove PII actions are saved automatically
Preannotation	• Pretag the possible PII entities by the trained ML[d] model when the users open a file
Hide completed files	• Filter completed files out from the data file list
File status management	• Switch the file status between “Edit” and “Complete”

[a]DEFT: deidentifying free text.

[b]UI: user interface.

[c]PII: personally identifiable information.

[d]ML: machine learning.

Ethics Approval

This study has obtained ethical approval from the South Eastern Sydney Local Health District Human Research Ethics Committee (reference 2019/ETH12625) and the Population Health Services Research Ethics Committee (reference number 2020/ETH01614). The ethics committees allow the data usage for this study without additional consent. The experiments in this study were conducted in the E-Research Institutional Cloud Architecture [41], a secure cloud computing infrastructure for individuals working with sensitive data.

Results

Overview

We selected 2 deidentification systems (MIST and NLM Scrubber) for functionality comparison with DEFT. We chose these systems because MIST has a similar design strategy of "human in the loop" to DEFT, and NLM Scrubber is an accessible open-source off-the-shelf deidentification system. We also considered INCEpTION, ezTag, and Prodigy, which are text annotation systems with interactive learning loops, because they can be used for the deidentification task with extra effort from technical or ML experts. The system features we compared are (1) support for project-based free text file management; (2) support for user access control; (3) support for customized PII (NER) types; (4) support for bulk file import; (5) support for automated preannotation based on a pretrained model; (6) support for auto save of the tagging actions; (7) support for interactive learning loop; (8) support for annotated data export; (9) support for deidentified data export; (10) suitability for nontechnical end users; (11) support for team collaboration; (12) web-based system; (13) central data storage; (14) off-the-shelf; and (15) autotag matches (all occurrences are automatically tagged when annotators tag a PII entity in the whole working free text). Figure 6 shows that DEFT, MIST, and INCEpTION support the user access control functionality, which is important in deploying the systems for team collaboration on the deidentification task. Otherwise, any devices on the same network can access the system via the system's URL to potentially access the identified data. Although ezTag provides a session-based login, it is not a secure way to control access because anyone who has the session URL can access the project freely. Both MIST and Prodigy require end users to use a command-line interface to configure the projects or data sets and retrain the models in the interactive learning loop. MIST and ezTag need to manually trigger the model retraining and preannotation from the command-line interface and the web UI, respectively. When starting the Prodigy system, a path to the free-text data needs to be configured. Therefore, it partially supports the project management and bulk file import. Different from other systems, MIST must manually save each tagging action by the user clicking on the save button. In both the INCEpTION and ezTag systems, importing the free-text data requires transferring the files from the original data location to the specified location using the web UI, while MIST uses the command-line interface to do the same thing. This could be a bottleneck when importing large volumes of data sets due to network delays. Because NLM Scrubber is an off-the-shelf desktop software, most of the comparison functionalities are not supported by it. All the systems except NLM Scrubber cannot be used "out-of-the-box." Only the MIST system provides "Autotag matches" functionality.

Fewer mouse clicks make the annotation process more efficient. We counted the mouse clicks of the add-PII and remove-PII operations for all the selected systems except NLM Scrubber, which is pretrained and doesn't support annotation. The "drag-and-release" action is counted as 1 mouse click. As shown in Table 2, DEFT and Prodigy had the fewest clicks for both operations. MIST and INCEpTION needed 3 mouse clicks to annotate a PII entity. However, MIST provides an "Autotag matches" functionality, which can automatically tag all the occurrences of the same word spans of the selected PII entity in the whole working file. ezTag needed the most mouse clicks to remove a tagged PII entity.

We evaluated the performance of our model by comparing it with 5 benchmark models [14,17,18,23,24] on the 2014 i2b2 data set. The microaverage strict entity–level scores and binary PII token–level scores are reported in Table 3. Our model achieved the highest strict entity–level recall and F_1-scores at 0.9563 and 0.9627, respectively. Table S1 in Multimedia Appendix 1 lists the hyperparameters used for model training. The microaverage scores by i2b2 category for strict entity matching are shown in Table S2 in Multimedia Appendix 1.

Figure 6. Comparison of the selected tools and DEFT. DEFT: deidentifying free text; MIST: MITRE Identification Scrubber Toolkit; NER: named entity recognition; NLM: National Library of Medicine; PII: personally identifiable information.

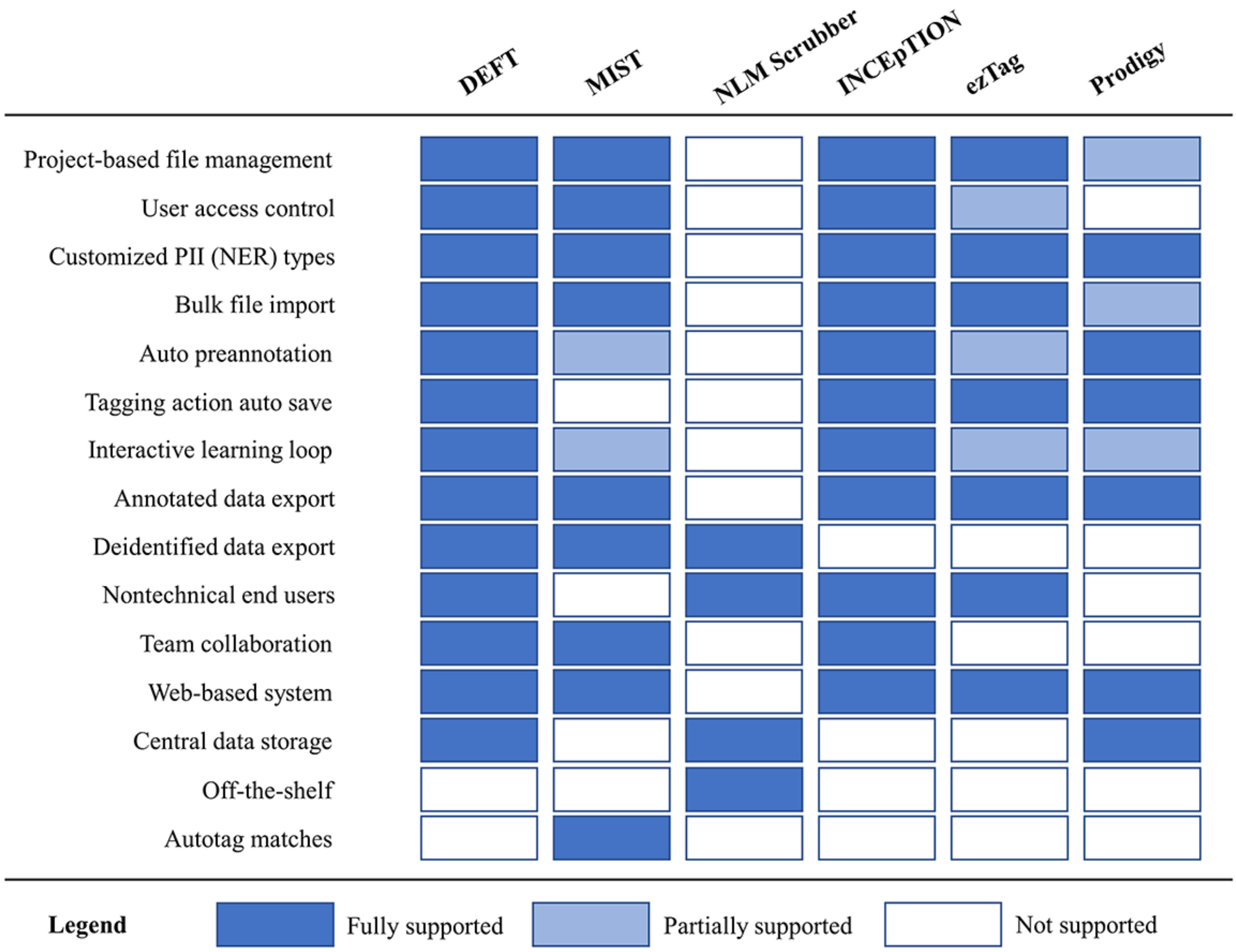


Table 2. Mouse click comparison of add-PII[a] and remove-PII operations.

Operation	DEFT[b]	MIST[c]	NLM[d] Scrubber	INCEpTION	ezTag	Prodigy
Add-PII[e]	1-2	3	N/A[f]	3	1-2	1-2
Remove-PII	1	2	N/A	2	3	1

[a]PII: personally identifiable information.

[b]DEFT: deidentifying free text.

[c]MIST: MITRE Identification Scrubber Toolkit.

[d]NLM: National Library of Medicine.

[e]The mouse-click number of add-PII operation for DEFT, ezTag, and Prodigy can be 1 or 2, depending on whether the relevant PII type has been selected or not.

[f]N/A: not applicable.

Table 3. Microaverage scores comparison on the 2014 i2b2 data set.

Model (reference)	Model architecture	Strict entity-level			Binary PII[a] token-level		
		Precision	Recall	F_1-score	Precision	Recall	F_1-score
Liu et al [24]	Ensemble model	0.9646	0.938	0.9511	0.993	0.9728	0.9828
Kim et al [23]	Ensemble model	0.9704[b]	0.9445	0.9573	0.9916	0.9806	0.9861
Tang et al [17]	BiLSTM-CRF[c] (BERT[d])	0.9599	0.9502	0.955	0.9902	0.9838[b]	0.987[b]
Catelli et al [18]	BiLSTM-CRF	0.9653	0.9506	0.9579	0.991	0.9755	0.9832
Liu et al [14]	Ensemble model	0.969	0.9559	0.9624	0.9943[b]	0.9786	0.9864
Our model	BiLSTM-CRF (RoBERTa)	0.9692	0.9563[b]	0.9627[b]	0.9892	0.9766	0.9829

[a]PII: personally identifiable information.

[b]The best result for each metric.

[c]BiLSTM-CRF: Bidirectional Long Short-Term Memory-Conditional Random Field.

[d]BERT: Bidirectional Encoder Representations from Transformers.

Use Case

To evaluate its efficiency and performance, we used DEFT to deidentify clinical notes for the project CardiacAI [42]. CardiacAI is a prospective data repository that collects EMR data for patients with cardiac issues who are admitted to a group of participating hospitals in New South Wales, Australia. The EMR data are linked to state-wide hospital and emergency department visits, the state's death registry, as well as mobile health remote monitoring data. The aim of this data collection is to enable collaborative research, facilitate the use and rapid translation of state-of-the-art tools and technologies, and ultimately drive improvement in patient care and outcomes. The CardiacAI data repository holds data from January 1, 2017 onwards. The repository currently holds EMRs of 44,201 patients with 61,721 individual hospitalizations to a cardiac, cardiothoracic, or vascular surgical specialty. There are 2,087,737 clinical documents including discharge summaries, progress notes, and other semistructured forms with an average of 34 (SD 44) documents per hospitalization. Discharge summaries have undergone some rule-based deidentification where the document header, which contains structured identifier fields, has been removed. However, embedded identifiers remain within the unstructured text.

The DEFT system was deployed on a workspace with 32 GB memory and 8 vCPU in the E-Research Institutional Cloud Architecture [41] environment. A project named "CardiacAI," and 2 data sets named "Discharge Summary" and "Progress Notes" were created in the system. According to the data content and research requirements, 5 PII types (PERSON, identification number [IDN], DOB, PHONE, ADDRESS) were defined for the project. At the beginning, we randomly selected 400 clinical notes including 200 discharge summaries and 200 progress notes to be manually annotated using the DEFT end user UI by 1 annotator who is an experienced health-domain data analyst. An intermediate NER model was automatically trained by DEFT using the 400 clinical notes with a split ratio of 0.8, 0.1, and 0.1 for training, validation, and test, respectively. Table 4 shows the model performance on the test set. Nearly 81% PII entities were person type and only 2 were address type. The model achieved a micro-F_1-score of 0.9432.

To compare annotation time with and without auto-preannotation, we randomly selected another 200 clinical notes that consisted of 100 discharge summaries and 100 progress notes and split them into 10 rounds, of which each round contained 10 discharge summaries and 10 progress notes. We also considered the text length when splitting the data to make sure that the compared rounds had similar total words as other rounds. The annotator manually tagged the PIIs in the first 5 rounds without auto-preannotation and continued to complete the last 5 rounds with auto-preannotation enabled. Each round was completed in a single session and the annotation time was recorded. As shown in Table 5, the total time of 5 rounds with-preannotation was about 58 minutes, which was approximately 40% less than the total time of 5 rounds without preannotation (~98 minutes). Moreover, the annotation of the last 5 rounds (with preannotation) was much quicker than the one of the first 5 rounds, despite there being more PII entities in the last 5 rounds. Through the 5-round comparison, the preannotation by the trained model sped up the annotation work by 43% (95% CI 33.9-52.1; P<.001).

To compare manual annotation performance without and with preannotation, a gold standard corpus was created by having 2 annotators independently review and correct the annotations of the 200 clinical notes. All disagreements were resolved through a consensus meeting between the 2 annotators. The performance results are shown in Table 6. Manual annotation with preannotation by the pretrained model achieved higher overall microaverage recall and F_1-scores by 0.013 and 0.0053, respectively. Although there were a few false negative and positive errors on the PHONE and ADDRESS types, manual annotation with preannotation demonstrated higher accuracy on the PERSON type (the majority type within the data set) than manual annotation without preannotation.

We reconfigured the retraining threshold to 200 to trigger the system to retrain the model using the whole 600 clinical notes. The results are shown in Table 7. Although the test sets were different between this training and the previous training, we do observe that the overall F_1-score increased by 0.0075, compared to the previous intermediate model trained on 400 clinical notes. The new model improved the F_1-score of DOB and PHONE PII types by large margins of 0.1367 and 0.1298, respectively. Moreover, the recall, which is the most important metric for the deidentification tasks, was improved for each PII type, except ADDRESS.

Table 4. The NER[a] model performance on the test set from the corpus with 400 clinical notes.

PII[b] type	Precision	Recall	micro-F_1-score	PII entity number
PERSON	0.9657	0.9505	0.958	444
IDN[c]	1	0.9649	0.9821	57
DOB[d]	0.875	0.8077	0.84	26
PHONE	0.7273	0.7273	0.7273	22
ADDRESS	0.3333	0.5	0.4	2
Overall	0.9519	0.9347	0.9432	551

[a]NER: named entity recognition.
[b]PII: personally identifiable information.
[c]IDN: identification number.
[d]DOB: date of birth.

Table 5. Annotation time comparison.

Without preannotation					With preannotation				
Round	Clinical note number	Annotation time (min)	PII[a] entity number	Word number	Round	Clinical note number	Annotation time (min)	PII entity number	Word number
1	20	13:03	81	2687	1a	20	6:28	113	2774
2	20	15:56	117	4089	2a	20	7:55	184	4153
3	20	19:08	155	5198	3a	20	9:31	179	5369
4	20	20:24	139	7133	4a	20	12:13	185	7404
5	20	29:05	189	11,503	5a	20	22:08	300	12,304
Total	100	97:36	681	30,610	Total	100	58:15	961	32,004

[a]PII: personally identifiable information.

Table 6. Performance comparison without and with preannotation.

Without preannotation				With preannotation			
PII[a] Type	Precision	Recall	micro-F_1-score	PII Type	Precision	Recall	micro-F_1-score
PERSON	0.9980	0.9675	0.9825	PERSON	0.9983	0.9851	0.9917
IDN[b]	1	1	1	IDN	1	1	1
DOB[c]	1	1	1	DOB	1	1	1
PHONE	1	1	1	PHONE	0.9667	0.9667	0.9667
ADDRESS	1	1	1	ADDRESS	0.8571	1	0.9231
Overall	0.9985[d]	0.9756	0.9869	Overall	0.9958	0.9886[d]	0.9922[d]

[a]PII: personally identifiable information.
[b]IDN: identification number.
[c]DOB: date of birth.
[d]The best result for each metric between performance without and that with preannotation.

Table 7. The NER[a] model performance on the test set from the corpus with 600 clinical notes.

PII[b] type	Precision	Recall	micro-F_1-score	PII entity number
PERSON	0.9423	0.9583	0.9502	528
IDN[c]	1	0.9762	0.988	84
DOB[d]	0.9844	0.9692	0.9767	65
PHONE	0.8571	0.8571	0.8571	35
ADDRESS	1	0.5	0.6667	6
Overall	0.9487	0.9526	0.9507	718

[a]NER: named entity recognition.
[b]PII: personally identifiable information.
[c]IDN: identification number.
[d]DOB: date of birth.

In order to evaluate the deidentification efficiency, we used the trained model to deidentify CardiacAI clinical notes that summarized the encounters of patients who have had heart failure, containing on average about 170 words. The whole process took about 95 hours to export 280,785 deidentified clinical notes (about 0.82 seconds per clinical note). There are on average about 5 PII entities per clinical note.

Discussion

Principal Findings

We have presented the design and implementation of DEFT, a simple web-based deidentification system with interactive learning loop. DEFT provides a task-focused web UI to end users so that annotation of PII can be done quickly and easily. In addition, an admin web UI is provided to manage the projects, users, and team collaboration on the deidentification task. The DEFT system can be deployed on a central server to provide deidentification services to multiple teams within the same network environment at the same time through user access control functionality.

The results of the functionality comparison showed that our DEFT system is better than the selected comparison systems on the deidentification task, in terms of the ability to customize the task, ease and security of management of users and files, automation of the interactive learning loop, and central data storage.

DEFT can easily customize PII types at the project level using the admin web UI. It provides flexibility for determining the level of deidentification according to risk assessment and analysis requirements of the project. The selected systems which we compared with DEFT have similar customization functionality, with the exception of NLM Scrubber. However, only INCEpTION and ezTag provide a web UI for this function. MIST and Prodigy uses either a configuration file or command line to predefine PII types. NLM Scrubber uses fixed HIPAA-compliant PII types, which restrict the adoption of the system outside of a HIPAA-covered country or organization [6].

DEFT has a 2-level file management structure (ie, project and data set) which allows the users to organize the data more flexibly, compared to the 1-level structure (ie, project) of the other systems. For example, the deidentification project may have many different types of clinical notes. Users can create different data sets for each type of clinical note. Like INCEpTION, DEFT can easily manage user credentials and assign user access to the project level through the admin web UI. However, in the MIST system, user management and access control need to be done from a command-line interface, and all the users share 1 project access key. When starting up the Prodigy system, the users need to configure an input data source. Therefore, the different project users cannot work on the system at the same time. Moreover, there is no user access control functionality in Prodigy. ezTag uses a session-based login to generate a unique URL for each user to manage the user access. However, the session can be accessed by anyone with the unique URL, so it is not suitable for annotating sensitive information in EMR free-text data.

DEFT supports a fully automated interactive learning loop. The model training process is triggered when the newly annotated files reach the preset retraining threshold. The retrained model is automatically loaded for the relevant project. When the annotator opens a nonannotated file, the model automatically preannotates it to provide suggestions for the annotator. In contrast, MIST, Prodigy, and ezTag need human manual operations during the learning loop. MIST retrains a model and preannotates the free-text files via a command-line interface. Moreover, the system needs to be restarted to load the retrained model. Similarly, Prodigy also needs different command lines to perform retraining and preannotation operations. Users of ezTag need to click the “Auto Annotation” and “Train” buttons to trigger the relevant tasks. The real-world use case study we present demonstrated that DEFT’s annotation speed can be increased by 43% with the automated preannotation in the learning loop. Furthermore, the DEFT system can use only 600 annotated clinical notes to achieve good performance with an F_1-score of 0.9507, which is greater than 0.95, the rule-of-thumb benchmark for evaluating the reliability of a deidentification system [16,40]. The low scores for ADDRESS PII entity as shown in Tables 4 and 7 may be caused by the lack of training samples, as there were only 44 and 55 ADDRESS entities in the training sets of 400 and 600 clinical notes, respectively. The model performance will continually improve as the amount of annotated free-text data increases. We also evaluated the annotation time and accuracy of manual annotation without and with preannotation. Using preannotation during the manual annotation process resulted in a time savings of 43% and a micro-F_1-score improvement of 0.005 compared to manual annotation performed without preannotation.

The design of DEFT allows the raw free-text data to be stored in a central data storage location which is under the control of the researchers or data custodians. The system-related data, such as project and data set information, users, PII index and type details, and the free-text file names, are saved in DEFT's database. All the raw data and exported data are managed centrally by the data manager. This not only protects the identified data from unauthorized access, but also avoids transferring large volumes of data through the network. A data source path is required when starting up the Prodigy system and therefore data can be stored at a central location. However, none of the other systems support this.

In the traditional annotation process, more than 2 annotators are needed to generate a gold standard data set. The interannotator agreement is measured to evaluate the annotation reliability between different annotators. Although a single annotator was used in this case study, the interactive learning loop enabled the model and the annotator to suggest and correct each other to continually improve the model performance. However, DEFT can easily perform the traditional annotation process. For example, the data manager creates a data set named "gold standard" in the project. In total, 2 annotators are required to tag the PIIs from all the files in the data set. Another annotator as an audit changes the file status to completion after reviewing the annotation results. The NER model will be trained on the gold-standard data set.

Limitations

In DEFT, we preannotate all the PII from the free text using the trained NER model. The annotators need to review the whole text to correct the preannotations. In contrast, INCEpTION and Prodigy combine preannotation with active learning which queries the user for feedback (accept, reject, or skip) on the annotation suggestions that are most informative to the model [30]. Active learning can achieve rapid and accurate annotation with less annotation time [43]. Integration of active learning within the learning loop could enhance the next version of DEFT. Moreover, we will explore fine-tuning pretrained transformer language models for NER in DEFT to improve the accuracy of deidentification. Furthermore, implementing "Autotag matches" in DEFT could have the potential to further decrease manual effort involved in the annotation process. Another limitation of DEFT is that it relies solely on deep learning models. The system's performance could potentially be enhanced by incorporating a hybrid method that combines knowledge-based methods (such as predefined regular expressions or knowledge dictionaries) and deep learning. The evaluation results in this study are derived from clinical notes which were annotated by a single annotator. This may introduce potential reliability issues when assessing the deidentification performance. To address this, an annotation process, which is manually annotated by 3 annotators (2 annotators and 1 adjudicator), is needed to be introduced in the next version of DEFT.

Conclusions

DEFT is a web-based deidentification system, which is designed for health domain researchers and data custodians to easily deidentify free-text data in EMRs with the support of an interactive learning loop. End users can perform all the operations through a well-designed web UI. DEFT has many good features to help manage and organize the deidentification tasks. In particular, the central data storage feature ensures that the identified data are protected properly in a central location without transfer through the network. The real-world use case demonstrated that DEFT can speed up the annotation process and quickly complete the deidentification work for large volumes of data with high reliability. The source code of the DEFT system is available at GitHub [44].

Acknowledgments

LL designed and implemented the DEFT system and its web application, conducted all the experiments, and wrote the paper. VB conducted all the annotation work using DEFT system. VB and BG provided the information about the CardiacAI project. OP-C, AN, VB, and LJ provided the supervision and suggestions and reviewed the paper. We thank the South Eastern Sydney Local Health District and the Illawarra Shoalhaven Local Health District for their contribution to the CardiacAI project. The study was funded by the Australian government and the Commonwealth Industrial and Scientific Research Organisation through Australian Government Research Training Program scholarship and Commonwealth Industrial and Scientific Research Organisation top up scholarship.

Data Availability

The public 2014 i2b2 data set repository can be accessed with registration on the website of Department of Biomedical Informatics at Harvard Medical School [45]. Access to CardiacAI data repository must be assessed and approved by the CardiacAI Data Governance Committee. Please visit the CardiacAI website [46] for more information about eligible research proposals.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemetary material.
[DOCX File , 468 KB - ijmr_v12i1e46322_app1.docx]

References

1. Ford E, Carroll JA, Smith HE, Scott D, Cassell JA. Extracting information from the text of electronic medical records to improve case detection: a systematic review. J Am Med Inform Assoc 2016;23(5):1007-1015 [FREE Full text] [doi: 10.1093/jamia/ocv180] [Medline: 26911811]
2. Kotevski DP, Smee RI, Field M, Nemes YN, Broadley K, Vajdic CM. Evaluation of an automated presidio anonymisation model for unstructured radiation oncology electronic medical records in an Australian setting. Int J Med Inform 2022;168:104880. [doi: 10.1016/j.ijmedinf.2022.104880] [Medline: 36272315]
3. Mercorelli L, Nguyen H, Gartell N, Brookes M, Morris J, Tam CS. A framework for de-identification of free-text data in electronic medical records enabling secondary use. Aust Health Rev 2022;46(3):289-293. [doi: 10.1071/AH21361] [Medline: 35546422]
4. The HIPAA privacy rule. U.S. Department of Health & Human Services. 2022. URL: https://www.hhs.gov/hipaa/for-professionals/privacy/index.html [accessed 2020-04-08]
5. The privacy act 1988. Federal Register of Legislation. Australia; 2022. URL: https://www.legislation.gov.au/Details/C2022C00321 [accessed 2022-10-29]
6. Hanauer D, Aberdeen J, Bayer S, Wellner B, Clark C, Zheng K, et al. Bootstrapping a de-identification system for narrative patient records: cost-performance tradeoffs. Int J Med Inform 2013;82(9):821-831. [doi: 10.1016/j.ijmedinf.2013.03.005] [Medline: 23643147]
7. Sweeney L. Replacing personally-identifying information in medical records, the scrub system. Proc AMIA Annu Fall Symp 1996:333-337 [FREE Full text] [Medline: 8947683]
8. Neamatullah I, Douglass MM, Lehman LWH, Reisner A, Villarroel M, Long WJ, et al. Automated de-identification of free-text medical records. BMC Med Inform Decis Mak 2008;8:32 [FREE Full text] [doi: 10.1186/1472-6947-8-32] [Medline: 18652655]
9. Menger V, Scheepers F, van Wijk LM, Spruit M. DEDUCE: a pattern matching method for automatic de-identification of Dutch medical text. Telemat Inform 2018;35(4):727-736. [doi: 10.1016/j.tele.2017.08.002]
10. He B, Guan Y, Cheng J, Cen K, Hua W. CRFs based de-identification of medical records. J Biomed Inform 2015;58(Suppl):S39-S46 [FREE Full text] [doi: 10.1016/j.jbi.2015.08.012] [Medline: 26315662]
11. Yadav S, Ekbal A, Saha S, Bhattacharyya P. Deep learning architecture for patient data de-identification in clinical records. 2016 Presented at: Proceedings of the Clinical Natural Language Processing Workshop (ClinicalNLP); December 11-17 2016; Osaka, Japan p. 32-41 URL: https://aclanthology.org/W16-4206/
12. Jiang Z, Zhao C, He B, Guan Y, Jiang J. De-identification of medical records using conditional random fields and long short-term memory networks. J Biomed Inform 2017;75S:S43-S53 [FREE Full text] [doi: 10.1016/j.jbi.2017.10.003] [Medline: 29032162]
13. Yang X, Lyu T, Li Q, Lee CY, Bian J, Hogan WR, et al. A study of deep learning methods for de-identification of clinical notes in cross-institute settings. BMC Med Inform Decis Mak 2019;19(Suppl 5):232 [FREE Full text] [doi: 10.1186/s12911-019-0935-4] [Medline: 31801524]
14. Liu L, Perez-Concha O, Nguyen A, Bennett V, Jorm L. De-identifying Australian hospital discharge summaries: an end-to-end framework using ensemble of deep learning models. J Biomed Inform 2022;135:104215 [FREE Full text] [doi: 10.1016/j.jbi.2022.104215] [Medline: 36195240]
15. Lee HJ, Wu Y, Zhang Y, Xu J, Xu H, Roberts K. A hybrid approach to automatic de-identification of psychiatric notes. J Biomed Inform 2017;75S:S19-S27 [FREE Full text] [doi: 10.1016/j.jbi.2017.06.006] [Medline: 28602904]
16. Stubbs A, Filannino M, Uzuner Ö. De-identification of psychiatric intake records: overview of 2016 CEGS N-GRID shared tasks track 1. J Biomed Inform 2017;75S:S4-S18 [FREE Full text] [doi: 10.1016/j.jbi.2017.06.011] [Medline: 28614702]
17. Tang B, Jiang D, Chen Q, Wang X, Yan J, Shen Y. De-identification of clinical text via Bi-LSTM-CRF with neural language models. AMIA Annu Symp Proc 2019;2019:857-863 [FREE Full text] [Medline: 32308882]
18. Catelli R, Casola V, De Pietro G, Fujita H, Esposito M. Combining contextualized word representation and sub-document level analysis through Bi-LSTM+CRF architecture for clinical de-identification. Knowl Based Syst 2021;213:106649. [doi: 10.1016/j.knosys.2020.106649]
19. Catelli R, Gargiulo F, Casola V, De Pietro G, Fujita H, Esposito M. Crosslingual named entity recognition for clinical de-identification applied to a COVID-19 Italian data set. Appl Soft Comput 2020;97:106779 [FREE Full text] [doi: 10.1016/j.asoc.2020.106779] [Medline: 33052197]
20. Ahmed A, Abbasi A, Eickhoff C. Benchmarking modern named entity recognition techniques for free-text health record deidentification. AMIA Jt Summits Transl Sci Proc 2021;2021:102-111 [FREE Full text] [Medline: 34457124]
21. Johnson AEW, Bulgarelli L, Pollard TJ. Deidentification of free-text medical records using pre-trained bidirectional transformers. New York, NY, United States: Association for Computing Machinery; 2020 Presented at: CHIL '20: Proceedings

of the ACM Conference on Health, Inference, and Learning; April 2-4, 2020; Toronto Ontario Canada p. 214-221. [doi: 10.1145/3368555.3384455]
22. Marimon M, Gonzalez-Agirre A, Intxaurrondo A, Rodriguez H, Martin JL, Villegas M, et al. Automatic de-identification of medical texts in Spanish: the MEDDOCAN track, corpus, guidelines, methods and evaluation of results. IberLEF@ SEPLN 2019:618-638 [FREE Full text]
23. Kim Y, Heider P, Meystre S. Ensemble-based methods to improve de-identification of electronic health record narratives. AMIA Annu Symp Proc 2018;2018:663-672 [FREE Full text] [Medline: 30815108]
24. Liu Z, Tang B, Wang X, Chen Q. De-identification of clinical notes via recurrent neural network and conditional random field. J Biomed Inform 2017;75S:S34-S42 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.023] [Medline: 28579533]
25. Neves M, Ševa J. An extensive review of tools for manual annotation of documents. Brief Bioinform 2021;22(1):146-163 [FREE Full text] [doi: 10.1093/bib/bbz130] [Medline: 31838514]
26. Giachelle F, Irrera O, Silvello G. MedTAG: a portable and customizable annotation tool for biomedical documents. BMC Med Inform Decis Mak 2021;21(1):352 [FREE Full text] [doi: 10.1186/s12911-021-01706-4] [Medline: 34922517]
27. Stenetorp P, Pyysalo S, Topić G, Ohta T, Ananiadou S, Tsujii J. BRAT: a web-based tool for NLP-assisted text annotation. 2012 Presented at: Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics; April 23-27 2012; Avignon, France p. 102-107 URL: https://aclanthology.org/E12-2021/
28. Syed M, Al-Shukri S, Syed S, Sexton K, Greer ML, Zozus M, et al. DeIDNER corpus: annotation of clinical discharge summary notes for named entity recognition using BRAT tool. Stud Health Technol Inform 2021;281:432-436 [FREE Full text] [doi: 10.3233/SHTI210195] [Medline: 34042780]
29. Kwon D, Kim S, Wei CH, Leaman R, Lu Z. ezTag: tagging biomedical concepts via interactive learning. Nucleic Acids Res 2018;46(W1):W523-W529 [FREE Full text] [doi: 10.1093/nar/gky428] [Medline: 29788413]
30. Klie JC, Bugert M, Boullosa B, de Castilho RE, Gurevych I. The inception platform: machine-assisted and knowledge-oriented interactive annotation. 2018 Presented at: Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations; August 20-26, 2018; Santa Fe, New Mexico, USA p. 5-9 URL: https://aclanthology.org/C18-2002/
31. Explosion. Prodigy. 2022. URL: https://prodi.gy/ [accessed 2022-10-29]
32. Aberdeen J, Bayer S, Yeniterzi R, Wellner B, Clark C, Hanauer D, et al. The MITRE identification scrubber toolkit: design, training, and assessment. Int J Med Inform 2010;79(12):849-859. [doi: 10.1016/j.ijmedinf.2010.09.007] [Medline: 20951082]
33. Heider PM, Obeid JS, Meystre SM. A comparative analysis of speed and accuracy for three off-the-shelf de-identification tools. AMIA Jt Summits Transl Sci Proc 2020;2020:241-250 [FREE Full text] [Medline: 32477643]
34. Amazon Comprehend Medical. Amazon Web Services. 2022. URL: https://aws.amazon.com/comprehend/medical/ [accessed 2022-10-30]
35. CliniDeID. Clinacuity. 2022. URL: https://www.clinacuity.com/clinideid/ [accessed 2022-10-30]
36. The National Library of Medicine's scrubber. National Library of Medicine. 2022. URL: https://scrubber.nlm.nih.gov/ [accessed 2022-10-30]
37. Akbik A, Bergmann T, Blythe D, Rasul K, Schweter S, Vollgraf R. FLAIR: an easy-to-use framework for state-of-the-art NLP. 2019 Presented at: Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations); June 2-7, 2019; Minneapolis, Minnesota p. 54-59 URL: https://aclanthology.org/N19-4010/ [doi: 10.18653/v1/n19-4010]
38. Biomedical and clinical language models. GitHub. 2023. URL: https://github.com/facebookresearch/bio-lm/blob/main/README.md [accessed 2023-04-01]
39. Uzuner Ö, Luo Y, Szolovits P. Evaluating the state-of-the-art in automatic de-identification. J Am Med Inform Assoc 2007;14(5):550-563 [FREE Full text] [doi: 10.1197/jamia.M2444] [Medline: 17600094]
40. Stubbs A, Kotfila C, Uzuner Ö. Automated systems for the de-identification of longitudinal clinical narratives: overview of 2014 i2b2/UTHealth shared task track 1. J Biomed Inform 2015;58(Suppl):S11-S19 [FREE Full text] [doi: 10.1016/j.jbi.2015.06.007] [Medline: 26225918]
41. ERICA—E-research institutional cloud architecture. University of New South Wales. 2022. URL: https://research.unsw.edu.au/erica [accessed 2022-12-30]
42. Blake V, Jorm L, Yu J, Lee A, Gallego B, Ooi SY. The Cardiac Analytics and Innovation (CardiacAI) data repository: an Australian data resource for translational cardiovascular research. ArXiv. Preprint posted online on April 18 2023 2023:1-37 [FREE Full text] [doi: 10.48550/arXiv.2304.09341]
43. Dobbie S, Strafford H, Pickrell WO, Fonferko-Shadrach B, Jones C, Akbari A, et al. Markup: a web-based annotation tool powered by active learning. Front Digit Health 2021;3:598916 [FREE Full text] [doi: 10.3389/fdgth.2021.598916] [Medline: 34713086]
44. Liu L. DEFT source code. GitHub. 2022. URL: https://github.com/leiboliu/deft [accessed 2023-02-07]
45. Department of Biomedical Informatics (DBMI). Harvard Medical School. 2023. URL: https://portal.dbmi.hms.harvard.edu/ [accessed 2023-05-25]
46. Cardiac Analytics and Innovation. 2022. URL: https://www.cardiacai.org [accessed 2022-12-26]

Abbreviations

BiLSTM-CRF: Bidirectional Long Short-Term Memory-Conditional Random Field
DEFT: deidentifying free text
DOB: date of birth
EMR: electronic medical record
HIPAA: Health Insurance Portability and Accountability Act
IDN: identification number
MIST: MITRE Identification Scrubber Toolkit
ML: machine learning
NER: named entity recognition
NLM: National Library of Medicine
PII: personally identifiable information
UI: user interface

Edited by A Mavragani; submitted 08.02.23; peer-reviewed by Y Tchouka, S Meystre; comments to author 05.04.23; revised version received 31.05.23; accepted 24.07.23; published 25.08.23.

Please cite as:
Liu L, Perez-Concha O, Nguyen A, Bennett V, Blake V, Gallego B, Jorm L
Web-Based Application Based on Human-in-the-Loop Deep Learning for Deidentifying Free-Text Data in Electronic Medical Records: Development and Usability Study
Interact J Med Res 2023;12:e46322
URL: https://www.i-jmr.org/2023/1/e46322
doi:10.2196/46322
PMID:37624624

Original Paper

Postgraduate-Year-1 Residents' Perceptions of Social Media and Virtual Applicant Recruitment: Cross-sectional Survey Study

Daniel L Plack[1], MD; Arnoley S Abcejo[1], MD; Molly B Kraus[2], MD; J Ross Renew[3], MD; Timothy R Long[1], MD; Emily E Sharpe[1], MD

[1]Department of Anesthesiology and Perioperative Medicine, Mayo Clinic, Rochester, MN, United States
[2]Department of Anesthesiology and Perioperative Medicine, Mayo Clinic, Phoenix, AZ, United States
[3]Department of Anesthesiology and Perioperative Medicine, Mayo Clinic, Jacksonville, FL, United States

Corresponding Author:
Emily E Sharpe, MD
Department of Anesthesiology and Perioperative Medicine
Mayo Clinic
200 1st St SW
Rochester, MN, 55905
United States
Phone: 1 507 284 9839
Email: sharpe.emily@mayo.edu

Abstract

Background: The dissemination of information about residency programs is a vital step in residency recruitment. Traditional methods of distributing information have been printed brochures, websites, in-person interviews, and increasingly, social media. Away rotations and in-person interviews were cancelled, and interviews were virtual for the first time during the COVID-19 pandemic.

Objective: The purpose of our study was to describe postgraduate-year-1 (PGY1) residents' social media habits in regard to residency recruitment and their perceptions of the residency programs' social media accounts in light of the transition to virtual interviews.

Methods: A web-based 33-question survey was developed to evaluate personal social media use, perceptions of social media use by residency programs, and perceptions of the residency program content. Surveys were sent in 2021 to PGY1 residents at Mayo Clinic in Arizona, Florida, and Minnesota who participated in the 2020-2021 interview cycle.

Results: Of the 31 program directors contacted, 22 (71%) provided permission for their residents to complete the survey. Of 219 residents who received the survey, 67 (30%) completed the survey. Most respondents applied to a single specialty, and greater than 61% (41/67) of respondents applied to more than 30 programs. The social media platforms used most regularly by the respondents were Instagram (42/67, 63%), Facebook (36/67, 54%), and Twitter (22/67, 33%). Respondents used the program website (66/67, 99%), residents (47/67, 70%), and social media (43/67, 64%) as the most frequent resources to research programs. The most commonly used social media platforms to research programs were Instagram (38/66, 58%), Twitter (22/66, 33%), and Doximity (20/66, 30%). The type of social media post ranked as most interesting by the respondents was "resident life outside of the hospital." In addition, 68% (39/57) of the respondents agreed or strongly agreed that their perception of a program was positively influenced by the residency program's social media account.

Conclusions: In this multispecialty survey of PGY1 residents participating in the 2020-2021 virtual interview season, respondents preferred Instagram to Twitter or Facebook for gathering information on prospective residency programs. In addition, the program website, current residents, and social media platforms were the top-ranked resources used by prospective applicants. Having an up-to-date website and robust social media presence, particularly on Instagram, may become increasingly important in the virtual interview environment.

(Interact J Med Res 2023;12:e42042) doi:10.2196/42042

KEYWORDS

COVID-19; resident match; social media; Twitter; Instagram; virtual interview; residency; medical education; dissemination; residency program; residency recruitment

Introduction

The dissemination of information about residency programs is a vital step in residency recruitment. Traditional methods of distributing information have been printed brochures, websites, in-person interviews, and increasingly, social media. The use of social media has expanded in medicine over recent years and has become a valuable resource in influencing resident recruitment, graduate medical education, professional development, and academic scholarship [1,2]. In 2012, only 15% of residency programs had a social media presence [3]. This contrasts with recent studies, which found 63%, 61%, and 55% of anesthesiology, pediatric, and general surgery residency programs, respectively, had some form of residency social media account in October 2020 [4-6]. In 2018, a Plastic and Reconstructive Surgery residency program conducted a survey of all applicants to their program and found 96% of respondents had at least one social media account and 73% followed a residency program on social media [7].

Moreover, the demand for social media integration into education and residency continues to evolve. In the 2021 Match cycle at a single institution, most residency applicants not only followed programs' social media accounts, but also preferred certain social media platforms—namely Instagram [8]. Twitter has found an established domain in medical education and dissemination of information across many, if not most, specialties [9]. More recently, Instagram use, with its ability to leverage picto- and videographic material without limitations on characters, has modernized some residency education curricula [10].

During the COVID 19 pandemic, away rotations and in-person interviews were canceled, and residency recruitment and interviews were virtual for the first time in history. A 2021 study of urology residency programs reported that program use of social media increased from 26%-50% prior to 2020 to 51%-75% in 2021 [11]. These investigators also described changing attitudes toward social media use by applicants, with a greater emphasis being placed on such resources in 2021 [11]. Of interest, the Coalition on Physician Accountability has made recommendations that, going forward, all specialties use virtual interviews to ensure equity and reduce the cost of travel and time away from school [12].

Our study sought to assess postgraduate-year-1 (PGY1) residents' social media use for the purposes of researching residency programs and the residents' perceptions of the effects of residency program–based social media accounts during the 2020-2021 residency recruitment cycle.

Methods

Survey Development

We conducted a survey of PGY1 residents in Mayo Clinic School of Graduate Medical Education residency training programs at all Mayo Clinic sites, including Rochester, Minnesota; Jacksonville, Florida; and Scottsdale, Arizona.

A web-based 33-question survey was developed based on surveys used in anesthesiology and urology to evaluate residency applicants' perceptions of social media [2,11]. The survey included questions regarding demographic characteristics, type of medical school attended, specialties applied to, and number of residency program applications submitted. In addition, the survey included questions on personal social media habits, the social media platforms candidates used to evaluate residencies, the type of content they were seeking, attitudes about programs' use of social media, and contact with programs over social media. The study survey was piloted with 5 PGY1 residents at residency programs unrelated to Mayo Clinic who assessed the survey for clarity, duration, and ease of reading. Recommendations for survey changes from the pilot group were incorporated into the final survey (Multimedia Appendix 1).

Recruitment

All 31 Mayo Clinic School of Graduate Medical Education residency program directors (PDs) were invited to share the survey with their residents (289 PGY1 residents in total). PDs could then decide if they would distribute the survey to their residents or provide the authors with permission to contact the residents directly to recruit them for the study. Residents were included if they were PGY1 and had participated in the 2020-2021 residency recruitment cycle. They were excluded if PDs declined to participate, or if they did not participate in the 2020-2021 residency recruitment cycle. A public link that generated anonymous responses in REDCap (Research Electronic Data Capture; version 11.1.120; Vanderbilt University) was provided for the PDs to send to their residents directly; otherwise, after approval by the PD, a recruitment email was sent to the PGY1 residents with a link to a REDCap survey . REDCap is a secure, web-based application designed to support data capture for research studies. For programs that allowed the research group to email the PGY1 residents directly to solicit participation, 4 weekly reminders were sent to nonresponders. For those who received the survey directly from their PD, no reminder emails were sent. Surveys were completed between September 4 and November 10, 2021.

Statistical Analysis

Nondemographic data were presented as a 5-point Likert scale. Respondent demographics and data regarding personal social media use and residency social media are presented as numbers and percentages.

Ethical Considerations

After approval by the Mayo Clinic Educational Research Committee, the study was reviewed and deemed exempt from ethics approval by the Mayo Clinic Institutional Review Board in Rochester, Minnesota.

Results

Response and Demographics

Of the 31 PDs who were contacted, 4 PDs elected to send the survey to their residents directly, and 18 PDs provided permission for the research group to contact their residents directly. A total of 219 of 289 (76%) PGY1 residents received the survey: 26 residents received the survey directly from their PDs, and 193 residents received the survey via REDCap. Of the 26 surveys that were sent out via PDs, 7 residents completed the survey. Of the 193 surveys that were sent directly, 60 residents completed the survey. This resulted in a total of 67 complete surveys for a response rate of 31% (67/219). Minor differences are present in the denominators of the data because not all survey respondents answered each question.

Demographics of respondents and residency application information are summarized in Table 1. Most respondents (57/67, 85.1%) applied to a single specialty. A majority of respondents applied to more than 30 residency programs: 61% (41/67) applied to more than 30 residency programs and 30% (20/67) applied to more than 60 residency programs. Of 27 medical and surgical specialties, 18 (67%) had at least one respondent. Regarding virtual or in-person interviews, only 2 individuals attended both in-person and virtual interviews, with the vast majority (65/67, 97%) having attended virtual interviews only.

Table 1. Respondent demographics (N=67).

Characteristics	Values, n (%)
Age on match day (year)	
<25	0 (0)
25-29	54 (81)
30-35	10 (15)
36-40	3 (5)
>40	0 (0)
Gender	
Female	34 (51)
Male	32 (48)
Another gender identity	0 (0)
Prefer to not identify	1 (2)
Type of medical school	
US allopathic	46 (69)
US osteopathic	10 (15)
International	11 (16)
Number of specialties applied	
1	57 (85)
2	10 (15)
>2	0 (0)
Specialty Applied	
Internal medicine	22 (33)
Family medicine	10 (15)
Anesthesiology	8 (12)
Pediatrics	8 (12)
Psychiatry	5 (8)
Dermatology	3 (5)
Emergency medicine	3 (5)
Pathology	3 (5)
Orthopedic Surgery	2 (3)
Otolaryngology	2 (3)
Surgery	2 (3)
Neurological surgery	1 (2)
Neurology	1 (2)
Obstetrics and gynecology	1 (2)
Ophthalmology	1 (2)
Physical medicine and rehabilitation	1 (2)
Plastic surgery	1 (2)
Urology	1 (2)
Number of individual programs applied	
<10	1 (2)
11-20	12 (18)
21-30	13 (19)

Characteristics	Values, n (%)
31-40	8 (12)
41-50	6 (9)
51-60	7 (10)
>60	20 (30)
Mayo Clinic site matched	
Florida or Jacksonville	2 (3)
Minnesota or Rochester	58 (87)
Arizona or Scottsdale	7 (10)
Race or ethnic origin	
Black or African American	2 (3)
American Indian or Alaskan Native	1 (2)
White	45 (67)
Asian	14 (21)
Native Hawaiian or Other Pacific Islander	0 (0)
Multiracial	0 (0)
Other or unknown	4 (6)
Ethnic Origin Hispanic (ie, a person of Hispanic ethnicity who may be of any race)	5 (8)
Prefer to not respond	2 (3)

PGY1 Residents' Social Media Use

Respondents were asked on which social media platforms they had accounts and which accounts they used regularly (ie, more than once a week; Table 2). Although 48% (32/67) of applicants had LinkedIn accounts and 54% (36/67) had Doximity accounts, only 3% (2/67) and 6% (4/67) indicated they regularly used LinkedIn and Doximity, respectively. Of 67 respondents, 42 (63%) regularly used Instagram, 54% (36/67) regularly used Facebook, and 33% (22/67) regularly used Twitter; 99% (66/67) of respondents used the program website, 70% (47/67) used the residents, and 64% (43/67) used social media as the most frequent resources to research programs. Regarding the specific social media or web-based forums used to research programs, 58% (38/66) of respondents most frequently used Instagram, 33% (22/66) used Twitter, and 30% (20/66) used Doximity to research programs (Table 3). Of the 52 respondents who began to follow a residency program on social media, 11 (21%) began to follow the program before recruitment season, 35 (67%) began to follow it during the recruitment season, and 6 (12%) began to follow it after the recruitment season.

Table 2. Respondents' social media use.

Characteristics	Values, n (%)
Social media accounts (n=67)	
Facebook	53 (79)
Instagram	50 (75)
Twitter	37 (55)
Snapchat	36 (54)
Doximity	36 (54)
LinkedIn	32 (48)
Reddit	23 (34)
TikTok	14 (21)
Discord	10 (15)
Student Doctor Network	4 (6)
Other	2 (3)
None	2 (3)
Tumblr	1 (2)
Yammer	1 (2)
Flickr	0 (0)
Social media regular use (n=67)	
Instagram	42 (63)
Facebook	36 (54)
Twitter	22 (33)
Snapchat	19 (28)
Reddit	18 (27)
TikTok	11 (16)
None	5 (8)
Doximity	4 (6)
Discord	3 (5)
LinkedIn	2 (3)
Tumblr	1 (2)
Other	1 (2)
Flickr	0 (0)
Student Doctor Network	0 (0)
Yammer	0 (0)
Communication with any residency program on social media during 2020-2021 interview cycle (n=67)	
Yes	9 (13)
No	58 (87)
Follow individual faculty or resident accounts (n=67)	
Yes	24 (36)
No	43 (64)
Reason to stop following residency program social media accounts (n=66)	
I matched at a different program	31 (47)
Too many posts	2 (3)
Content not interesting to me	2 (3)

Characteristics	Values, n (%)
Lack of professionalism	1 (2)
Poor social media etiquette	0 (0)
I did not stop following any programs	29 (44)
Other	1 (2)[a]

[a]Free-text response to “Other” selection: “was not invited to interview.”

Table 3. Respondents’ social media use related to the residency program’s accounts.

Characteristics	N (%)
Resources used to research prospective programs (n=67)	
Residency program websites	66 (99)
Residents	47 (70)
Social media (eg, Facebook, Twitter, and Instagram)	43 (64)
Other medical students	41 (61)
Web-based town hall style event hosted by residency program	35 (52)
Web-based forums (eg, Student Doctor Network and Reddit)	33 (49)
Doximity	30 (45)
Attending physicians	27 (40)
YouTube	18 (27)
Other	4 (6)
None	0 (0)
Social media platforms used to research prospective programs (n=66)	
Instagram	38 (58)
Twitter	22 (33)
Doximity	20 (30)
Reddit	18 (27)
Facebook	16 (24)
None	13 (20)
Student Doctor Network	6 (9)
Discord	5 (8)
Other	2 (3)
TikTok	0 (0)
Snapchat	0 (0)
LinkedIn	0 (0)
Yammer	0 (0)
Tumblr	0 (0)
Flickr	0 (0)

PGY1 Residents’ Perception on Residency’s Social Media Accounts

Instagram was the social media platform used most frequently to research prospective programs (Table 3). Respondents (27/55, 49%) rated “resident life outside the hospital” as the type of social media post that was the most interesting during the 2020-2021 interview cycle (Figure 1). The popularity of other categories is ranked in Table 4. The program’s Instagram account (46/67, 69%) was rated the most helpful social media platform or media type for dissemination of future information for prospective residents, followed by the program’s Instagram stories (34/67, 51%) and Twitter account (33/67, 49%; Table 5).

A total of 68% (39/57) of respondents agreed or strongly agreed that residency programs’ social media accounts positively influenced their perception of a program, whereas 53% (31/59) disagreed or strongly disagreed that lack of social media presence negatively influenced their perception of a program; in other words, lack of social media presence was not seen in a negative light (Table 6). In addition, 67% (41/61) of respondents agreed or strongly agreed that residency programs should not initiate contact with applicants over social media.

Figure 1. The highest ranked type of social media post from residency programs during the 2020-2021 interview cycle.

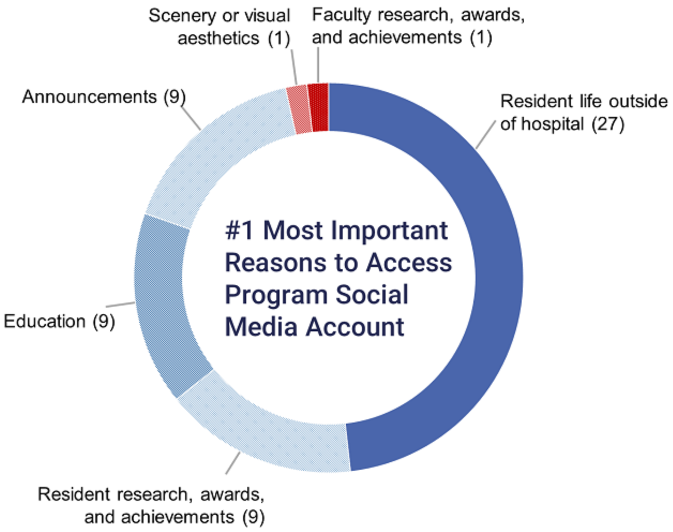


Table 4. Rank order of types of residency program's social media posts.

Characteristics	1 (most Interesting)	2	3	4	5	6 (least Interesting)
Resident life outside the hospital (n=55)	27 (49)	11 (20)	3 (5)	6 (11)	3 (6)	5 (9)
Resident research, awards, and achievements (n=56)	9 (16)	12 (21)	10 (18)	10 (18)	11 (20)	4 (7)
Faculty research, awards, and achievements (n=55)	1 (2)	6 (11)	13 (24)	11 (20)	10 (18)	14 (26)
Education (n=54)	9 (17)	12 (22)	11 (20)	12 (22)	7 (13)	3 (6)
Announcements (n=57)	9 (16)	7 (12)	9 (16)	4 (7)	17 (30)	11 (19)
Scenery or visual aesthetic (n=60)	1 (2)	8 (13)	11 (18)	12 (20)	10 (17)	18 (30)

Table 5. Social media platform and media type helpful for future dissemination of program information (N=67).

Media type	Values, n (%)
Residency program's Instagram account	46 (69)
Residency program's Instagram stories	34 (51)
Residency program's Twitter account	33 (49)
Instagram 1-day account takeover by residents	32 (48)
Residency program's Facebook account	17 (25)
YouTube	17 (25)
Residency program's Twitter "Tweetorials"	15 (22)
Instagram 1-day account takeover by faculty	14 (21)
Residency program's Instagram live sessions with faculty or residents	11 (16)
None	10 (15)
TikTok	5 (8)
Individual residents' Twitter account	4 (6)
Individual faculty's Twitter account	4 (6)
Other	2 (3)

Table 6. Respondents' reactions related to the residency programs' social media use.

Statement	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
"When I look at residency programs on social media, I am looking for specific information" (n=56)	2 (4)	10 (18)	18 (32)	25 (44)	1 (2)
"When I look at residency programs on social media, I am browsing for relaxation, not interested in a specific piece of information" (n=58)	5 (9)	16 (28)	14 (24)	23 (40)	0 (0)
"Information I found specifically on social media positively influenced my perception of a program" (n=57)	2 (4)	5 (9)	11 (19)	32 (56)	7 (12)
"Information I found specifically on social media negatively influenced my perception of a program" (n=58)	4 (7)	18 (31)	19 (33)	15 (26)	2 (3)
"The lack of a social media presence of a program negatively influenced my perception of a program" (n=59)	11 (19)	20 (34)	9 (15)	18 (31)	1 (2)
"A residency program's social media account is a good representation of how their residency program actually is" (n=59)	4 (7)	14 (24)	30 (51)	10 (17)	1 (2)

Professionalism on Social Media

In regard to professional use of social media, 60% (40/67) of respondents indicated that they had received education on how to maintain a professional social media account, and 73% (49/67) indicated they considered their professional reputation when posting. However, 65% (39/60) of respondents disagreed or strongly disagreed that residency programs should use social media networks to evaluate applicants.

Discussion

Principal Results

We conducted a multispecialty survey of PGY1 residents who participated in the 2020-2021 virtual interview season in a large sponsoring institution that spans 3 separate geographical regions of the country to ascertain social media use and perceptions on residency programs' social media accounts. Approximately, 61% (41/67) of surveyees applied to 31 or more residency programs.

We found an overwhelming preference for Instagram over Twitter or Facebook for gathering information on residency programs and gaining insight into the daily life of a resident. At the start of the pandemic, Instagram was already the primary social media platform preferred by the generation aged 18-34 years [13]. Moreover, this Instagram preference trends through other specialties, education pathways, and outreach efforts [14-16]. Understanding Instagram use and behaviors therein can positively influence a residency program's social media reach and user interaction—defined by likes, views, and shares [17]. The authors speculate Instagram is the current preferred social media platform for investigating potential programs because it has easily accessible photographic, videographic, and free simple editing functionality. These functions facilitate prompt display of attractive program attributes through incorporation of captions, music, and interactive displays for any user.

Comparison With Prior Work

Our results demonstrating the utility of social media for resident recruitment is consistent with prior efforts. Czawlytko et al [18] demonstrated that 71% of prospective radiology residents viewed a program's social media account to learn about the program, without specifying the specific platform. Although Instagram was the most frequently used tool in our effort, Cox et al [19] have demonstrated significant expansion in general surgery programs' Twitter use since the start of the pandemic. Similarly, emergency medicine residency programs increased social media use by 34% in 2020, and the authors felt the emphasis on web-based platforms in the setting of the pandemic was a significant catalyst [20].

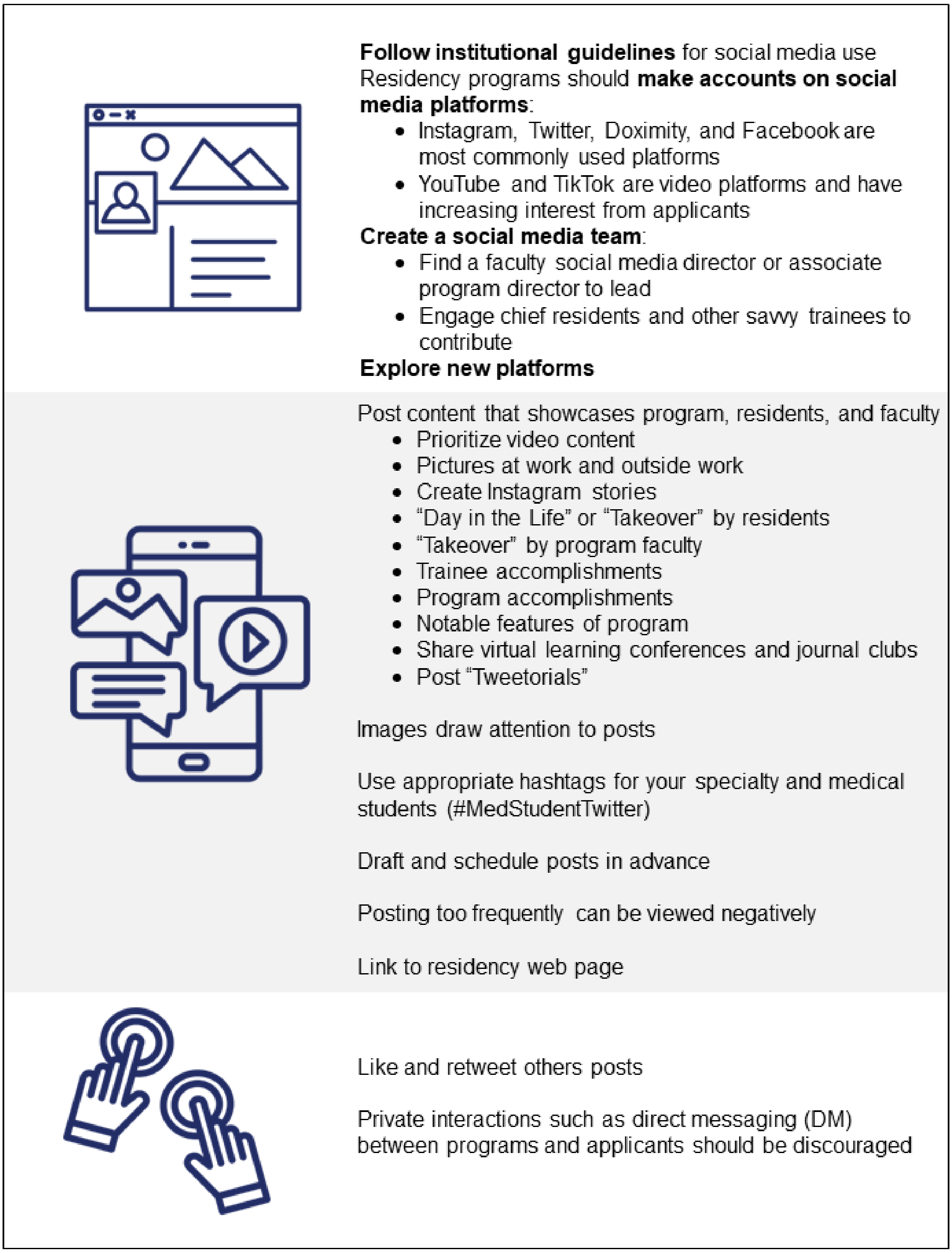
The majority of our respondents considered professionalism while maintaining their own personal social media accounts—a phenomenon that is also in line with past efforts [21]. Although others have demonstrated the use of social media to exchange peer medical information [22], our results support the notion that such platforms are similarly important to promoting residency programs to potential candidates.

Recommendations

Given the results of our survey, programs should consider focusing their attention on information distributed through their website, residents, and social media. Considering that students used Instagram, Twitter, and Doximity with the highest frequency to research programs, residency programs might direct their efforts to those social media platforms. Where the in-person campus tour previously took place on interview day, social media can help bridge the gap by helping prospective students answer the question "What would it be like to work and live here day-to-day?" Students were most likely to follow the programs' social media accounts during the interview cycle, highlighting that period as the most important time for residency program accounts to be active, posting new content, and engaging its followers. For content, respondents were most interested in "resident life outside the hospital." Posts about local life, extracurricular activities, and a "day in the life of a resident" may be best received by prospective students. Programs should note that although prospective residents are consumers of social media, they do not want to be contacted by programs over social media. Although the presence of an active social media account can positively influence a recruit's attitude toward a program, we found that the reciprocal—the lack of an account—equally may or may not have a negative impact on their attitude toward a program.

As program directors and residency leadership plan for future virtual interview cycles, a strong social media presence will be important to reach applicants who regularly use social media. Social media alone will not replace other elements of the interview and recruitment process, including website, emails, information sessions, and video interviews, but having a robust social media presence as a fundamental element to residency recruitment will become increasingly important [23]. Recommendations for creating a profile, type of content to post, and interacting on social media are provided in Figure 2.

Figure 2. Recommendations for creating a profile, type of content to post, and interacting on social media.



Limitations

This study has several limitations. A pilot survey was conducted, but more formal survey validation tools were not used. The survey response rate of 30.6% may reflect a nonresponse bias. Additionally, the low absolute response rate from various specialties prevented any meaningful analysis of trends between specialties. Participants were asked into which specialty they matched, but some may have recorded their current status as an intern in medicine or surgery rather than their specialty after internship. Recall bias may be present, given that participants reflected on what avenues for information they used after being prompted by specific answer choices. The term "Doximity" was used rather than the term "Doximity Residency Navigator," and respondents may interpret these as different entities. The study also evaluates residents at a single institution in 3 geographically distinct areas, with the geographic Midwest being vastly overrepresented. The data may be influenced by the geographic location into which the respondents matched. Future efforts should evaluate the implementation of social media recruitment strategies and associated outcomes, such as applicant numbers, Match Day success, and resident perceptions of program fit.

Conclusions

In this multispecialty survey of residents who participated in the 2020-2021 virtual interview season, we found the program website, residents, and social media, specifically Instagram, were the top ways that prospective residents researched residency programs. If recruitment is to remain virtual, programs need to be innovative in their efforts to showcase their learning environment to prospective residents. Social media platforms, particularly Instagram, and to a lesser extent, Twitter, can be useful tools. The virtual residency interview has changed the dynamic for programs and applicants alike.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Survey sent to postgraduate-year-1 residents at Mayo Clinic in Arizona, Florida, and Minnesota asking for personal social media use, social media use in regard to residency programs, and perceptions of residency programs' content during the 2020-2021 interview cycle.

[PDF File (Adobe PDF File), 55 KB - ijmr_v12i1e42042_app1.pdf]

References

1. Economides JM, Choi YK, Fan KL, Kanuri AP, Song DH. Are we witnessing a paradigm shift? PRS GO 2019;7(8):e2288. [doi: 10.1097/gox.0000000000002288]
2. Renew JR, Ladlie B, Gorlin A, Long T. The impact of social media on anesthesia resident recruitment. JEPM 2019 Jan 1;21(1):1-7. [doi: 10.46374/volxxi-issue1-renew]
3. Schweitzer AL, Ross JT, Klein CJ, Lei KY, Mackey ER. An electronic wellness program to improve diet and exercise in college students: a pilot study. JMIR Res Protoc 2016 Feb 29;5(1):e29 [FREE Full text] [doi: 10.2196/resprot.4855] [Medline: 26929118]
4. Clay Pruett J, Deneen K, Turner H, Kozar T, Singh NP, King TW, et al. Social media changes in pediatric residency programs during COVID-19 pandemic. Acad Pediatr 2021 Sep;21(7):1104-1107. [doi: 10.1016/j.acap.2021.06.004] [Medline: 34126258]
5. Lee DC, Kofskey AM, Singh NP, King TW, Piennette PD. Adaptations in anesthesiology residency programs amid the COVID-19 pandemic: virtual approaches to applicant recruitment. BMC Med Educ 2021 Aug 31;21(1):464 [FREE Full text] [doi: 10.1186/s12909-021-02895-2] [Medline: 34465325]
6. Rohde SC, White EM, Yoo PS. Residency program use of social media in the COVID-19 era: an applicant's perspective. J Surg Educ 2021 Jul;78(4):1066-1068. [doi: 10.1016/j.jsurg.2020.12.011] [Medline: 33358933]
7. Steele TN, Galarza-Paez L, Aguilo-Seara G, David LR. Social media impact in the Match: a survey of current trends in the United States. Arch Plast Surg 2021 Jan 20;48(1):107-113 [FREE Full text] [doi: 10.5999/aps.2020.00836] [Medline: 33503753]
8. Naaseh A, Thompson S, Tohmasi S, Wiechmann W, Toohey S, Wray A, et al. Evaluating applicant perceptions of the impact of social media on the 2020-2021 residency application cycle occurring during the COVID-19 pandemic: survey study. JMIR Med Educ 2021 Oct 05;7(4):e29486 [FREE Full text] [doi: 10.2196/29486] [Medline: 34591779]
9. Panda A, Sharma A, Dundar A, Packard A, Aase L, Kotsenas A, et al. Twitter use by academic nuclear medicine programs: pilot content analysis study. JMIR Form Res 2021 Nov 08;5(11):e24448 [FREE Full text] [doi: 10.2196/24448] [Medline: 34747708]

10. Koenig JFL, Buentzel J, Jung W, Truemper L, Wurm-Kuczera RI. Using Instagram to enhance a hematology and oncology teaching module during the COVID-19 pandemic: cross-sectional study. JMIR Med Educ 2021 Nov 15;7(4):e30607 [FREE Full text] [doi: 10.2196/30607] [Medline: 34779777]
11. Ho P, Margolin E, Sebesta E, Small A, Badalato GM. #AUAMatch: The impact of COVID-19 on social media use in the urology residency match. Urology 2021 Aug;154:50-56 [FREE Full text] [doi: 10.1016/j.urology.2021.05.019] [Medline: 34033828]
12. Advancing health care education and practice. Coalition for Physician Accountability. URL: https://physicianaccountability.org/ [accessed 2023-02-23]
13. 5WPR 2020 Consumer Culture Report. URL: https://www.5wpr.com/new/wp-content/uploads/pdf/5W_consumer_culture_report_2020final.pdf [accessed 2023-02-23]
14. Abbas MJ, Jildeh TR, Khalil LS, Buckley P, Mumuni SP, Washington KJ, et al. Social media use continues to increase among orthopaedic residency programs in the United States. Arthrosc Sports Med Rehabil 2021 Dec;3(6):e1761-e1767 [FREE Full text] [doi: 10.1016/j.asmr.2021.08.004] [Medline: 34977631]
15. Gleason A, Singh G, Keffer L, Nepomnayshy D. General surgery going viral: current trends in social media utilization by general surgery residency programs. J Surg Educ 2021 Nov;78(6):e62-e67. [doi: 10.1016/j.jsurg.2021.10.014] [Medline: 34782270]
16. Johnson JL, Bhatia N, West DL, Safdar NM. Leveraging social media and web presence to discuss and promote diversity, equity, and inclusion in radiology. J Am Coll Radiol 2022 Jan;19(1 Pt B):207-212. [doi: 10.1016/j.jacr.2021.09.020] [Medline: 35033313]
17. Plack DL, Sharpe EE, Wanderman RL, Ripoll JG, Abcejo AS. Getting the first thousand-optimizing instagram residency content to increase followers during the COVID-19 pandemic. J Educ Perioper Med 2021 Jun 1;23(2):E660 [FREE Full text] [doi: 10.46374/volxxiii_issue2_abcejo] [Medline: 34104674]
18. Czawlytko C, Smith E, Awan O, Resnik C, Hossain R. The effect of virtual interviews and social media on applicant decision-making during the 2020-2021 resident match cycle. Acad Radiol 2022 Jun;29(6):928-934 [FREE Full text] [doi: 10.1016/j.acra.2021.05.028] [Medline: 34244069]
19. Cox JS, Wehrle CJ, Mejias C, Devarakonda AK, McKenzie JA, Arora TK. General surgery twitter during COVID-19: tweets, trends, and implications for recruitment strategies. Am Surg 2021 Dec 22:31348211060414. [doi: 10.1177/00031348211060414] [Medline: 34937400]
20. Baldwin CS, DeMarinis AR, Singh NP, Khoury CA. Evaluation of Emergency Medicine Residency Programs' use of social media in the setting of the COVID‐19 pandemic. JACEP Open 2022 Jan 06;3(1):1-5. [doi: 10.1002/emp2.12637]
21. Viskić J, Jokić D, Marelić M, Machala Poplašen L, Relić D, Sedak K, et al. Social media use habits, and attitudes toward e-professionalism among medicine and dental medicine students: a quantitative cross-sectional study. Croat Med J 2021 Dec 31;62(6):569-579 [FREE Full text] [doi: 10.3325/cmj.2021.62.569] [Medline: 34981689]
22. Hazzam J, Lahrech A. Health care professionals' social media behavior and the underlying factors of social media adoption and use: quantitative study. J Med Internet Res 2018 Nov 07;20(11):e12035 [FREE Full text] [doi: 10.2196/12035] [Medline: 30404773]
23. Bhayani RK, Fick L, Dillman D, Jardine DA, Oxentenko AS, O'Glasser A. Twelve tips for utilizing residency program social media accounts for modified residency recruitment. MedEdPublish 2020 Aug 26;9(1):178. [doi: 10.15694/mep.2020.000178.1]

Abbreviations

PGY1: postgraduate-year-1
PD: program director
REDCap: Research Electronic Data Capture

Edited by A Mavragani; submitted 19.08.22; peer-reviewed by M Kapsetaki, M Earl, R George, E Malinzak; comments to author 25.01.23; revised version received 10.02.23; accepted 13.02.23; published 21.03.23.

Please cite as:
Plack DL, Abcejo AS, Kraus MB, Renew JR, Long TR, Sharpe EE
Postgraduate-Year-1 Residents' Perceptions of Social Media and Virtual Applicant Recruitment: Cross-sectional Survey Study
Interact J Med Res 2023;12:e42042
URL: https://www.i-jmr.org/2023/1/e42042
doi:10.2196/42042
PMID:36943340

Review

Supporting Midwifery Students During Clinical Practice: Results of a Systematic Scoping Review

Hafaza Amod[1], PhD; Sipho Wellington Mkhize[1], PhD

[1]School of Nursing and Public Health, College of Health Sciences, University of KwaZulu-Natal, Durban, South Africa

Corresponding Author:
Hafaza Amod, PhD
School of Nursing and Public Health
College of Health Sciences, University of KwaZulu-Natal
238 Masizi Kunene Road
Umbilo
Durban, 4041
South Africa
Phone: 27 31 260 4000 ext 3037
Email: amodh@ukzn.ac.za

Abstract

Background: Midwifery educators are highly concerned about the quality of clinical support offered to midwifery students during clinical placement. The unpreparedness of midwifery practitioners in mentorship roles and responsibilities affects the competence levels of the next-generation midwives being produced.

Objective: The aim of this paper is to highlight various clinical support interventions to support midwifery students globally and propose a framework to guide mentorship training in South Africa.

Methods: This paper adopts a mixed methodology approach guided by the Arksey and O'Malley framework. Keywords such as midwifery students, clinical support, mentorship, preceptorship, and midwifery clinical practice were used during the literature search. The review included primary quantitative, qualitative, and mixed methods design papers published between 2010 and 2020, and studies on clinical support interventions available to midwifery students during clinical placement. The search strategy followed a 3-stage system of title, abstract, and full-text screening using inclusion and exclusion criteria. All included papers were quality appraised with a mixed methods appraisal tool. Extracted data were analyzed and presented in themes following a thematic content analysis approach.

Results: The screening results attained 10 papers for data extraction. In total, 7 of the 10 (70%) studies implemented a mentorship training program, 2 (20%) used a training workshop, and 1 (10%) used an intervention guide to support midwifery students in clinical practice. Of these 10 papers, 5 were qualitative, 4 mixed methods, and 1 quantitative in approach. In total, 9 of the 10 (90%) studies were conducted in high-income countries with only 1 study done in Uganda but supported by the United Kingdom. The quality of included papers ranged between 50% and 100%, showing moderate to high appraisal results. Significant findings highlighted that the responsibility of mentorship is shared between key role players (midwifery practitioners, students, and educators) and thus a 3-fold approach to mentorship. Mentorship training and support are essential to strengthen the clinical support of midwifery students during placement. The main findings produced 2 main themes and 2 subthemes each. The main themes included strengthening partnerships and consultation; and providing mentor support through training. The 4 subthemes were: establishing stronger partnerships between nursing education institutions and clinical facilities; improving consultation between midwifery educators, practitioners, and students; the quality of clinical support depends on the training content; and the training duration and structure. Hence, the researchers proposed these subthemes in a framework to guide mentorship training.

Conclusions: Mentorship training and support for midwifery practitioners will likely strengthen the quality of midwifery clinical support. A framework to guide mentorship training will encourage midwifery educators to develop and conduct mentorship training with ease. More studies using quantitative approaches in research and related to midwifery clinical support are required in African countries.

International Registered Report Identifier (IRRID): RR2-10.2196/29707

(Interact J Med Res 2023;12:e36380) doi:10.2196/36380

KEYWORDS

clinical support; mentorship training program; midwifery clinical education; midwife; midwifery; mentor; mentorship; clinical education training; clinical support; midwifery student; South Africa; Africa; framework; medical education

Introduction

Supporting students placed at various clinical facilities is an essential component of learning during clinical practice. In clinical education programs, such as midwifery, clinical placement is a perfect opportunity to achieve the skills necessary to become a safe and competent practitioner. The quality of midwifery students graduating is the responsibility of both midwifery practitioners and educators [1]. The midwifery module in the undergraduate nursing program is a hands-on module that expects midwifery students to spend most of their module time in clinical practice [2]. Therefore, midwifery educators rely on midwifery practitioners to clinically prepare students for role-taking, hoping that new graduates become competent, safe, and independent practitioners.

Recent challenges in the health care system and its effects on the quality of clinical support offered to midwifery students have become a significant concern for midwifery educators globally [3-5]. High student enrollment rates have subsequently increased the teaching workloads of midwifery educators [6]. Additionally, challenges related to developments in nursing programs and the unexpected disruptions experienced during the COVID-19 pandemic have increased midwifery educators' academic and clinical responsibilities. The corresponding increase in the number of students placed at clinical sites has also become a challenge for midwifery practitioners. Uncertainties about mentoring roles, negative feelings about teaching, time constraints, and dire staff shortages and resources have negatively affected the clinical support of midwifery students [3,6,7].

However, global efforts using various clinical support models, such as mentorship, preceptorship, and clinical supervision, have shown positive outcomes on midwifery students' clinical learning and support [8-11]. Mentorship is a highly recommended means to provide the support that students require [9,11,12], and mentorship training programs to support midwifery practitioners in mentoring roles have shown numerous benefits globally [11,13]. Mentorship in maternity units is a direct relationship between the mentor (midwifery practitioner) and the mentee (midwifery student). Midwifery practitioners who are either not trained or inadequately supported in mentorship roles experience difficulties in supervising students [12,14,15] and, as a result, feel unprepared to share the responsibility of mentoring students [2,16,17]. Lack of support for mentors in maternity departments is a global challenge [18]. Clear guidance on how to conduct mentorship training and a need to identify interventions to support midwifery practitioners in mentorship is likely to improve the clinical support of midwifery students in clinical practice. Disregarding mentorship improvements poses the risk of employing unprepared and unsafe practitioners who are detrimental to health care outcomes. This review aims to identify clinical support interventions for midwifery students globally and develop a framework to guide mentorship training in South Africa.

Methods

Study Design

This systematic scoping review followed a protocol developed to analyze the evidence on interventions to strengthen the clinical support of midwifery students during clinical placements [4]. The review followed a population, concept, and context framework [19]. The review focused on the concept of the clinical support available to midwifery students (population) in clinical placements in a global context.

Identifying the Research Question

This review answers the research question what interventions are available to strengthen the current clinical support for midwifery students globally? By identifying and analyzing the clinical support interventions available on a global platform, the researchers desired to integrate these interventions to develop a new framework to guide mentorship training in South Africa.

Search Strategy

The retrieval of records was through database searching conducted between September 2019 and March 2020. Hence, this review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart [20]. The search strategy included keywords midwifery students, clinical support, mentorship, and midwifery clinical practice. The search was refined to English and confined to the last 10 years (January 2010 to August 2020) to ensure only current and updated clinical support interventions for this review. The review included a hand search through the main published papers and citations from the "related literature" list.

Electronic databases used for this review included (1) EBSCOHost (CINAHL, MEDLINE, Health Source: Nursing/Academic Edition) using boolean terms such as midwifery and clinical support, midwifery and mentorship, midwifery and clinical supervision, and midwifery and preceptorship; (2) PubMed and Science Direct included MeSH (Medical Subject Headings) terms such as midwifery students and clinical support, or midwifery students and mentorship, or midwifery students and clinical supervision; and (3) Google and Google Scholar used keywords such as midwifery students in undergraduate nursing programs, midwifery students and clinical support, mentorship in midwifery, and midwifery clinical practice and clinical supervision.

The librarian assisted with retrieving full-text papers not found on the website. All researchers kept an electronic record of retrieved papers.

Study Selection Process

The search strategy followed a 3-stage system of title screening, abstract screening, and full-text screening. The selection included qualitative, quantitative, and mixed methods papers published in peer-review journals. All selected papers were exported to an EndNote (Clarivate, 2020) library. Duplications were removed from the list. The primary investigator and an independent collaborator screened all saved abstracts using a standardized Google Form as a tool. Both the primary investigator and the independent collaborator applied the inclusion and exclusion criteria developed for this search.

Inclusion criteria include (1) only primary studies conducted between 2010 and 2020; (2) papers that used qualitative, quantitative, or mixed methods approaches; (3) papers that present programs, training, or interventions related to clinical support such as mentorship, preceptorship, and clinical supervision; and (4) papers available in the English language. Exclusion criteria include (1) studies that did not include a program, training, or intervention; (2) papers that were reviews; and (3) studies related to nurses in the general, community, and psychiatry nursing disciplines.

All papers selected from the abstract-screening stage were eligible for a full-text paper screening process using another standardized Google Form. Both the primary investigator and the research collaborator worked independently to screen all retrieved papers and compiled a report of both the abstract and full-text screening. A third reviewer (the research supervisor) was available to resolve any discrepancies; however, there were none at the time. The involvement of 3 reviewers prevented bias in the selection of papers. All selected papers from the screening process were saved in an EndNote software folder.

Quality Appraisal

All included studies were quality appraised using a mixed methods appraisal tool [21]. The intention was to retrieve high-quality papers related to the topic, avoid reading flawed literature, and prevent bias or untrustworthy information, which is the essence of conducting a systematic scoping review.

Data Charting and Analysis

This review identified papers, which included clinical support interventions. The data charting variables, included (1) the author's name, (2) the year of publication, (3) the aims of the study, (4) intervention outcomes, and (5) the most significant findings.

A desktop review of included papers was followed by a thematic content analysis approach [21]. Data were organized into meaning units, coded, and presented as themes and subthemes.

Ethics Approval

Ethical approval was obtained from the Human and Social Science Research Ethics Committee of the University of KwaZulu-Natal (HSS/1509/018M).

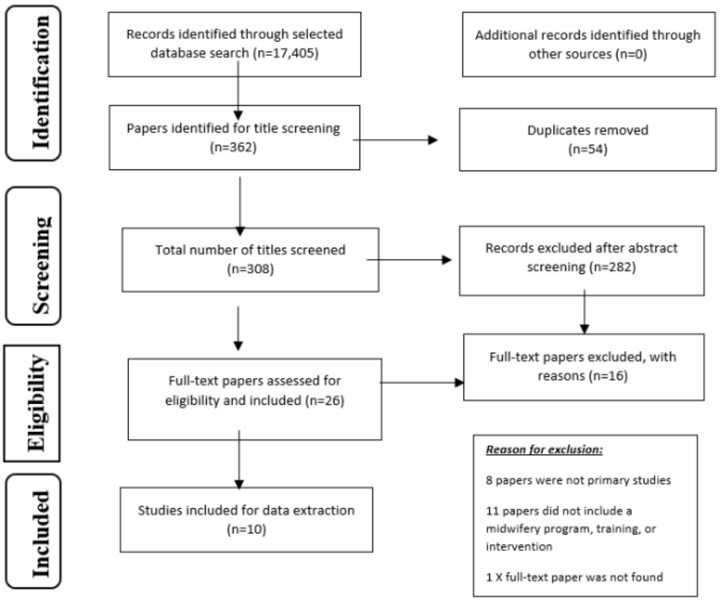
Results

The results are presented as the screening results and the data extraction results.

Screening Results

The researcher selected only papers from primary studies for this review and adopted the PRISMA flowchart [20]. The result of the screening process is shown in Figure 1. Screening results include the study characteristics (the research approaches and the study settings) and the quality of included papers.

Figure 1. The screening results presented in a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram [20].



Characteristics and Quality of the Included Papers

Research Approaches

There were 5 qualitative papers [22-26], 4 mixed methods papers [11,13,27,28], and 1 quantitative paper [29].

Study Settings

From across the globe, 4 studies were in the United Kingdom [13,22,23,27], 1 in Scotland [26], 2 in Australia [24,28], 1 in the United States [25], 1 in New Zealand [29], and lastly, 1 study in Uganda partnered with the United Kingdom [11].

Quality Assessment

Ten papers remained for data extraction, and the quality assessment of these papers was according to the research approaches selected in these primary studies. Hence, the mixed methods appraisal tool was selected to audit these papers. In total, 5 of the 10 papers (50%) were qualitative, of which 3 scored 100% and 2 scored 75%, showing high-quality values. The quality of 4 mixed method designs showed scores between 50% and 100%, and the remaining quantitative design paper scored 100%. These results indicated that all 10 papers were of high quality and complemented the purpose of conducting a systematic scoping review.

Data Extraction Results

Ten papers published between 2010 and 2020 remained for data extraction. The objective of this review was to identify interventions to support midwifery students during clinical practice. Data charting variables were applied to extract data during this stage. Table 1 shows the data extraction results.

Table 1. Data extraction.

Authors and year	Aims of the study	Intervention used	Significant findings
Broad et al [22], 2011	To support preregistration midwifery students during clinical placement	A transition model of preceptorship	• The intervention facilitated midwifery students learning in practice through the guidance of a preceptor. • Increased confidence and competence of newly qualified midwives. • Contributed to staff retention, increase co-operation, and quality of care given. • Increased investment in health care and education.
Barker et al [23], 2011	To train and support the role of mentors in assessing clinical competence of midwifery students	Mentor support by PEFs[a]	• Support for mentors is critical to improve student facilitation and support in clinical practice. • Protected time was necessary for SOMs[b] to attend workshops. • The intervention showed that better patient care outcomes increased collaboration between mentors, PEFs, and university and improved mentor assessment skills.
Durham et al [27], 2012	To develop skills in mentorship using a developmental program	A developmental program to support mentors	• The program promoted high standards of mentoring knowledge and skills and improved understanding and accountability of the mentorship roles. • A tripartite role benefitted the institution and the SOM.
Clements et al [24], 2012	To evaluate the core elements of a transition support program for newly qualified midwives from undergraduate and postgraduate nursing program	A transition support program for midwives	• A structured support during this transitional phase is necessary to ensure quality and safe practice of midwives. • Supernumerary time was highly valued but not always available. • Midwives appreciated study days, which allowed them to share their clinical experiences and debrief. • The program promoted peer midwife to midwife support.
Thunes and Sekse [25], 2015	To gain a better understanding of midwifery students' first encounter in the maternity wards and what was essential to them in the learning environment	A planned clinical practice approach	• Student-mentor relationships are crucial for students' achievements and learning outcomes. • Midwifery students need to feel valued and included in the team, learning was based on students' expectations, understanding, and previous experience. Mutual engagement with mentors is necessary.
Dixon et al [29], 2015	To explore the retention of new graduates in midwifery practice following participation in the MFYP[c] program	An MFYP program	• The program provided mentor support to new midwifery graduate and increased their confidence in the first year of practice as a registered midwife.
Moran and Banks [26], 2016	To explore the experiences and the value of "SOMs"	SOMs and the value they hold to this role	• Mentors valued their role and found it to be essential for the supervision of midwifery students during clinical practice. • Students value mentors for continuity, feedback, and planning.
Hogan et al [28], 2017	To explore the benefits of a peer mentoring program for midwifery students	A peer mentoring program in midwifery clinical placement	• Benefits to the mentee—reduced anxiety of first-year students, smoother transition to clinical practice, mentors were encouraging, understanding, reassuring, and positive. • Benefits to the mentor—building communication skills, self-confidence, and increased employability.
Kemp et al [11], 2018	To develop a model of mentorship for Ugandan midwifery students to improve the quality of midwifery care	The MOMENTUM[d] project 2015-2017	• Showed improved knowledge, skills, and attitudes of students and mentors. • Improved audit scores at clinical sites. • Improved confidence; however, mentors did not assess students' clinical skills in practice.

Authors and year	Aims of the study	Intervention used	Significant findings
Tweedie et al [13], 2019	To evaluate the model of coaching and collaborative learning and the role of the clinical education midwife	Collaborative coaching and learning model adapted from the CLiP[e] model by Lobo et al [30], 2014	• Improved students' confidence in knowledge and clinical and communication skills. • Student support through a clinical education midwife. • Ensured partnership between HEI[f] and hospitals.

[a]PEF: practice education facilitator.

[b]SOM: sign-off mentor.

[c]MFYP: Midwifery First Year of Practice.

[d]MOMENTUM: Developing a Model of Mentorship for Ugandan Midwifery.

[e]CLiP: Collaborative Learning in Practice.

[f]HEI: higher education institution.

Answering the Research Question

The objective of this review was to identify interventions available to strengthen the current clinical support for midwifery students globally. Interventions identified in this review included training programs, workshops, and one intervention guideline.

Synthesis of Screening Results

Overview

In total, 7 of the 10 studies (70%) implemented mentorship or preceptorship programs [11,13,22,24,26,28,29]. Two studies (20%) conducted a training workshop [23,27], while only 1 study (10%) included an intervention guideline [25]. These interventions supported either midwifery students or clinical mentors during clinical placements. The benefits of using clinical support interventions showed improvements in students' confidence levels, competence, and readiness for role-taking; it also revealed benefits for the clinical mentor in terms of improved mentorship knowledge, skills, and accountability [11,28]. Beyond these benefits, clinical support interventions show improved patient care outcomes [22,23] and collaborations between clinical facilities and nursing education institutions (NEIs) [13,23].

Meta-analysis of the significant findings was conducted to identify how interventions can be combined, adapted, and integrated to produce a more robust conclusion on strengthening midwifery students' clinical support during practice. Six codes emanated from the significant findings, as seen in Textbox 1. The third reviewer verified the findings and the constructed codes. These codes included academic-service partnerships, collaboration and consultation, clinical support methods, clinical support guidelines, clinical support materials, and course content.

These constructed codes were further analyzed to identify a more intense understanding of how to strengthen mentorship in midwifery. The review adopted a thematic content analysis approach [21]. Overall, 2 themes, with 2 subthemes each, emerged from the analysis. These themes are essential to guide mentorship program development and sustainability.

Textbox 1. Coding of significant findings.

Academic-service partnership

- Partnership between the clinical placement facility and the higher education institution is essential when designing an intervention for clinical support of midwifery students
- Partnership includes liaison between various stakeholders such as university educators, nurse managers, government personnel (if necessary), Mentors or preceptors or clinical facilitators, and expert advisory groups

Collaboration and consultation

- Continuous collaboration between the university and the hospital through a link lecturer is important
- Consultation with clinical mentors, midwifery students, students' support services, quality assurance teams, and previous cohort of students

Clinical support methods

- Presentations ranging from 3 hours to half-day workshops for clinical mentors included case scenarios, Objective Structured Clinical Examinations for evaluating mentor knowledge and skills
- Structured clinical support program for students, which includes student rotation plans, supernumerary time, and study days. Includes support for clinical mentors from universities, colleges, colleagues, and senior managers
- Structured Midwifery First Year of Practice program for newly qualified midwives
- 10-day study program validated by the Nursing and Midwifery Council (NMC) guidelines
- Peer-to-peer mentoring—3-hour training of third-year students (clinical mentor)
- Pilot sampling of intervention was adopted in 2 studies

Clinical support guidelines

- NMC Standards framework for nursing and midwifery education (2018) [31]
- Australian and New Zealand Support Services Association Incorporated guidelines

Materials used in clinical support training sessions

- Workbooks, portfolios, booklets, information pack, and a toolkit

Course content

- Role of the clinical mentor and mentee—named preceptors
- Outline of the program—practical component or areas of practice or placement schedules or clinical rotations, study days or skills education days, relationship building, communication skills, feedback, and debriefing opportunities
- Professional issues—NMC guidelines or standards for mentors
- Responsibilities or role expectations of clinical mentors—include boundary restrictions
- Self-care or support services available and referrals

Theme 1: Strengthening Partnerships and Consultation

Overview

The included papers revealed that improved partnerships and consultations were vital in supporting students during clinical placement. This theme developed from 2 subthemes: establishing stronger partnerships between NEIs and clinical facilities and improving consultation between midwifery educators, practitioners, and students.

Subtheme 1.1: Establishing Stronger Partnerships Between Nursing Education Institutions and Clinical Facilities

In 2011, the transition model of preceptorship began through regular meetings between the nurse managers and heads of departments at NEIs [22]. This strategy aimed to link the education and practice setting through a preceptorship model, which assisted midwifery students in achieving the required clinical practice standards. This highlighted that collaboration between the health facility, the facilitator, and the NEI is the cornerstone for success in mentorship [13], especially when negotiating protected time for mentors to attend workshops [23] or conduct mentor skills training [27]. Support from liaison facilitators employed at hospital facilities and educators of higher education facilities helped mentors to gain confidence in teaching and supervising students in practice. Hence, strengthening partnerships between NEIs and clinical facilities will facilitate continued collaborations and thus improve the clinical support of midwifery students. The idea was well-supported in other studies included in this review [11,26,28].

Subtheme 1.2: Improving Consultation Between Midwifery Educators, Practitioners, and Students

The review revealed that main stakeholders such as the nurse managers, regional or placement coordinators, clinical preceptors or mentors, midwifery practitioners, practice educators, or link lecturers have their roles in supporting midwifery students in clinical placements. Six papers showed that knowing the role of the mentor or preceptor, a named preceptor, contact details, clinical rotation, study days, and supernumerary time were factors that influenced the degree of clinical support offered to students by midwifery practitioners [22-24,26-28]. In addition, the continuity in students' support by the same preceptor with a planned or structured clinical plan influenced students' learning outcomes [13,25]. These authors further recommended that mutual engagement, shared knowledge, and shared goals are imperative to improving students' learning outcomes. Continuous relations between midwifery educators and practitioners should be encouraged because both share the responsibility of mentoring midwifery students during clinical practice.

Mentor relationships affect the students' perceptions of clinical practice. Students felt they depended on mentors to teach, show, and help them [25], and mentors, too, became optimistic. They showed interest in students' expectations and engaged with students through good teamwork and communication [26]. Furthermore, the mentor roles were valued because they played an essential role. Therefore, describing the mentors' role and expectations is critical in the training program, and this should be clear at the training program's onset [27].

Theme 2: Providing Mentor Support Through Training

Overview

Midwifery practitioners in clinical placements often feel unprepared to teach students due to the lack of training and support that is available to them. Without the necessary support and training, midwifery practitioners cannot fulfill a mentor's expected roles and responsibilities. Hence, mentor support and training are vital ingredients to improve the clinical support of midwifery students during placement. Two subthemes, namely, the quality of clinical support depends on the training content; and the training duration and structure.

Subtheme 2.1: The Quality of Clinical Support Depends on the Training Content

Durham et al's [27] study showed that a developmental training program to support mentors in their role focused on the content of the course and included a theory and practical component to support this training. The training content may include discussions on roles and responsibilities, professional issues, and boundaries to mentorship [22]. Therefore, mentorship training programs should include the policies and guidelines that govern midwifery education, practice, and training. In this review, 9 of the 10 studies (90%) were in first-world countries and guided by the Nursing and Midwifery Council (NMC). One peer mentoring study used the "Australian and New Zealand Support Services Association Incorporated guidelines" for peer mentoring. The training included the program's aims and objectives, the available resources, and a program evaluation [28]. According to Thunes and Sekse [25], mentorship training programs should have a planned clinical practice approach that emphasizes students' knowledge, skills, and learning needs to provide an overview of the mentors' expectations. Therefore, training courses for mentors should include information regarding student expectations of the midwifery curriculum, clinical practice requirements, and competencies to be achieved during clinical practice. Midwifery practitioners should be familiar with student requirements outlined in midwifery clinical workbooks and portfolios [27] or clinical booklets [28] to assist students to meet these requirements timeously. The content of training programs becomes critical to the success of mentorship. The information offered should ensure that training attendees become knowledgeable and skilled in their expected roles and responsibilities.

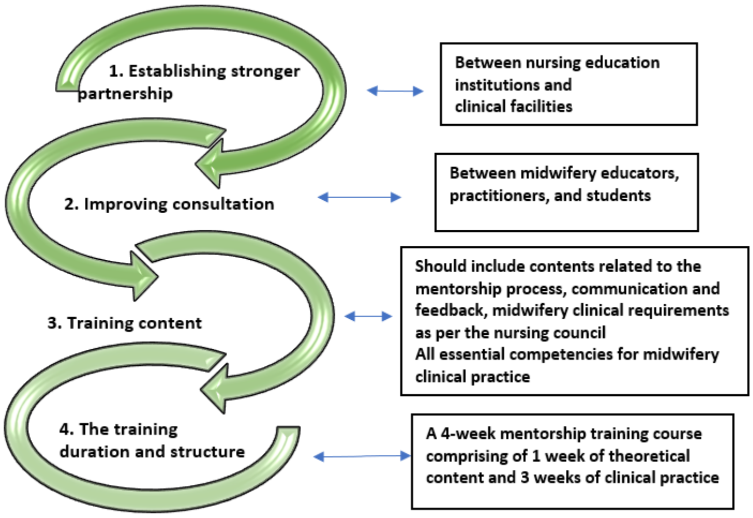
Subtheme 2.2: The Training Duration and Structure

This review identified clinical support interventions that range from a 3-hour face-to-face training session to a 10-day study program and extended to a 12-month program. Training sessions were either informal or unplanned or formal and planned and took place in the clinical placement site. Findings showed that mentors involved in informal, shorter, or fragmented training sessions could not attend all the sessions as they experienced challenges with leaving the wards and received poor support from senior colleagues and managers [23,24]. A well-planned and structured mentorship training program contributed to better clinical support outcomes [25,28]. Hence, the timing of mentorship training programs is vital to consider in line with ensuring that the program is well-planned, formalized, and nonfragmented.

Developing a Framework

Themes identified in this systematic scoping review are the significant results emanating from tried and tested interventions of previous studies. Therefore, the results that were recurrently seen in the included studies guided the researchers to identify core considerations when planning and developing mentorship training programs. These 4 subthemes are foundational for supporting any mentorship training program, and hence, the researchers propose these subthemes as a framework (see Figure 2) to guide mentorship training.

Figure 2. A framework to guide mentorship.



Discussion

Principal Results

This review identified interventions to support midwifery students during clinical practice. Included interventions were mentor support programs, mentorship models, models of preceptorship, mentor developmental programs, and collaborative learning in practice models. Findings showed that mentorship was the most practiced intervention in supporting students globally, producing benefits to both students and mentors. Additionally, the benefits of mentorship extended to improved patient care outcomes and collaborations between NEI and clinical facilities. However, mentorship training and support for midwifery practitioners who undertake the mentor role are not well established, and concerns over graduates' competence are worrisome. Therefore, it is necessary to follow clear guidance in developing successful mentorship training programs. The analysis of included papers highlighted essential aspects to consider when developing mentorship training programs. This involves strengthening partnership and consultation by establishing more robust relationships between NEI and clinical facilities and subsequently improving consultations between midwifery educators, practitioners, and students. Providing mentor support through training is essential, and therefore, the training content, structure, and duration of the mentorship training should accommodate clinical expectations.

Comparison of Prior Work

The quality of clinical support for midwifery students is a concern despite efforts toward improvements. This review showed that mentorship is the blueprint for supporting midwifery students to achieve the expected competence needed to become safe and independent practitioners. Mentorship benefits are seen globally, especially in many developed countries, and are effective in clinically preparing students for role-taking [12]. Similarly, this review presented that the benefits of mentorship extend from midwifery students to practitioners, academics, and patients or clients. Therefore, nurse managers and heads of NEIs should support midwifery practitioners and educators, respectively, in this shared mentorship responsibility. Hence, partnerships and collaborations between NEIs and clinical placements are necessary.

Continued consultation opportunities contribute to a better understanding of students' clinical expectations [22,23]. In 2011, trained sign-off mentors assisted midwifery students in achieving the requirements for clinical practice. However, these mentors experienced numerous challenges and felt inadequately prepared and supported in the role [23]. Subsequently, practice education facilitators were employed to support sign-off mentors in their roles [23].

In South Africa, midwifery educators and practitioners share the responsibilities of mentoring midwifery students during clinical placement. Improving consultation between midwifery educators (from NEIs) and midwifery practitioners (from clinical facilities) is needed to improve students' support. Student-centered learning approach in higher education institutions promotes student responsibility and accountability for own learning outcomes. As a result, midwifery students understand that establishing good mentorship relationships with midwifery educators and practitioners is crucial in achieving clinical learning outcomes. In an attempt to review the current midwifery preregistration programs, the NMC supports and empowers students to become active or self-directed learners [32] as does the South African Nursing Council (SANC) [2,32].

The findings from this review highlighted the importance of conducting a well-structured mentorship training program. These programs should align with the learning objectives stipulated by nursing councils and NEIs. Hence, maintaining strong partnerships and regular consultation between relevant stakeholders (NEIs and clinical facilities) is necessary to improve the clinical support of midwifery students. Furthermore, the training program's content should contain the students' learning objectives, the process of mentorship, essential midwifery competencies, assessment and support materials, contact details of midwifery educators, and guidelines to follow during the mentorship process. Through content-specific and contextualized mentorship training programs and support, midwifery practitioners should be able to carry out mentorship roles and responsibilities with ease.

Empowering midwifery practitioners through mentorship training and support is advantageous to the quality of service provided at a clinical facility. Yet, clinical challenges remain a barrier to attend training workshops conducted off-site. Besides, too lengthy training programs are also an inconvenience in fragmented working schedules. Therefore, on-site, short-term, on-the-job mentorship training approaches that integrate theory-related instruction are likely to complement a “hands-on” approach in clinical mentorship.

Strengths

Conducting systematic scoping reviews is a major strength in research as it ensures that only high-quality papers are included for data extraction. The review applied a mixed methodology, which provided a more detailed analysis of the findings. This review aims to identify the various interventions to strengthen midwifery clinical support and proposes a framework to guide mentorship training. The framework to guide mentorship training (Figure 2) is an investment to midwifery education and practice.

Limitations

The limitations of the study are as follows. First, this review was restricted to clinical support interventions available to midwifery students only and may have limited the clinical support interventions available across nursing disciplines. However, the selected population of this review was midwifery students only and hence did not affect the study results. Second, a restricted timeframe over the last decade (2010 to 2020) may have excluded older but more applicable models of interventions. In view of this limitation, the results may have been short-played. Third, the review excluded the implications of mentorship to other categories of nurses, and hence, this should be explored further in future studies.

Future Directions

The results of this review are likely to assist program developers and midwifery educators to participate in mentorship training and support programs. Strengthening mentorship through training opportunities for midwifery practitioners creates a platform to network and collaborates for the betterment of midwifery clinical practice and education. Given the limited papers retrieved from African countries in this review, there is a need for more research studies and publications on midwifery clinical education in African countries.

Conclusions

Across the globe, mentorship training programs were the most common clinical support available to midwifery students. Mentorship in maternity departments is crucial, and mentors require the support of their colleagues, senior managers, and midwifery educators to ensure mentorship success. The ultimate success of mentorship lies in improved patient care outcomes. Therefore, mentorship training and support for midwifery students should not be side-lined because the safety of our patients is in the hands of these students currently in training.

Mentorship training and support programs alone are insufficient to meet role players’ needs. It is important to strengthen partnerships between NEIs and clinical facilities as it allows midwifery educators to become involved in the training and support of midwifery practitioners ceasing consultation and collaboration opportunities. By expanding and promoting engagements between midwifery students, practitioners, and educators, mentorship in midwifery becomes an equally important 3-fold shared responsibility, and this is the goal mentorship program developers want to achieve.

So, mentorship program developers want to advance the scope of mentorship. Attempts to revive mentorship training opportunities are necessary. Despite global attempts to strengthen mentorship, the competence of midwifery students produced remains a significant concern. The framework to guide mentorship training proposed in this review is likely to encourage midwifery educators to pursue more mentorship training opportunities with ease and hence, improve the quality of midwifery clinical education.

A structured mentorship training program to support midwifery practitioners in their mentorship roles and responsibilities is necessary to make improvements in the quality of clinical support. Midwifery students who are well-supported during clinical placement assures that the next generation of midwives are safe and competent practitioners who are likely to contribute to positive maternal health outcomes globally and in South Africa.

Acknowledgments

The authors would like to thank Dr Claudine Muraraneza for her contribution during the screening process. The University of KwaZulu-Natal supported the publication of this paper.

Authors' Contributions

HA conceptualized and prepared the review under the guidance of SWM. Both authors contributed to reviewing the draft manuscript and approved the final version of this results paper.

Conflicts of Interest

None declared.

References

1. Gray M. Midwifery mentorship: what do we know about the mentors' perspective of the role? Australian Midwifery News. Informit. 2021. URL: https://search.informit.org/doi/10.3316/informit.562037703067807 [accessed 2021-06-29]
2. The Nursing Act 33 of 2005: regulations relating to the approval of and the minimum requirements for the education and training of a learner leading to registration in the categories professional nurse and midwife. South African Nursing Council. 2005. URL: https://www.sanc.co.za/wp-content/uploads/2020/06/R174-Reg-cpn.pdf [accessed 2020-08-16]
3. Bradshaw C, Noonan M, Barry M, Atkinson S. Working and learning: post-registration student midwives' experience of the competency assessment process. Midwifery 2013 May;29(5):519-525. [doi: 10.1016/j.midw.2012.04.010] [Medline: 23149236]
4. Amod H, Mkhize SW, Muraraneza C. Analyzing evidence on interventions to strengthen the clinical support for midwifery students in clinical placements: protocol for a systematic scoping review. JMIR Res Protoc 2021 Sep 21;10(9):e29707 [FREE Full text] [doi: 10.2196/29707] [Medline: 34546180]
5. Abuosi AA, Kwadan AN, Anaba EA, Daniels A, Dzansi G. Number of students in clinical placement and the quality of the clinical learning environment: a cross-sectional study of nursing and midwifery students. Nurse Educ Today 2022 Jan;108:105168. [doi: 10.1016/j.nedt.2021.105168] [Medline: 34656035]
6. Phuma-Ngaiyaye E, Bvumbwe T, Chipeta MC. Using preceptors to improve nursing students' clinical learning outcomes: a Malawian students' perspective. Int J Nurs Sci 2017 Apr 10;4(2):164-168 [FREE Full text] [doi: 10.1016/j.ijnss.2017.03.001] [Medline: 31406737]
7. Jamshidi N, Molazem Z, Sharif F, Torabizadeh C, Najafi Kalyani M. The challenges of nursing students in the clinical learning environment: a qualitative study. Sci World J 2016;2016:1846178 [FREE Full text] [doi: 10.1155/2016/1846178] [Medline: 27366787]
8. Sidebotham M, Fenwick J. Midwifery students' experiences of working within a midwifery caseload model. Midwifery 2019 Jul;74:21-28. [doi: 10.1016/j.midw.2019.03.008] [Medline: 30921548]
9. Stefaniak M, Dmoch-Gajzlerska E. Mentoring in the clinical training of midwifery students—a focus study of the experiences and opinions of midwifery students at the Medical University of Warsaw participating in a mentoring program. BMC Med Educ 2020 Oct 30;20(1):394 [FREE Full text] [doi: 10.1186/s12909-020-02324-w] [Medline: 33126872]
10. Lethale S, Makhado L, Koen MP. Factors influencing preceptorship in clinical learning for an undergraduate nursing programme in the North West Province of South Africa. Int J Africa Nurs Sci 2019;10:19-25 [FREE Full text] [doi: 10.1016/j.ijans.2018.11.006]
11. Kemp J, Shaw E, Musoke MG. Developing a model of midwifery mentorship for Uganda: the MOMENTUM project 2015-2017. Midwifery 2018 Apr;59:127-129. [doi: 10.1016/j.midw.2018.01.013] [Medline: 29425895]
12. Gray O, Brown D. Evaluating a nurse mentor preparation programme. Br J Nurs 2016;25(4):212-217. [doi: 10.12968/bjon.2016.25.4.212] [Medline: 26911167]
13. Tweedie K, Yerrell J, Crozier K. Collaborative coaching and learning in midwifery clinical placements. Br J Midwifery 2019 May 02;27(5):324-329 [FREE Full text] [doi: 10.12968/bjom.2019.27.5.324]
14. Clark L, Casey D. Support for mentors—an exploration of the issues. Br J Nurs 2016 Nov 10;25(20):1095-1100. [doi: 10.12968/bjon.2016.25.20.1095] [Medline: 27834526]
15. Wells L, McLoughlin M. Fitness to practice and feedback to students: a literature review. Nurse Educ Pract 2014 Mar;14(2):137-141. [doi: 10.1016/j.nepr.2013.08.006] [Medline: 24148671]
16. Mbakaya BC, Kalembo FW, Zgambo M, Konyani A, Lungu F, Tveit B, et al. Nursing and midwifery students' experiences and perception of their clinical learning environment in Malawi: a mixed-method study. BMC Nursing 2020;19:87 [FREE Full text] [doi: 10.1186/s12912-020-00480-4] [Medline: 32943984]
17. Jokelainen M, Turunen H, Tossavainen K, Jamookeeah D, Coco K. A systematic review of mentoring nursing students in clinical placements. J Clin Nurs 2011 Oct;20(19-20):2854-2867. [doi: 10.1111/j.1365-2702.2010.03571.x] [Medline: 21429055]
18. Simane-Netshisaulu KG. Student to midwife transition: newly qualified midwives' experiences in Limpopo province. Health SA Gesondheid 2022;27:1992 [FREE Full text] [doi: 10.4102/hsag.v27i0.1992] [Medline: 36483506]
19. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32 [FREE Full text] [doi: 10.1080/1364557032000119616]
20. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
21. Vaismoradi M, Jones J, Turunen H, Snelgrove S. Theme development in qualitative content analysis and thematic analysis. J Nurs Educ Pract 2016 Jan 15;6(5):100-110 [FREE Full text] [doi: 10.5430/jnep.v6n5p100]
22. Broad P, Walker J, Boden R, Barnes A. Developing a 'model of transition' prior to preceptorship. Br J Nurs 2011;20(20):1298-1301. [doi: 10.12968/bjon.2011.20.20.1298] [Medline: 22068004]
23. Barker M, Blacow L, Cosgrove S, Howorth N, Jackson G, McMahon J. Implementation of 'sign-off' mentorship: different perspectives. Br J Nurs 2011;20(19):1252-1255. [doi: 10.12968/bjon.2011.20.19.1252] [Medline: 22067838]
24. Clements V, Fenwick J, Davis D. Core elements of transition support programs: the experiences of newly qualified Australian midwives. Sex Reprod Healthc 2012 Dec;3(4):155-162. [doi: 10.1016/j.srhc.2012.08.001] [Medline: 23182448]

25. Thunes S, Sekse RJ. Midwifery students first encounter with the maternity ward. Nurse Educ Pract 2015 May;15(3):243-248. [doi: 10.1016/j.nepr.2015.01.012] [Medline: 25701290]
26. Moran M, Banks D. An exploration of the value of the role of the mentor and mentoring in midwifery. Nurse Educ Today 2016 May;40:52-56. [doi: 10.1016/j.nedt.2016.02.010] [Medline: 27125150]
27. Durham WJ, Kingston P, Sykes C. Implementing a sign off mentor preparation workshop—a tripartite approach. Nurse Educ Today 2012 Apr;32(3):273-277. [doi: 10.1016/j.nedt.2011.03.014] [Medline: 21497960]
28. Hogan R, Fox D, Barratt-See G. Peer to peer mentoring: outcomes of third-year midwifery students mentoring first-year students. Women Birth 2017 Jun;30(3):206-213. [doi: 10.1016/j.wombi.2017.03.004] [Medline: 28366500]
29. Dixon L, Calvert S, Tumilty E, Kensington M, Gray E, Lennox S, et al. Supporting New Zealand graduate midwives to stay in the profession: an evaluation of the Midwifery First Year of Practice programme. Midwifery 2015 Jun;31(6):633-639. [doi: 10.1016/j.midw.2015.02.010] [Medline: 25819705]
30. Lobo C, Arthur A, Lattimer V. Collaborative Learning in Practice (CLiP) for pre-registration nursing students. NHS. University of East Anglia. 2014. URL: https://healthacademy.lancsteachinghospitals.nhs.uk/app/uploads/2020/09/UEA-CLiP-Paper_-15-Sept-14.pdf [accessed 2023-04-12]
31. Standards for student supervision and assessment. Nursing and Midwifery Council. 2018. URL: https://www.nmc.org.uk/standards-for-education-and-training/standards-for-student-supervision-and-assessment/ [accessed 2020-01-31]
32. Standards to support learning and assessment in practice. Nursing and Midwifery Council. 2018. URL: https://www.nmc.org.uk/standards-for-education-and-training/standards-to-support-learning-and-assessment-in-practice/ [accessed 2020-06-03]

Abbreviations

MeSH: Medical Subject Headings
NEI: nursing education institution
NMC: Nursing and Midwifery Council
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SANC: South African Nursing Council

Edited by T Leung; submitted 12.01.22; peer-reviewed by A Joseph, S Hertling; comments to author 22.09.22; revised version received 03.11.22; accepted 14.12.22; published 21.04.23.

Please cite as:
Amod H, Mkhize SW
Supporting Midwifery Students During Clinical Practice: Results of a Systematic Scoping Review
Interact J Med Res 2023;12:e36380
URL: https://www.i-jmr.org/2023/1/e36380
doi:10.2196/36380
PMID:37083750

Original Paper

Intervention for Intraoperative Teaching in Anesthesiology Using Weekly Keyword Program: Development and Usability Study

George Tewfik[1], MD; Rotem Naftalovich[1], MD; Carlos Rodriguez-Aponte[1], MD; Bishoy Ezzat[1], MD

Rutgers New Jersey Medical School, Newark, NJ, United States

Corresponding Author:
George Tewfik, MD
Rutgers New Jersey Medical School
185 South Orange Ave
Newark, NJ, 07103
United States
Phone: 1 908 868 4292
Email: glt31@njms.rutgers.edu

Abstract

Background: Learning in the operating room (OR) for residents in anesthesiology is difficult but essential for successful resident education. Numerous approaches have been attempted in the past to varying degrees of success, with efficacy often judged afterward using surveys distributed to participants. The OR presents a particularly complex set of challenges for academic faculty due to the pressures required by concurrent patient care, production pressures, and a noisy environment. Often, educational reviews in ORs are personnel specific, and instruction may or may not take place in this setting, as it is left to the discretion of the parties without regular direction.

Objective: This study aims to determine if a structured intraoperative keyword training program could be used to implement a curriculum to improve teaching in the OR and to facilitate impactful discussion between residents and faculty. A structured curriculum was chosen to allow for the standardization of the educational material to be studied and reviewed by faculty and trainees. Given the reality that educational reviews in the OR tend to be personnel specific and are often focused on the clinical cases of the day, this initiative sought to increase both the time and efficiency of learning interactions between learners and teachers in the stressful environment of the OR.

Methods: The American Board of Anesthesiology keywords from the Open Anesthesia website were used to construct a weekly intraoperative didactic curriculum, which was distributed by email to all residents and faculty. A weekly worksheet from this curriculum included 5 keywords with associated questions for discussion. The residents and faculty were instructed to complete these questions on a weekly basis. After 2 years, an electronic survey was distributed to the residents to evaluate the efficacy of the keyword program.

Results: A total of 19 teaching descriptors were polled for participants prior to and following the use of the intraoperative keyword program to assess the efficacy of the structured curriculum. The survey results showed no improvement in intraoperative teaching based on respondent perception, despite a slight improvement in teaching time, though this was statistically insignificant. The respondents reported some favorable aspects of the program, including the use of a set curriculum, suggesting that greater structure may be beneficial to facilitate more effective intraoperative teaching in anesthesiology.

Conclusions: Although learning is difficult in the OR for residents, the use of a formalized didactic curriculum, centered on daily keywords, does not appear to be a useful solution for residents and faculty. Further efforts are required to improve intraoperative teaching, which is well known to be a difficult endeavor for both teachers and trainees. A structured curriculum may be used to augment other educational modalities to improve the overall intraoperative teaching for anesthesia residents.

(Interact J Med Res 2023;12:e42060) doi:10.2196/42060

KEYWORDS

resident teaching; intraoperative teaching; educational strategies; teaching; anesthesiology; education; efficacy; survey; electronic; medical residents; operation

Introduction

Education is a critical component of residency training; however, learning in the operating room (OR) for residents in anesthesiology is difficult and unstructured [1]. Various techniques have been used in the past to improve intraoperative teaching in anesthesiology. Faculty in residency programs have attempted to educate residents using traditional lectures on set topics, problem-based learning discussions, or case debriefing [1-3]. Attempts have been made to improve anesthesia education in the OR using a systematic approach to curriculum development and clearly defining study topics. Walsh et al [1] used a stepwise progression from a generalized needs assessment to a targeted needs assessment, defining goals and objectives and using various educational strategies and implementation.

Anesthesia education efficacy is difficult to assess due to the subjective nature of teaching and the rare measurable data points of formal examinations. Survey-based assessments to ascertain efficacy have been deployed in the past. Wakatsuki et al [4] used this methodology to conclude that in teaching, incorporating autonomy, reasoning, literature, prior knowledge, flexibility, reflection, as well as real-time feedback and teach back are most efficacious.

An important consideration when discussing intraoperative teaching is the maintenance of safety in patient care and vigilance for ongoing procedures. The practice of reading intraoperatively during periods of maintenance anesthesia [5] has been observed to have no significant effect on vigilance or responsiveness to adverse events. Other perceived barriers to successful intraoperative education include clinical production pressure on anesthesiologists [6].

Intraoperative teaching is also difficult for surgical services, in which faculty and residents spend the majority of their time engaging in patient care. A surgical study by Iwaszkiewicz et al [7] showed that faculty efforts to maintain a positive attitude toward teaching, establishing a calm and courteous environment, and providing “hands on” learning for residents contributed to improved perceptions by residents regarding intraoperative teaching [7]. Past studies have shown that acute stress is nearly ubiquitous in surgery and in the OR specifically, affecting both surgical performance and patient safety [8]. Interestingly, more recent studies have shown acute stress in the OR to cause both negative and positive effects on clinical performance [9]. Formalized training for faculty using evidence-based teaching frameworks has also been used with success [10]. Moreover, simulation has been used effectively by orthopedic surgical training programs to teach skills to trainees and residents [11].

Another interesting approach to intraoperative education for residents is the “briefing, intraoperative teaching, debriefing” model [12]. This model describes the use of a briefing to identify objectives for the case, intraoperative teaching focused on these objectives, and a debriefing after the case to reflect upon the events that have transpired [12]. Nonetheless, significant barriers have been identified in surgical literature regarding the gap in perception between residents and faculty regarding preparation for intraoperative learning and perioperative feedback, limiting the efficacy of perioperative education [13]. This discordance extends to large differences in the perception between trainees and faculty regarding both the quantity and quality of intraoperative teaching, though Timberlake et al [14] recommend a structured approach to perioperative teaching before, during, and after surgical cases.

Methods

Overview

The American Board of Anesthesiology keywords (archived online by the joint Open Anesthesia–International Anesthesia Research Society partnership) were used for a new intraoperative learning curriculum for the Department of Anesthesiology at Rutgers New Jersey Medical School [15].

Each week, 5 keywords were selected at random from the American Board of Anesthesiology keywords list, and a series of questions (3-5 per keyword) distributed based primarily upon the information on the Open Anesthesia website. The questions were open-ended to promote conversation between residents and faculty. The residents were instructed to choose a keyword each day to discuss with their assigned intraoperative faculty and to make that determination the night before so that both faculty and residents could study the topic ahead of time. Keywords were sent via email to residents and faculty on each Friday for the following week.

The keyword program began in March of 2019, and instructions were given in detail both at the beginning of the program and at regular intervals. The program continued for 2 years prior to evaluation by resident surveys. The survey to assess the efficacy of the keyword program was a modified version of the Anesthesia Theater Education Environment Measure (ATEEM) questionnaire [16]. The ATEEM questionnaire was modified into 19 questions assessing the efficacy of intraoperative teaching [16]. The Likert scale was used, scoring each category 1-5 from “Strongly Agree” to “Strongly Disagree.” Residents were asked to answer these questions comparing and contrasting days in the OR room when the keyword program was used for teaching and days when no keywords were discussed. Several additional questions were also added to the survey to assess the differences in time spent teaching, residents’ perceptions on the most successful overall modalities of intraoperative teaching, and the most effective characteristics of the keyword program for teaching.

Note that despite the distribution of the weekly keywords to all members of the department, keyword discussions did not occur between the residents and faculty daily. This was due to changes in staffing, changes in cases or OR assignments, emergent cases, or an inability of the faculty member to remain in the OR during maintenance anesthesia, when most intraoperative teaching occurs.

Ethics Approval

Institutional review board of experimental protocols was approved by Rutgers University (reference number Pro2019001411). All methods were carried out in accordance with the relevant guidelines and regulations. All participants signed informed consent for participation in this study. No compensation was provided for the study participants. Moreover, study data were deidentified prior to analysis.

Results

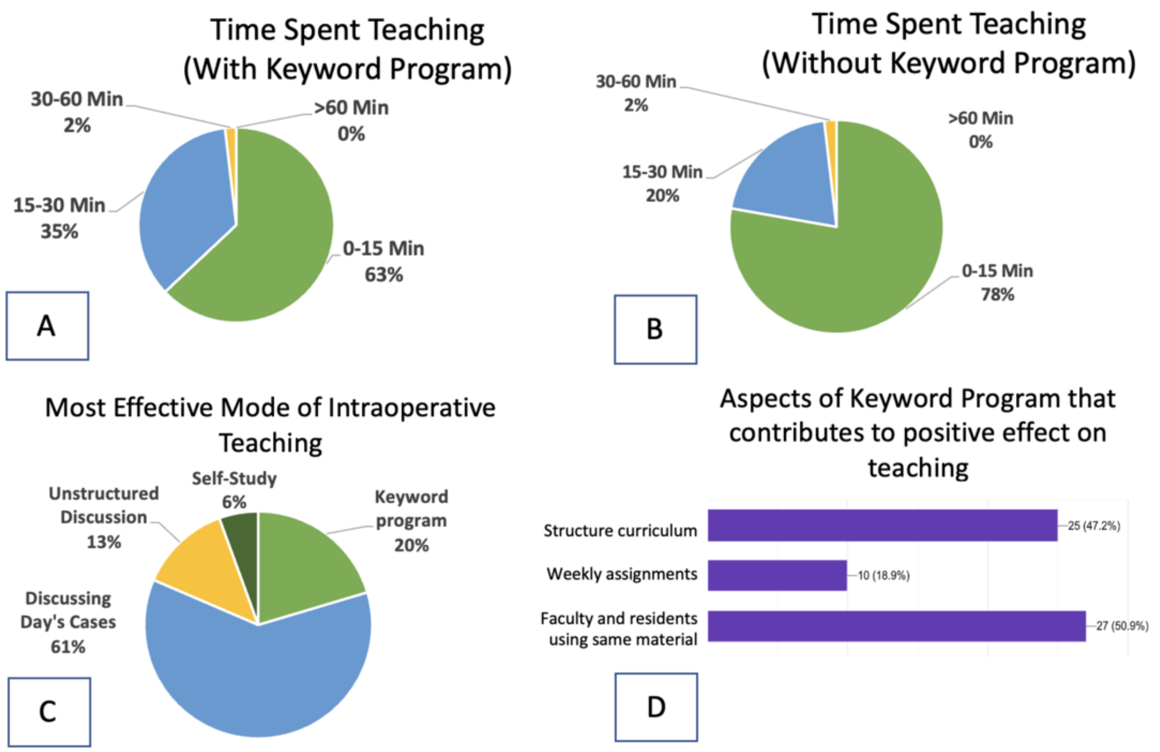
The program was initiated in July of 2019 and continued for 2 years until June 2021. Surveys were distributed to residents in June 2021, and 54 responses were recorded for the surveys, accounting for 90% of all residents during this time period. The results of the responses to the 19 questions (Textbox 1)—comparing days in which the keyword program was and was not used—were assessed for differences using the paired 2-tailed sample *t* test. No statistically significant changes were found between the 2 groups of responses, indicating no effect for the program. The residents were asked to rate each of these descriptors on the Likert scale for days when keywords were used and for days when keywords were not used for intraoperative teaching. Using an value of .05, no category demonstrated significant difference between the 2 groups.

Figure 1 includes additional survey results that demonstrate a statistically insignificant increase in time spent teaching, the most effective mode of teaching, and potential aspects of the keyword program found to contribute positively to intraoperative teaching. The respondents who stated that 0-15 minutes a day were spent on resident education decreased from 78% (42/54) to 63% (34/54) when keywords were integrated into the day's instruction, and they increased from 20% (11/54) to 35% (19/54) for those indicating that 15-30 minutes a day were spent on learning (Figure 1A,B). Only 20% (11/54) of the respondents indicated that the structured keyword program was the most effective tool for intraoperative learning, with 61% (33/54) reporting that the discussion of the current clinical case was more efficacious and conducive to learning (Figure 1C). The aspects of the keyword program that were found to be most helpful for intraoperative learning include using a structured curriculum (25/53, 47.2%) and using the same curriculum for faculty and residents to study (27/53, 50.9%; Figure 1D).

Textbox 1. Teaching descriptors assessed using the Anesthesia Theater Education Environment Measure tool in the survey form following 2 years of the keyword intraoperative teaching program.

The teaching helps to develop my confidence.
I receive effective supervision from the clinical teachers.
Teaching is done at appropriate times not affecting vigilance.
I receive teaching anesthetic specialty areas targeted at my learning needs.
The teacher helps to develop my competence.
My clinical teachers are accessible for advice.
I experience friendly relations with my teachers in the operating room.
The clinical teachers in this hospital interact well with trainees.
My clinical teachers promote an atmosphere of mutual respect.
I have an appropriate level of clinical responsibility.
My clinical teachers are clear in their teaching.
I am clear about the learning objectives of teaching sessions in the operating room.
I receive the necessary clinical supervision.
I have a good collaboration with anesthesia staff.
I have the opportunity for on-the-job learning.
My clinical teachers have established good rapport with me.
I am encouraged to participate in the theatre setting.
There is a systematic clinical training program.
I feel able to ask the question I want.

Figure 1. A-B: the differences in time spent teaching on days with or without keywords (though no statistical significance found). 1C: survey results when residents were asked about the most efficacious mode of intra-operative teaching. 1D: respondents' choices regarding the aspects of the keyword program that they deemed to contribute positively to intra-operative teaching.



Discussion

Principal Findings

This investigation sought to improve intraoperative teaching by incorporating a structured curriculum with keywords for residents in anesthesiology and the faculty. Unfortunately, there was no demonstrable positive effect to this intervention. There was essentially no statistically significant difference in the responses by the residents to a survey when comparing intraoperative teaching with and without the use of keywords. This may demonstrate a failure on the part of the initiative to make meaningful improvements to intraoperative teaching. The survey used for the study was an ATEEM tool, which is a validated structure to assess education in anesthesiology intraoperatively, and it was modified to assess the efficacy of the keyword program in this study.

In Figure 1, resident survey respondents noted an overall decrease in time spent during intraoperative teaching when keywords were not used by residents and faculty. The respondents who believed teaching constituted 15-30 minutes of the day fell from 35% (19/54) to 20% (11/54) when keywords were not used, and the category of 0-15 minutes increased from 63% (34/54) to 78% (42/54) in this cohort. Figure 1C shows that resident survey respondents believe that despite the years-long implementation of this keyword program, the most efficacious form of intraoperative teaching is discussing the cases of the day. This may suggest that intraoperative teaching is more effective when didactic material matches the clinical case that is commanding the resident's attention during the workday.

Nonetheless, the residents identified several characteristics of the keyword program that they believe contributed to a positive effect of keywords on intraoperative teaching in Figure 1D, including the fact that keywords forced residents and faculty to use the same educational material simultaneously, and that the keywords used a structured curriculum. These aspects of the keyword program may offer possibilities for future educational interventions to improve the intraoperative instruction of resident trainees.

Intraoperative learning is a notoriously difficult task for educators of residents in medicine. Past studies have attempted to use such modalities as traditional lectures, problem-based learning discussions, and case debriefing, as well as the targeted assessment of residents [1-3]. Our study attempted to use a set curriculum to teach residents in the OR theater, and to assess the efficacy of this program with a validated survey-based approach [4]. It is unclear exactly why this approach failed, but it is possible that focusing on an unrelated topic during a surgical procedure may not have been feasible due to the aforementioned clinical production pressure, which has been demonstrated to be a barrier to successful intraoperative education [6]. It is also more than likely that any efforts to improve education in the OR have a large barrier to success due to the acute stress caused by the environment [8].

Limitations

Limitations of this study include a lack of assessment of the percentage of time when the intraoperative keyword curriculum was used by learners. The keywords were sent to both residents and faculty weekly, with no mechanism in place to ensure a successful adherence to the program. This was deemed too difficult due to clinical production pressure, patient emergencies, call burden, vacations, and off-site rotations. Nonetheless, this is a significant limitation, because survey respondents may be included who did not participate in the program or use the keywords in a meaningful manner during the study period.

Conclusions

After using this intraoperative keyword teaching program for more than 2 years, this study revealed that it had a minimal effect on intraoperative teaching between the anesthesiologist resident trainees and faculty. Consideration should be given to alternate methodologies to improve intraoperative teaching for learners in the anesthesiology residency. The results presented in this study may suggest characteristics of a future intervention that may be more successful in improving intraoperative education. Nearly half of the respondents agreed that the use of a structured curriculum contributed positively to education. Perhaps the use of a structured curriculum that is custom tailored to a resident's current rotation (instead of an arbitrary schedule, which was employed in this study) may be more efficacious for residents' educational enhancement. It was also noted by the survey respondents that the coordination of educational material between the residents and faculty was advantageous for learning. This could be incorporated into a policy in which the resident and faculty member plan their educational discussions ahead of time to provide both the teacher and the learner the opportunity to review a chosen topic before their review together in the OR.

Data Availability

All data generated or analyzed during this study are included in this article.

Conflicts of Interest

None declared.

References

1. Walsh DP, Neves SE, Wong VT, Mitchell JD. Formation of an Intraoperative Educational Curriculum for Anesthesiology Residents Using a Systematic Approach. A A Pract 2020 Oct;14(12):e01330 [FREE Full text] [doi: 10.1213/XAA.0000000000001330] [Medline: 33094949]
2. Rudolph J, Simon R, Dufresne RL, Raemer DB. There's no such thing as "nonjudgmental" debriefing: a theory and method for debriefing with good judgment. Simul Healthc 2006;1(1):49-55. [doi: 10.1097/01266021-200600110-00006] [Medline: 19088574]
3. Viola L, Young DA. How to Teach Anesthesia in the Operating Room. Int Anesthesiol Clin 2016;54(3):18-34. [doi: 10.1097/AIA.0000000000000103] [Medline: 27285070]
4. Wakatsuki S, Tanaka P, Vinagre R, Marty A, Thomsen JLD, Macario A. What Makes for Good Anesthesia Teaching by Faculty in the Operating Room? The Perspective of Anesthesiology Residents. Cureus 2018 May 01;10(5):e2563-e2563 [FREE Full text] [doi: 10.7759/cureus.2563] [Medline: 29974018]
5. Slagle, Weinger MB. Effects of intraoperative reading on vigilance and workload during anesthesia care in an academic medical center. Anesthesiology 2009 Feb;110(2):275-283 [FREE Full text] [doi: 10.1097/ALN.0b013e318194b1fc] [Medline: 19194155]
6. Haydar B, Schwartz AJ, Baker K, Ambardekar AP. Academic Anesthesiologists Perceive Significant Internal Barriers to Intraoperative Teaching in a Cross-Sectional Survey. JEPM 2019 Jan 1;21(1):1-14. [doi: 10.46374/volxxi-issue1-haydar]
7. Iwaszkiewicz M, Darosa DA, Risucci DA. Efforts to enhance operating room teaching. J Surg Educ 2008 Nov;65(6):436-440. [doi: 10.1016/j.jsurg.2008.07.006] [Medline: 19059174]
8. Georgiou K, Larentzakis A, Papavassiliou AG. Surgeons' and surgical trainees' acute stress in real operations or simulation: A systematic review. Surgeon 2017 Dec;15(6):355-365. [doi: 10.1016/j.surge.2017.06.003] [Medline: 28716368]
9. Tjønnås MS, Guzmán-García C, Sánchez-González P, Gómez EJ, Oropesa I, Våpenstad C. Stress in surgical educational environments: a systematic review. BMC Med Educ 2022 Nov 15;22(1):791 [FREE Full text] [doi: 10.1186/s12909-022-03841-6] [Medline: 36380334]
10. Gardner A, Timberlake MD, Dunkin BJ. Faculty Development for the Operating Room: An Examination of the Effectiveness of an Intraoperative Teaching Course for Surgeons. Ann Surg 2019 Jan;269(1):184-190. [doi: 10.1097/SLA.0000000000002468] [Medline: 28817439]
11. Atesok K, Hurwitz S, Anderson DD, Satava R, Thomas GW, Tufescu T, et al. Advancing Simulation-Based Orthopaedic Surgical Skills Training: An Analysis of the Challenges to Implementation. Adv Orthop 2019 Sep 02;2019:2586034-2586037 [FREE Full text] [doi: 10.1155/2019/2586034] [Medline: 31565441]

12. Roberts N, Williams RG, Kim MJ, Dunnington GL. The briefing, intraoperative teaching, debriefing model for teaching in the operating room. J Am Coll Surg 2009 Feb;208(2):299-303. [doi: 10.1016/j.jamcollsurg.2008.10.024] [Medline: 19228544]
13. Rose JS, Waibel BH, Schenarts PJ. Disparity between resident and faculty surgeons' perceptions of preoperative preparation, intraoperative teaching, and postoperative feedback. J Surg Educ 2011 Nov;68(6):459-464. [doi: 10.1016/j.jsurg.2011.04.003] [Medline: 22000531]
14. Timberlake M, Mayo HG, Scott L, Weis J, Gardner AK. What Do We Know About Intraoperative Teaching?: A Systematic Review. Ann Surg 2017 Aug;266(2):251-259. [doi: 10.1097/SLA.0000000000002131] [Medline: 28059834]
15. Keywords. Open Anesthesia. 2019. URL: http://www.openanesthesia.org/aba-keywords/ [accessed 2023-04-21]
16. Holt MC, Roff S. Development and validation of the Anaesthetic Theatre Educational Environment Measure (ATEEM). Med Teach 2004 Sep 12;26(6):553-558. [doi: 10.1080/01421590410001711599] [Medline: 15763835]

Abbreviations

ATEEM: Anesthesia Theater Education Environment Measure
OR: operating room

Edited by A Mavragani; submitted 19.08.22; peer-reviewed by D Grech, J Bruthans, J Gamble; comments to author 12.12.22; revised version received 22.03.23; accepted 13.04.23; published 18.05.23.

Please cite as:
Tewfik G, Naftalovich R, Rodriguez-Aponte C, Ezzat B
Intervention for Intraoperative Teaching in Anesthesiology Using Weekly Keyword Program: Development and Usability Study
Interact J Med Res 2023;12:e42060
URL: https://www.i-jmr.org/2023/1/e42060
doi:10.2196/42060
PMID:37200082

Original Paper

Professional Relationship Between Physicians and Journalists in Bangladesh: Web-Based Cross-Sectional Study

Md Aminul Islam[1,2], MSS; Md Golam Rabbani[2,3], MSS; Zamilur Rahaman[4], MD; Taufique Joarder[5], DrPH

[1]Department of Media Studies and Journalism, University of Liberal Arts Bangladesh, Dhaka, Bangladesh
[2]Public Health Foundation, Bangladesh, Dhaka, Bangladesh
[3]Health Economics Unit, Health Services Division, Ministry of Health and Family Welfare, Dhaka, Bangladesh
[4]Monno Medical College, Monno City, Manikganj, Dhaka, Bangladesh
[5]SingHealth Duke-NUS Global Health Institute, Singapore, Singapore

Corresponding Author:
Md Aminul Islam, MSS
Department of Media Studies and Journalism
University of Liberal Arts Bangladesh
688 Beribadh Road, Mohammadpur
Dhaka, 1209
Bangladesh
Phone: 880 1718674830
Email: aminul.vu@gmail.com

Abstract

Background: A health care system is intertwined with multiple stakeholders, including government institutions, pharmaceutical companies, patients, hospitals and clinics, health care professionals, health researchers and scientific medical experts, patients and consumer organizations, and media organizations. Physicians and journalists are the key actors who play a significant role in making health care services and health information accessible to the people of a country.

Objective: The aim of this study was to explore the tensions and alliances between physicians and journalists in Bangladesh, along with identifying strategies that could potentially improve the often contentious relationship and quality of medical journalism.

Methods: We conducted a web-based cross-sectional survey using the snowball sampling technique from September 2021 to March 2022. Adult Bangladeshi citizens belonging to the two selected professional groups (physicians and journalists), who understood the survey content, and agreed to participate in the survey were considered eligible for inclusion in the study. Both descriptive and logistic regression analyses, including the Mann-Whitney *U* test and Wilcoxon signed-rank test, were performed to determine the differences between groups concerning selected perception-related variables, and the associations of perceptions about lack of trust in each other's knowledge, skills, and professional integrity with background characteristics.

Results: A total of 419 participants completed the survey, including 219 physicians and 200 journalists. Among physicians, 53.4% (117/219) reported lower trust toward journalists' professional domain and expertise, whereas 43.5% (87/200) of journalists had lower trust toward physicians' professional domain and expertise. In terms of perception about not having respect for each other, the median value for the physicians was 5 (strongly agree), whereas it was only 3 (agree) for the journalists. We also found that male physicians (adjusted odds ratio [AOR] 0.45, compared with female physicians) and medical officers (AOR 0.30, compared with specialists) had significantly higher odds of lacking trust in journalists' knowledge, skills, and professional integrity. When rating the statement "Regular professional interaction between journalists and doctors may improve the relationship between the professional groups," most physicians (186/219, 84.9%) chose "neither agree nor disagree," whereas most journalists (106/200, 53.0%) stated that they "slightly agree."

Conclusions: Both physicians and journalists in Bangladesh have negative perceptions of each other's professions. However, physicians have a more negative perception of journalists than journalists have of the physicians. Strategies such as a legal framework to identify medical-legal issues in reporting, constructive discussion, professional interaction, and capacity-building training programs may significantly improve the relationship between physicians and journalists.

(Interact J Med Res 2023;12:e44116) doi:10.2196/44116

KEYWORDS
medicine; media; physician; journalist; communication; health policy; Bangladesh

Introduction

Background

A health care system is intertwined with multiple stakeholders, including government institutions, pharmaceutical companies, patients, hospitals and clinics, health care professionals, health researchers and scientific medical experts, patients and consumer organizations, and media organizations. Physicians and journalists are the key actors who play a significant role in making health care services and health information accessible to the people of a country. Media translate complex health issues, health policies, scientific medical innovations, and research updates for the public, patients, practitioners, and policy makers. People's health-related behaviors [1]; beliefs, attitudes, and actions [2]; and perceptions of the quality of health care services are influenced and shaped by media content. Therefore, media play a crucial role in framing public health debates [3] and the public health policy process [4,5].

Nevertheless, health and media are closely connected in many other ways. Briggs and Hallin [6] examined the relationship between media and medicine. They argued that news coverage of health issues plays a fundamental role in constructing wider cultural understandings of health and disease. Moreover, Stroobant et al [7] argued that health news is coproduced by health and media professionals. However, authors of previous studies explored the relationship between health and media through the lens of certain professional domains. In most cases, they adopted either the media perspective [6] or a health care perspective [8] exclusively. Thus, it is evident that the current discourse on the relationship is divided into two lines of thought. On the one hand, media professionals often argue that doctors do not know how to express themselves in a way that nonmedical professionals can understand, they do not appreciate journalists' skills and/or act as if they are superior or omnipotent, they try to take over/dominate the journalistic process, and they have a personal agenda when collaborating with the media. On the other hand, health professionals claim that the media cannot be trusted as journalists often report health-related issues in a biased, sensational, and inaccurate way; they do not understand the complexity of health care and the health care business; and are often responsible for breaches of confidentiality or privacy, or choose to misquote health professionals [9].

It is well established that media can play an influential role in promoting health behavior [10-14], the use of health care services [12,15-18], building people's trust in the health care system [19-23], and advancing health literacy [24,25] in a country. A cordial relationship between physicians and journalists is therefore crucial for improved health care delivery and the public health of a country. However, physicians and medical scientists often argue that the media frequently negatively portray the health sector and health care professionals and use exaggerated headlines that lower the quality of medical messages in the media. On the other side, media professionals argue that health care professionals do not cooperate with them in communicating medical messages properly. However, work on incorporating both perspectives is scant. In particular, there is no research on the topic in the context of Bangladesh.

Against this backdrop, the aim of this study was to explore the professional relationship between physicians and journalists in Bangladesh. The goal is to identify strategies that could potentially improve this often contentious relationship by characterizing the tensions and alliances between medicine and the media in Bangladesh and the quality of medical journalism in the country.

Theoretical Framework

The aim of this study was to examine the professional relationship between physicians and journalists through the theoretical lenses of biocommunicability [26], biomediatization [6], and boundary-work theory [27]. The notion of biocommunicability refers to the ways biomedical knowledge is created, circulated, and received [6,26]. According to this concept, media and medicine are two distinct but intensively interactional entities. By contrast, the biomediatization concept implies that biomedicine and the media are not two separate entities but are rather deeply intertwined. Both medicine and media contribute to the production of medical knowledge, the practice of medicine, and public health [6,7]. How media construct and communicate health knowledge affects perceptions of particular diseases, public health policies, clinic practices, and public reactions because the media frame health news through multiple social, economic, cultural, political, and moral lenses [6,28].

The concept of boundary work was first introduced by Thomas F Gieryn in 1983, which refers to an ideological demarcation between scientific and nonscientific fields [27]. Gieryn argued that various professional groups and occupations construct social boundaries that distinguish some intellectual activities. They put up such boundaries in arguing for their power, authority, control, credibility, expertise, prestige, and material resources. Moreover, they play rhetorical games for their objectivity and the need for autonomy. As an analytical instrument, this concept is particularly useful in understanding the professional relationship between physicians and journalists [29-32]. Thus, to understand the tensions and alliances between medicine and the media in Bangladesh, we sought to identify the prejudices physicians and journalists have against one another toward finding possible solutions that could improve the mutual relationship and the quality of medical journalism in the country.

Medical journalism refers to journalistic communication of issues related to health, medicine, and the health care system. In essence, medical journalism is another form of journalistic writing about science [33,34], representing an art and craft of telling complex stories on structural, institutional, political, financial, and ethical issues in health, medicine, and health care [35] in a way that enables a lay person to easily understand these issues.

Methods

Study Design, Setting, and Participants

This was a cross-sectional online survey. To achieve the objectives of this study, a quantitative approach was adopted. As the aim was to capture the perceptions of both physicians and journalists working in Bangladesh, two separate surveys were created, each comprising common variables and participant-specific variables. The surveys were conducted between September 2021 and March 2022. The call for participation was made on social media and by email.

Recruitment Procedure

We collected data through an anonymous web-based survey using social media platforms and email. Two semistructured questionnaires (for physicians and journalists, respectively) were designed using the Google survey tool (Google Forms). The generated link was shared with physicians and journalists identified through the snowball technique. The link was also shared with study participants via social media groups. Through the link, the study participants could access the relevant questionnaire as well as read a brief description of the study, with its objectives, implications, and data management guidelines. Informed consent was obtained from all participants through the same web link. After providing consent, a participant was able to access the remainder of the questionnaire, which also included the contact addresses of the research team and an Ethical Review Committee member, allowing them to reach out for further queries or clarification regarding the study. The participants were not required to provide any personal or identifiable information on the questionnaires. To maintain data quality, the research team checked the data regularly to determine whether there were any inconsistencies.

We collected data from professional physicians/registered physicians/clinical practitioners such as senior consultants, junior consultants, teaching professionals of medical colleges, and residential medical officers/medical officers or equivalent who work in primary, secondary, and tertiary government hospitals and medical colleges, as well as private clinics and private medical colleges across the country. In addition, any registered journalist working in print, television, or online news platforms was considered eligible for the journalist survey. All Bangladeshi citizens aged 20 years or above that belonged to the two selected professional groups, understood the survey content, and agreed to participate in the survey were considered eligible for inclusion in the study. A total of 419 of 528 participants completed the survey, with a response rate of 79.35%, including 219 physicians and 200 journalists.

Study Instruments

We developed the physician and journalist perception questionnaires following the existing literature, after which we customized these items to the Bangladeshi context and translated them into Bangla. The questionnaires included sociodemographic and profession-related questions. The physician perception questionnaire included questions on their experience of professional interactions with journalists, perceptions of the impact of media on the health care sector, perceptions of the importance of a good relationship between medicine and media, prejudice about media and journalists, and suggestions for improving the relationship. The journalist perception questionnaire comprised questions on their perceptions about health care professionals, knowledge about health and medical reporting, the experience of interactions with physicians, perceptions about the importance of a good relationship between medicine and media, prejudice about physicians and health care professionals, and suggestions for improving the relationship. The questionnaires are provided in Multimedia Appendix 1.

Statistical Analysis

We performed both descriptive and inferential statistical analyses. The descriptive analysis focused on frequencies (n) and percentages, and the Mann-Whitney *U* test and Wilcoxon signed-rank test were performed to determine the differences between groups (physicians and journalists) concerning selected perception-related variables. Internal consistency of the perception variables between the two groups was tested using the Cronbach α coefficient. The Cronbach α for the common perception (7 items) variables between the two groups was .776, indicating a satisfactory internal consistency level [36,37]. Moreover, we performed a reliability test for the physician perception–related variables toward journalists (12 items) and the journalist perception–related variables (9 items) toward physicians. The Cronbach α score for the physician and journalist groups was .893 and .814, respectively, indicating a satisfactory internal consistency level [36,37]. Multiple ordered logistic regression analyses jointly considering all the explanatory variables were performed to assess the association between the background characteristics of the two study groups with a common perception variable, formed based on the rating of the survey item “Not having trust in each other’s knowledge, skills, and professional integrity” on a 5-point Likert scale anchored at 1=“strongly disagree” and 5=“strongly agree.” A *P* value <.05 was considered statistically significant. We analyzed the data using Stata SE, version 15.0 (StataCorp LLC).

Ethics Considerations

The Ethical Review Committee (PHFBD-ERC: 12/2020) of the Public Health Foundation, Bangladesh approved our study protocol, procedures, consent statement, and study tools. All respondents were informed in Bengali about their rights related to their voluntary participation in the study. Participants who gave consent to willingly participate in the survey would click the “Continue” button and would then be directed to complete the self-administered questionnaire. Respondents were assured of the anonymity of the data they provided.

Results

Background Characteristics of Physicians and Journalists

A total of 219 physicians and 200 journalists residing in Bangladesh completed the questionnaire. Among the physicians, the mean age was 38.86 (SD 9.94) years and the mean duration of professional experience was 12.68 (SD 8.56) years. Nearly three-quarters of the participants identified as male; 31.05%

worked as specialists and 12.79% as consultants. In addition, 43.38% of the participants were currently based in Dhaka, the capital city of Bangladesh (Table 1). Among the journalists, the mean age was 33.97 (SD 9.24) years and the mean duration of professional experience was 9.85 (SD 8.19) years; nearly three-quarters were male. In addition, 48.5% worked as a reporter and 61.5% were currently working in Dhaka (Table 1).

Table 1. Background characteristics of the study participants (N=419).

Variables	Physicians (n=219)	Journalists (n=200)
Age (years), mean (SD)	38.86 (9.94)	33.97 (9.24)
Years of professional experience, mean (SD)	12.68 (8.56)	9.85 (8.19)
Sex, n (%)		
Male	159 (72.6)	149 (74.5)
Female	59 (26.9)	48 (24.0)
Prefer not to say	1 (0.5)	3 (1.5)
Physicians' professional title, n (%)		
Medical officer	111 (50.7)	N/A[a]
Junior consultant	12 (5.5)	N/A
Consultant	28 (12.8)	N/A
Specialist	68 (31.1)	N/A
Journalists' professional title, n (%)		
Correspondent	N/A	53 (26.5)
Reporter	N/A	97 (48.5)
News editor	N/A	42 (21.0)
Others (eg, anchor, media manager)	N/A	8 (4.0)
Current working place, n (%)		
Capital city (Dhaka)	95 (43.4)	123 (61.5)
Other divisional city	43 (19.6)	24 (12.0)
District	56 (25.6)	35 (17.5)
Upazila[b] and Union[c]	25 (11.4)	18 (9.0)

[a]N/A: not applicable.

[b]An administrative division in Bangladesh, functioning as a subunit of a district.

[c]The smallest rural administrative and local government unit in Bangladesh.

Physicians' Perceptions Toward Journalists and Journalists' Perceptions Toward Physicians

The perceptions of the two professional groups toward each other were assessed based on seven common domains (Table 2). Both physicians and journalists ranked their perceptions toward each other using a 5-point Likert scale, with 1 denoting "strongly disagree" or having a very negative perception and 5 indicating "strongly agree" or having a very positive perception. The complete data are provided in Multimedia Appendix 2.

Six out of seven variables regarding physicians' perception toward journalists and journalists' perception toward physicians were statistically significant at the 5% level (P<.05). Among physicians, 53.4% reported that they had lower trust toward journalists' professional domain and expertise, whereas among journalists, 43.5% of participants had lower trust toward physicians' professional domain and expertise. Moreover, 56.2% of physicians were of the view that journalists have a "very low" level of professionalism, whereas only 32.5% of the journalists perceived physicians to have a "very low" level of professionalism.

When rating the statement "Journalists often do not have respect for physicians as a professional group," 50.7% of physicians "strongly agreed." In contrast, only 10.5% of journalists "strongly agreed" with the same statement regarding the physicians. Similarly, 51.6% of physicians "strongly agreed" with the statement "Journalists do not have trust in the knowledge, skills, and professional integrity of physicians," whereas only 4.5% of journalists "strongly agreed" when rating the statement "Physicians do not have trust in media and journalists in the country."

When rating the statement "Journalists tend to believe that they are superior to physicians as a professional group," 86.8% of physicians "strongly agreed." In contrast, only 19.0% of journalists "strongly agreed" with the corresponding statement regarding the physicians. In addition, 82.2% of physicians "strongly agreed" with the statement "When reporting on the health sector, journalists often tend to serve the purpose of vested business interests, not representing the truth." In contrast, only 1.0% of journalists "strongly agreed" with the statement "Physicians often prescribe medicine or tests to ensure the interests of pharmaceutical companies." Finally, 89.5% of physicians, as opposed to 18.5% of journalists, expressed their strong agreement with the statement "The relationship between physicians and journalists is not good."

Table 2. Physicians' perceptions toward journalists and journalists' perceptions toward physicians in Bangladesh.

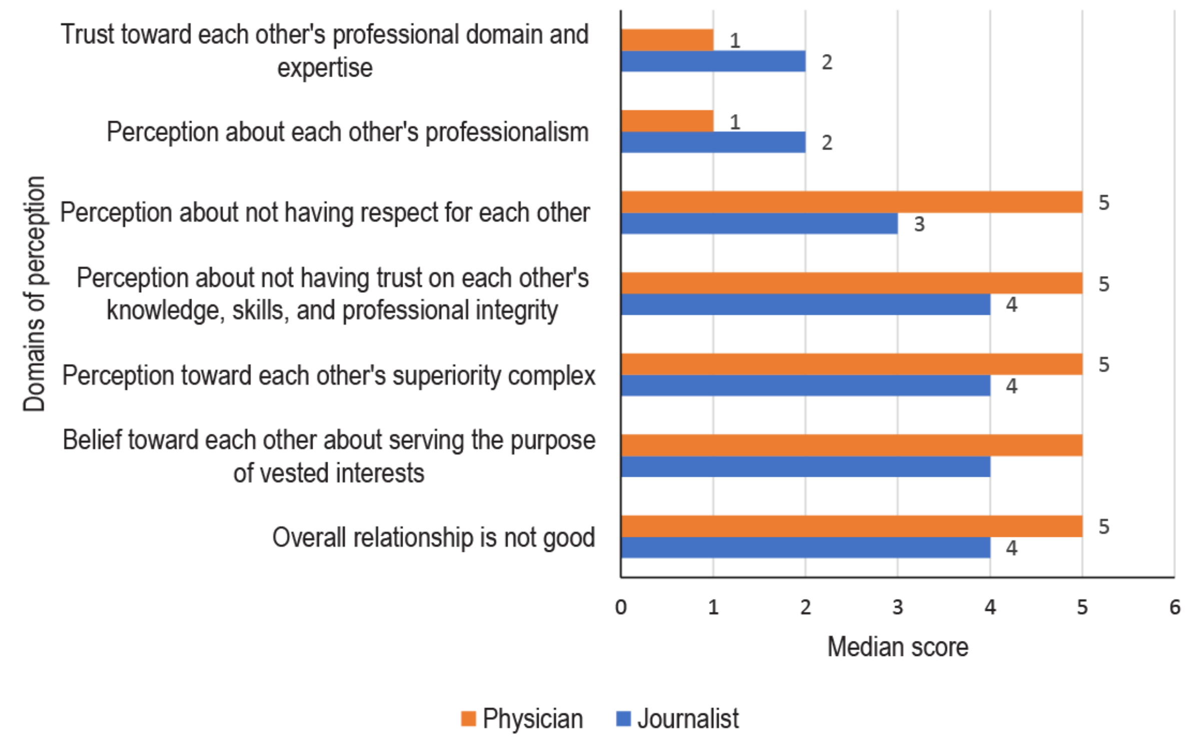
Question	Physicians (n=219), n (%)					Journalists (n=200), n (%)					*P* value[a]
	1[b]	2[c]	3[d]	4[e]	5[f]	1	2	3	4	5	
Trust toward each other's professional domain and expertise	117 (53.4)	34 (15.5)	57 (26.0)	5 (2.3)	6 (2.7)	87 (43.5)	70 (35.0)	39 (19.5)	3 (1.5)	1 (0.5)	.72
Perception about each other's professionalism	123 (56.2)	54 (24.7)	38 (17.4)	2 (0.9)	2 (0.9)	65 (32.5)	77 (38.5)	47 (23.5)	7 (3.5)	4 (2.0)	<.001
Perception about not having respect for each other	1 (0.5)	31 (14.2)	21 (9.6)	55 (25.1)	111 (50.7)	7 (3.5)	47 (23.5)	49 (24.5)	76 (38.0)	21 (10.5)	<.001
Perception about not having trust in each other's knowledge, skills, and professional integrity	0 (0)	31 (14.2)	20 (9.1)	55 (25.1)	113 (51.6)	4 (2.0)	35 (17.5)	49 (24.5)	103 (51.5)	9 (4.5)	<.001
Perception toward each other's superiority complex	0 (0)	4 (1.8)	3 (1.4)	22 (10.1)	190 (86.8)	7 (3.5)	15 (7.5)	17 (8.5)	123 (61.5)	38 (19.0)	<.001
Belief toward each other about serving the purpose of vested interests	0 (0)	1 (0.5)	10 (4.6)	28 (12.8)	180 (82.2)	2 (1.0)	15 (7.5)	28 (14.0)	104 (52.0)	51 (25.5)	<.001
Overall relationship is not good	0 (0)	0 (0)	2 (0.9)	21 (9.6)	196 (89.5)	1 (0.5)	19 (9.5)	27 (13.5)	116 (58.0)	37 (18.5)	<.001

[a]Mann-Whitney *U* test.
[b]Very low/strongly disagree.
[c]Slightly low/slightly disagree.
[d]Neither low/agree nor high/disagree.
[e]Slightly high/slightly agree.
[f]Very high/strongly agree.

Median Values of the 5-Point Likert Scale Ratings for Different Perception Aspects

In terms of the statements "Trust toward each other's professional domain and expertise" and "Perception about each other's professionalism," the median value for the physicians was 1, whereas it was 2 for the journalists. In terms of "Perception about not having respect for each other," the median value for the physicians was 5, whereas it was 3 for the journalists. In terms of other variables such as "Not having trust in each other's knowledge, skills, and professional integrity"; "Perception toward each other's superiority complex"; "Beliefs toward each other about serving the purpose of vested interests"; and "Overall relationship is not good," the median value for the physicians was 5, whereas it was 4 for the journalists (Figure 1).

Figure 1. Median values of the 5-point Likert scale on different domains of perceptions (1=strongly disagree and 5=strongly agree) for journalists and physicians.



Factors Associated With Lack of Trust in Each Other's Knowledge, Skills, and Professional Integrity

The ordered logistic regression analysis showed that several factors—sex and the professional title for physicians, and age, designation, and current workplace for journalists—were significantly associated with the lack of trust in each other's knowledge, skills, and professional integrity (Table 3).

In terms of the background characteristics of physicians, male physicians had significantly higher odds of lacking trust in journalists' knowledge, skills, and professional integrity compared to their female counterparts. In terms of professional title, medical officers had significantly higher odds of lacking trust in journalists' knowledge, skills, and professional integrity compared to the specialists.

On the other side, in terms of the background characteristics of journalists, a 1-year increase in journalist age increased the odds of lacking trust in physicians' knowledge, skills, and professional integrity by 0.14. The reporters and news editors had higher odds of lacking trust in physicians' knowledge, skills, and professional integrity compared to the reference category, correspondents. In terms of current workplace, journalists from other city corporations and from the district level had higher odds of lacking trust in physicians' knowledge, skills, and professional integrity compared with their reference category, journalists from Dhaka.

Table 3. Association of perceptions about not having trust in each other's knowledge, skills, and professional integrity with background characteristics of physicians and journalists.

Variables	Physicians (n=219)[a]		Journalists (n=200)[b]	
	AOR[c] (95% CI)	*P* value	AOR (95% CI)	*P* value
Age (years)	1.1 (1-1.21)	.06	1.14 (1.07-1.22)	<.001
Years of professional experience	0.91 (0.82-1.02)	.10	0.93 (0.87-1.00)	.06
Sex				
Male (reference)	N/A[d]	N/A	N/A	N/A
Female	0.45 (0.25-0.81)	.008	1.88 (0.96-3.67)	.07
Prefer not to say	0 (0-0)	.99	4.58 (0.44-47.85)	.20
Physicians' professional title				
Medical officer	N/A	N/A	N/A	N/A
Junior consultant	0.35 (0.11-1.15)	.09	N/A	N/A
Consultant	0.43 (0.17-1.12)	.09	N/A	N/A
Specialist	0.3 (0.13-0.71)	.006	N/A	N/A
Journalists' professional title				
Correspondent	N/A	N/A	N/A	N/A
Reporter	N/A	N/A	3.19 (1.42-7.15)	.005
News editor	N/A	N/A	3.34 (1.39-8.06)	.007
Others (eg, anchor, media manager)	N/A	N/A	3.44 (0.75-15.76)	.11
Current working place				
Capital city (Dhaka)	N/A	N/A	N/A	N/A
Others divisional city	0.59 (0.29-1.19)	.14	4.26 (1.67-10.88)	.002
District	0.73 (0.38-1.41)	.35	3.56 (1.51-8.38)	.004
Upazila[e] and below	0.55 (0.21-1.41)	.21	1.2 (0.43-3.36)	.73

[a]Model parameters: Likelihood ratio (χ^2_{10})=24.62, *P*=.006; Pseudo R^2=0.047.

[b]Model parameters: Likelihood ratio (χ^2_{10})=39.54, *P*=.001; Pseudo R^2=0.081.

[c]AOR: adjusted odds ratio.

[d]N/A: not applicable.

[e]An administrative division in Bangladesh, functioning as a subunit of a district.

Physicians' Perceptions Toward Journalists

Table 4 presents the physicians' perceptions of journalists in Bangladesh, measured through 12 variables, 6 of which were statistically significant ($P<.05$). A few of the significant findings are summarized below.

Most physicians "strongly agreed" with the statements "Journalists often prepare news stories on their own first and then talk to physicians," "Journalists often present health and medical information in a sensational way," "Journalists often use the term 'wrong treatment' without considering the context or details," and "Journalists often tend to publish news stories on health care professionals and health care services without adequate verification."

Table 4. Physicians' perceptions toward journalists (N=219).

Variable	Strongly disagree, n (%)	Slightly disagree, n (%)	Neither agree nor disagree, n (%)	Slightly agree, n (%)	Strongly agree, n (%)	*P* value[a]
Journalists often write and publish news on the health sector without having adequate knowledge about it	0 (0)	9 (4.1)	14 (6.4)	33 (15.1)	163 (74.4)	.26
Journalists often write and publish news stories based on their preconceived ideas	0 (0)	2 (0.9)	28 (12.8)	34 (15.5)	155 (70.8)	.44
Journalists often prepare news stories on their own first and then talk to physicians	1 (0.5)	2 (0.9)	36 (16.4)	30 (13.7)	150 (68.5)	<.001
Journalists often do not try to understand the real situation or the underlying meaning of a medical situation; rather, they are more interested in what they want to know	1 (0.5)	3 (1.4)	5 (2.3)	28 (12.8)	182 (83.1)	.63
In most cases, journalists present a distorted picture of health professionals and health care services	0 (0)	4 (1.8)	4 (1.8)	27 (12.3)	184 (84.0)	.48
The media always publish biased information on the health sector and health professionals	0 (0)	2 (0.9)	5 (2.3)	25 (11.4)	187 (85.4)	.13
Journalists often present health and medical information in a sensational way	1 (0.5)	1 (0.5)	2 (0.9)	25 (11.4)	190 (86.8)	.002
Journalists often use the term "wrong treatment" without considering the context or details	0 (0)	0 (0)	1 (0.5)	17 (7.8)	201 (91.8)	.004
In most cases, journalists incorrectly quote physicians or health care professionals in their news stories	0 (0)	0 (0)	6 (2.7)	23 (10.5)	190 (86.8)	.32
Journalists often tend to publish news stories on health care professionals and health care services without adequate verification	0 (0)	1 (0.5)	2 (0.9)	22 (10.1)	194 (88.6)	<.001
I am afraid of talking to journalists as they do not know how to ask questions objectively/neutrally	1 (0.5)	7 (3.2)	8 (3.7)	22 (10.1)	181 (82.7)	.01
Journalists tend to believe that most physicians are not qualified and inhumane	0 (0)	1 (0.5)	8 (3.7)	22 (10.1)	188 (85.8)	<.001

[a]Wilcoxon signed-rank test.

Journalists' Perceptions Toward Physicians

Table 5 presents the journalists' perceptions of physicians in Bangladesh, measured through 9 variables, 7 of which were statistically significant (*P*<.05). In particular, most journalists "slightly agreed" with the statements "During an interview or in case of communicating information relevant to a news story, most physicians tend not to give enough time to journalists"; "While talking to media, physicians use jargon and difficult terms that are not understandable for ordinary persons"; "During an interview, physicians often try to dominate over journalists"; and "Physicians often try to avoid media and journalists as a result of their professional supremacy attitude."

Table 5. Journalists' perceptions toward physicians (N=200).

Variables	Strongly disagree, n (%)	Slightly disagree, n (%)	Neither agree nor disagree, n (%)	Slightly agree, n (%)	Strongly agree, n (%)	*P* value[a]
Physicians often tend to believe that journalists do not have adequate knowledge about the country's health care system	5 (2.5)	24 (12.0)	35 (17.5)	119 (59.5)	17 (8.5)	.22
When contacting for any information relevant to a story, physicians often pretend that they are too busy	3 (1.5)	26 (13.0)	36 (18.0)	108 (54.0)	27 (13.5)	.02
Physicians often seem not to be confident while appearing in media or talking to journalists	1 (0.5)	27 (13.5)	43 (21.5)	120 (60.0)	9 (4.5)	.70
Most physicians are not skilled in giving an interview or talking to journalists	4 (2.0)	41 (20.5)	36 (18.0)	100 (50.0)	19 (9.5)	.02
During an interview or in case of communicating information relevant to a news story, most physicians tend to not give enough time to journalists	4 (2.0)	19 (9.5)	35 (17.5)	128 (64.0)	14 (7.0)	.002
While talking to the media, physicians use jargon and difficult terms that are not understandable to ordinary people	2 (1.0)	17 (8.5)	32 (16.0)	109 (54.5)	40 (20.0)	.01
Physicians often do not feel the need to present medical information in a simple, straightforward manner	3 (1.5)	24 (12.0)	33 (16.5)	116 (58.0)	24 (12.0)	<.001
During an interview, physicians often try to dominate journalists	18 (9.0)	25 (12.5)	48 (24.0)	96 (48.0)	13 (6.5)	<.001
Physicians often try to avoid media and journalists as a result of their professional supremacy attitude	3 (1.5)	13 (6.5)	34 (17.0)	124 (62.0)	26 (13.0)	.01

[a]Wilcoxon signed-rank test.

The Way Forward From the Perspectives of Physicians and Journalists

When rating the statement "Regular professional interaction between journalists and doctors may improve the relationship between the professional groups," most physicians (85%) chose "neither agree nor disagree," whereas most journalists (53%) stated that they "slightly agree" (Figure 2).

When rating the statement "Necessary training may improve the relationship between the professional groups," most physicians (90%) chose "strongly agree," whereas most journalists (50%) opted for "slightly agree" (Figure 3).

Figure 2. Responses to the statement "More interaction between physicians and journalists may improve the relationship between the professional groups.".

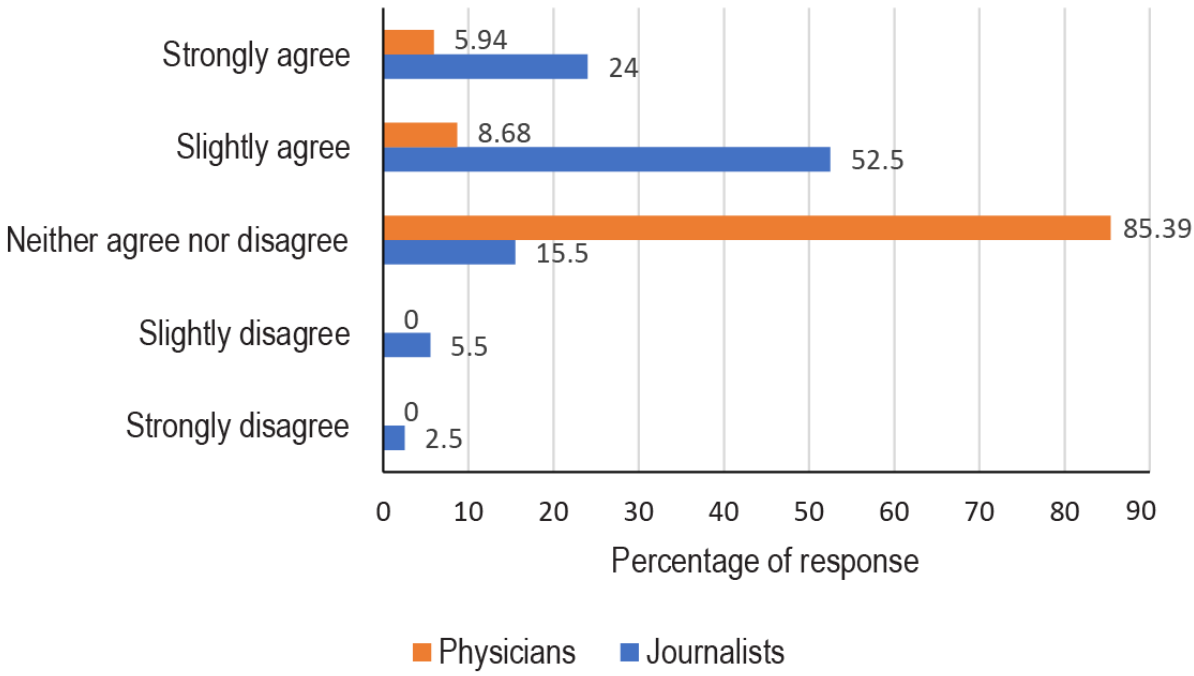
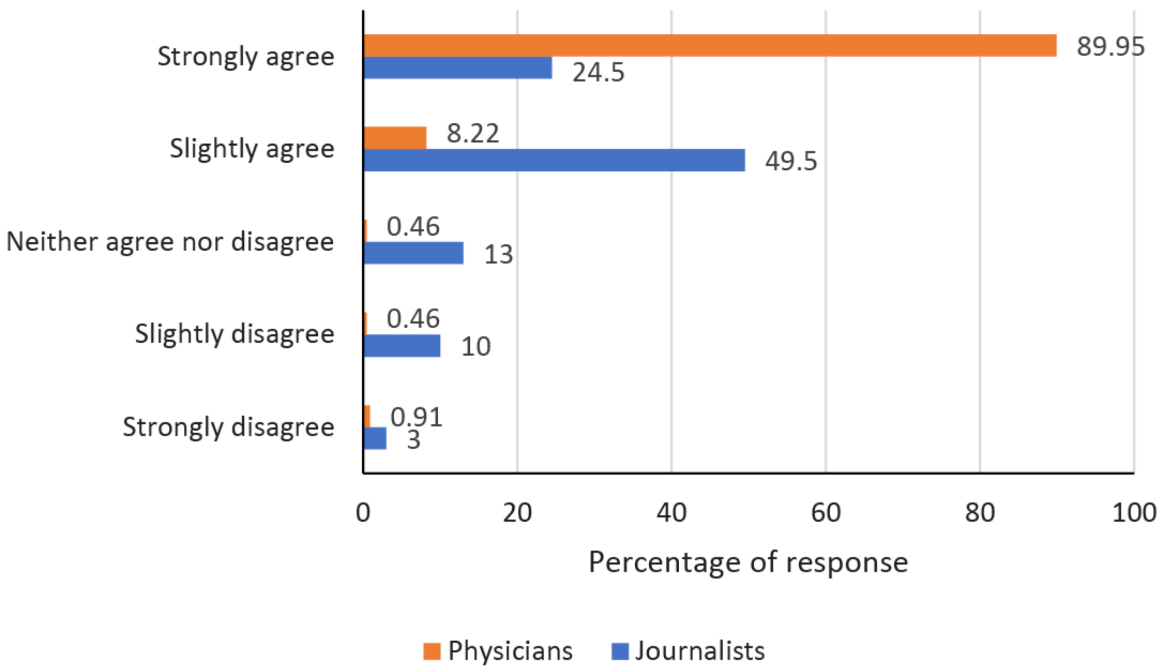


Figure 3. Responses to the statement "Necessary training may improve the relationship between the professional groups.".



Discussion

Principal Findings

From the overall results, it is evident that both physicians and journalists in Bangladesh have a negative perception of each other. However, physicians have a more negative perception of the journalists than the journalists have of the physicians. We also found that the attitude of male and junior physicians toward the journalists is more negative compared with the attitudes of other physician groups.

The negative attitude of the professional groups has been exemplified by the variables such as trust toward the opposite professional group; perception regarding professionalism and not having respect for each other; and not having trust in each other's knowledge, skills, and professional integrity. The high level of negative perceptions of physicians may be attributed to several factors. For example, in medical college, students are given an impression of superiority over other professional groups. This phenomenon has been observed by Zaman [38] in the Bangladeshi context. There is a historical link between traditional enmity and distrust toward each other [39-41]. It is also possible that a lack of knowledge, understanding, and training about medical and health issues among journalists often leads to misreporting on these issues, which harms the professional credentials of physicians. The issue of misreporting was echoed in the opinions of physicians that participated in this study, as they believe that journalists often present health and medical information in a sensational way, often use the term “wrong treatment” without considering the context or details, and tend to publish news stories on health care professionals and health care services without adequate verification. A similar perception of physicians toward journalists was observed by Ahlmén-Laiho et al [42,43] in the context of Finland. However, these authors argued that journalists’ experience of collaboration with physicians was positive. There are several cases of journalists portraying incidents of patients dying at a hospital or a clinic as an outcome of the ignorance and negligence of physicians, which could be a reason for the enmity from the physicians’ perspective. The negative portrayal of individual physicians, and the health care system as a whole, may affect public trust in physicians and the health care system [44-46]. In another study, Ahlmén-Laiho et al [42] found that physicians often do not trust health information published or broadcasted in news media.

Our results indicate that physicians in Bangladesh are not comfortable talking to journalists, while the journalists are skeptical of physicians’ communication skills. This might be due to the lack of communication skills among physicians, as the medical training curriculum does not adequately include behavioral science and communication skills, especially on how to face media. This failure often leads to negative perceptions of physicians among journalists, which is reflected in their news reporting. The negative perception may stem from the low standard of general journalistic practices, particularly in reporting on medical and health issues. Journalists often work on tight deadlines and write overstated headlines while covering health issues, compromising the relationship between professional groups [47]. However, Leask et al [1] argued that the relationship could be improved through physicians’ increased awareness of journalists’ work culture and daily routines, being available when journalists request an interview or any piece of information for their news stories, providing them with necessary resources, and building relationships with specialist health reporters.

Our results show that the attitude of junior physicians is more negative toward journalists compared to that of senior physicians, aligning with previous findings in the United Kingdom [48]. This attitude might be attributed to the communication skills and experience of facing journalists in professional encounters. Perhaps junior physicians might have less developed skills in facing media. However, journalists may also be less critical of senior physicians that have greater experience in facing media and journalists. We also found that male physicians and medical officers are more negative toward journalists compared to specialists. The same explanation may apply to this finding.

Overall, our results show that the relationship between two professional groups—journalists and physicians—is not good, concurring with the results obtained in previous studies [49]. The differences in work cultures often lead to negative perceptions. For example, in their study on 600 medical experts in 21 countries, Larsson and colleagues [47] found that the nature of journalists’ work, short deadlines, writing populist headlines, their choice of topics or angles in news stories, and their level of medical knowledge are some of the barriers to overcome to improve the quality of medical reporting. However, the negative attitudes of the two professional groups toward each other are harmful to the quality of care and may undermine their professional motivation.

Empirical evidence indicates that the lack of adequate communication between the two professional groups is one of the key barriers to the dissemination of public health information in a country. An improved relationship between the two professional groups to enhance their understanding of each other’s work culture is thus required. Medical colleges should incorporate issues related to communication skills with both patients and news media in the medical training of their graduates. On the other side, journalism schools should incorporate medical and health issues in their curricula so that future journalists can be equipped with the necessary medical knowledge. Moreover, the government should formulate a legal system to address medical negligence to ensure evidence-based representation of health issues.

Limitations

This study has some limitations. First, the participants of this study were only selected among users of social media platforms and email. Thus, there may be response bias, and only those concerned more about medicine and media among physicians and journalists may have participated in the study. Second, the study is based on a nonrepresentative sample size. Therefore, the results may not be generalizable to other physicians and journalists who are not social media or email users. Third, this study was exploratory, which did not allow for making any causal inferences.

Strengths

Despite these limitations, the study’s strength lies in the fact that it provides the first scientific evidence on the relationship between physicians and journalists in Bangladesh, to our best knowledge. Another key strength of the study is that it has considered the perspectives of both physicians and journalists to reveal the professional relationship between these two groups.

Further qualitative research is nonetheless still needed to understand this phenomenon in greater depth. Qualitative formative research will help design interventions to enrich professional skill sets, responsiveness, and communication strategies to improve and maintain a sustainable healthy relationship.

Conclusion

Both physicians and journalists in Bangladesh have negative perceptions of each other. The perception of physicians toward journalists appears to be more negative than the perception of journalists toward physicians. Our findings suggest that several strategies could be adopted to improve the existing unhealthy relationship between these two important professional groups in Bangladesh and other contexts. First, a legal framework is needed to identify medical-legal issues in reporting. Moreover, policy makers should take initiatives related to constructive discussion, professional interaction, and capacity-building training programs, which may significantly improve the relationship between physicians and journalists.

Acknowledgments

The authors are grateful to the Public Health Foundation, Bangladesh, and its members for providing scientific guidance for conducting the research. The authors are also grateful to the physicians and journalists for their voluntary participation in this study. No funding was received at any stage of this research, including study design, data collection, data analysis, writing the manuscript, and publication of this work.

Data Availability

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Authors' Contributions

MAI contributed to conceptualizing the research idea. MAI, MGR, and TJ contributed to the study design. MAI and ZR contributed to developing the study tools and data collection. MGR and TJ contributed to the data analysis. MAI coordinated the whole research, including data collection, data curation, and drafting and writing the manuscript. All authors contributed to the literature search, and writing, revising, and finalizing the manuscript. All authors read, revised, and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study questionnaire.
[DOCX File , 29 KB - ijmr_v12i1e44116_app1.docx]

Multimedia Appendix 2
Study data file.
[XLSX File (Microsoft Excel File), 72 KB - ijmr_v12i1e44116_app2.xlsx]

References

1. Leask J, Hooker C, King C. Media coverage of health issues and how to work more effectively with journalists: a qualitative study. BMC Public Health 2010 Sep 08;10:535 [FREE Full text] [doi: 10.1186/1471-2458-10-535] [Medline: 20822552]
2. Viswanath K, Blake KD, Meissner HI, Saiontz NG, Mull C, Freeman CS, et al. Occupational practices and the making of health news: a national survey of US Health and medical science journalists. J Health Commun 2008 Dec;13(8):759-777. [doi: 10.1080/10810730802487430] [Medline: 19051112]
3. Henderson L, Hilton S. The media and public health: where next for critical analysis? Crit Public Health 2018 Jun 29;28(4):373-376. [doi: 10.1080/09581596.2018.1482663]
4. Koon AD, Hawkins B, Mayhew SH. Framing and the health policy process: a scoping review. Health Policy Plan 2016 Jul;31(6):801-816 [FREE Full text] [doi: 10.1093/heapol/czv128] [Medline: 26873903]
5. Rowbotham S, McKinnon M, Marks L, Hawe P. Research on media framing of public policies to prevent chronic disease: a narrative synthesis. Soc Sci Med 2019 Sep;237:112428. [doi: 10.1016/j.socscimed.2019.112428] [Medline: 31357110]
6. Briggs C, Hallin D. Making health public: How news coverage is remaking media, medicine, and contemporary life. Milton Park, UK: Routledge; 2016.
7. Stroobant J, Van den Bogaert S, Raeymaeckers K. When medicine meets media: how health news is co-produced between health and media professionals. Journalism Stud 2018 Nov 01;20(13):1828-1845. [doi: 10.1080/1461670X.2018.1539344]
8. Van den Bogaert S, Stroobant J, Bracke P. (Dis)entangling medicine and media: a qualitative analysis of the relationship between the fields of healthcare and journalism. Health Sociol Rev 2018 Oct 22;28(1):69-84. [doi: 10.1080/14461242.2018.1537131]
9. Fitzgerald PE, Wilson L. Relations with news media: perceptions of healthcare executives. Hosp Top 1999;77(4):20-23. [doi: 10.1080/00185869909596534] [Medline: 11009939]
10. Quattrin R, Filiputti E, Brusaferro S. Health promotion campaigns and mass media: looking for evidence. Primary Health Care 2015;05(01):1. [doi: 10.4172/2167-1079.1000190]

11. Dunlop SM, Wakefield M, Kashima Y. Pathways to persuasion: cognitive and experiential responses to health-promoting mass media messages. Commun Res 2009 Dec 01;37(1):133-164. [doi: 10.1177/0093650209351912]
12. Nehme Z, Cameron PA, Akram M, Patsamanis H, Bray JE, Meredith IT, et al. Effect of a mass media campaign on ambulance use for chest pain. Med J Aust 2017 Jan 16;206(1):30-35. [doi: 10.5694/mja16.00341] [Medline: 28076734]
13. Van Teijlingen E, Simkhada P, Luce A, Hundley V. J Manmohan Memorial Inst Health Sci 2016 Sep 26;2:70-75. [doi: 10.3126/jmmihs.v2i0.15799]
14. Maryon-Davis A. Using the mass media to promote health. InnovAiT 2012 Nov 02;5(12):767-773. [doi: 10.1093/innovait/ins191]
15. Uzun SU, Zencir M. Reliability and validity study of the Turkish version of cyberchondria severity scale. Curr Psychol 2018 Sep 18;40(1):65-71. [doi: 10.1007/s12144-018-0001-x]
16. Cheng J, Benassi P, De Oliveira C, Zaheer J, Collins M, Kurdyak P. Impact of a mass media mental health campaign on psychiatric emergency department visits. Can J Public Health 2016 Oct 20;107(3):e303-e311 [FREE Full text] [doi: 10.17269/cjph.107.5265] [Medline: 27763847]
17. Grilli R, Ramsay C, Minozzi S. Mass media interventions: effects on health services utilisation. Cochrane Database Syst Rev 2002(1):CD000389. [doi: 10.1002/14651858.CD000389] [Medline: 11869574]
18. Elrod JK, Fortenberry JL. Advertising in health and medicine: using mass media to communicate with patients. BMC Health Serv Res 2020 Sep 15;20(Suppl 1):818 [FREE Full text] [doi: 10.1186/s12913-020-05599-3] [Medline: 32928190]
19. Jamshidi E, Nedjat S, Nedjat S, Nikooee S, Rostamigooran N, Majdzadeh R. How to utilize tacit knowledge in health organizations: an Iranian perspective. Med J Islam Repub Iran 2018 Sep 30;32:116-685 [FREE Full text] [doi: 10.14196/mjiri.32.116] [Medline: 30815411]
20. Dhanani LY, Franz B. The role of news consumption and trust in public health leadership in shaping COVID-19 knowledge and prejudice. Front Psychol 2020;11:560828 [FREE Full text] [doi: 10.3389/fpsyg.2020.560828] [Medline: 33192827]
21. Gille F, Smith S, Mays N. Towards a broader conceptualisation of 'public trust' in the health care system. Soc Theory Health 2016 Nov 17;15(1):25-43. [doi: 10.1057/s41285-016-0017-y]
22. Tokuda Y, Fujii S, Jimba M, Inoguchi T. The relationship between trust in mass media and the healthcare system and individual health: evidence from the AsiaBarometer Survey. BMC Med 2009 Jan 22;7(1):4 [FREE Full text] [doi: 10.1186/1741-7015-7-4] [Medline: 19161600]
23. Gille F, Smith S, Mays N. Why public trust in health care systems matters and deserves greater research attention. J Health Serv Res Policy 2015 Jan;20(1):62-64. [doi: 10.1177/1355819614543161] [Medline: 25038059]
24. Manganello JA. Health literacy and adolescents: a framework and agenda for future research. Health Educ Res 2008 Oct;23(5):840-847. [doi: 10.1093/her/cym069] [Medline: 18024979]
25. Levin-Zamir D, Lemish D, Gofin R. Media Health Literacy (MHL): development and measurement of the concept among adolescents. Health Educ Res 2011 Apr;26(2):323-335. [doi: 10.1093/her/cyr007] [Medline: 21422003]
26. Briggs C, Hallin D. Biocommunicability. Soc Text 2007;25(4):66. [doi: 10.1215/01642472-2007-011]
27. Gieryn TF. Boundary-work and the demarcation of science from non-science: strains and interests in professional ideologies of scientists. Am Sociol Rev 1983;48(6):781-795 [FREE Full text] [doi: 10.2307/2095325]
28. Schwitzer G, Mudur G, Henry D, Wilson A, Goozner M, Simbra M, et al. What are the roles and responsibilities of the media in disseminating health information? PLoS Med 2005 Jul 26;2(7):e215 [FREE Full text] [doi: 10.1371/journal.pmed.0020215] [Medline: 16033311]
29. Kenix L. Journalistic boundary work at a time of professional crisis. Aust Journal Rev 2016;38(1):32 [FREE Full text]
30. Revers M. Journalistic professionalism as performance and boundary work: Source relations at the state house. Journalism 2013 Mar 19;15(1):37-52. [doi: 10.1177/1464884913480459]
31. Baca M. Professional boundaries and dual relationships in clinical practice. J Nurse Practit 2011 Mar;7(3):195-200. [doi: 10.1016/j.nurpra.2010.10.003]
32. Jones I. The theory of boundaries: impact on interprofessional working. J Interprof Care 2007 Jun;21(3):355-357. [doi: 10.1080/13561820701257383] [Medline: 17487712]
33. Collins S. Elements of storytelling in medical journalism. Medical Writing 2015 Dec 23;24(4):222-224. [doi: 10.1179/2047480615z.000000000328]
34. Whelan J. Medical journalism: another way to write about science. Medical Writing 2015 Dec 23;24(4):219-221. [doi: 10.1179/2047480615z.000000000327]
35. Linden T. A delicate balance-ethical standards for physician-journalists. Virtual Mentor 2011 Jul 01;13(7):490-493 [FREE Full text] [doi: 10.1001/virtualmentor.2011.13.7.pfor1-1107] [Medline: 23134800]
36. Taber KS. The use of Cronbach's alpha when developing and reporting research instruments in science education. Res Sci Educ 2017 Jun 7;48(6):1273-1296. [doi: 10.1007/s11165-016-9602-2]
37. van Griethuijsen RALF, van Eijck MW, Haste H, den Brok PJ, Skinner NC, Mansour N, et al. Global patterns in students' views of science and interest in science. Res Sci Educ 2014 Nov 9;45(4):581-603. [doi: 10.1007/s11165-014-9438-6]
38. Zaman S. Broken limbs, broken lives. Ethnography of a hospital ward in Bangladesh. Amsterdam: Het Spinhuis; 2005.
39. Zaman S. Native among the Natives: physician anthropologist doing hospital ethnography at home. J Contemp Ethnogr 2008 Feb 04;37(2):135-154 [FREE Full text] [doi: 10.1177/0891241607312495]

40. Zaman S. Poverty and violence, frustration and inventiveness: hospital ward life in Bangladesh. Soc Sci Med 2004 Nov;59(10):2025-2036. [doi: 10.1016/j.socscimed.2004.03.007] [Medline: 15351470]
41. Zaman S. Silent saviours: family members in a Bangladeshi hospital. Anthropol Med 2013;20(3):278-287. [doi: 10.1080/13648470.2013.827426] [Medline: 24004293]
42. Ahlmén-Laiho U, Suominen S, Järvi U, Tuominen R. Reliability of health information in the media as defined by Finnish physicians. 2018 Presented at: Well-Being in the Information Society. Fighting Inequalities: 7th International Conference; August 27-29, 2018; Turku, Finland. [doi: 10.1007/978-3-319-97931-1_15]
43. Ahlmén-Laiho U, Suominen S, Järvi U, Tuominen R. Finnish health journalists' perceptions of collaborating with medical professionals. In: Saranto K, Castrén M, Kuusela T, Hyrynsalmi S, Ojala S, editors. Safe and Secure Cities. WIS 2014. Communications in Computer and Information Science, vol 450. Cham: Springer; 2014.
44. Peters D, Youssef FF. Public trust in the healthcare system in a developing country. Int J Health Plann Manage 2016 Apr;31(2):227-241. [doi: 10.1002/hpm.2280] [Medline: 25533779]
45. Van der Schee E. Public trust in health carexploring the mechanisms. Nivel. 2016. URL: https://www.nivel.nl/sites/default/files/bestanden/Proefschrift_Public_trust_in_health_care_schee.pdf [accessed 2023-06-09]
46. Williams SN, Armitage CJ, Tampe T, Dienes K. Public perceptions and experiences of social distancing and social isolation during the COVID-19 pandemic: a UK-based focus group study. BMJ Open 2020 Jul 20;10(7):e039334 [FREE Full text] [doi: 10.1136/bmjopen-2020-039334] [Medline: 32690752]
47. Larsson A, Appel S, Sundberg CJ, Rosenqvist M. Medicine and the media: medical experts' problems and solutions while working with journalists. PLoS One 2019;14(9):e0220897 [FREE Full text] [doi: 10.1371/journal.pone.0220897] [Medline: 31513581]
48. Goldacre MJ, Evans J, Lambert TW. Media criticism of doctors: review of UK junior doctors' concerns raised in surveys. BMJ 2003 Mar 22;326(7390):629-630 [FREE Full text] [doi: 10.1136/bmj.326.7390.629] [Medline: 12649237]
49. Stamm K, Williams JW, Noël PH, Rubin R. Helping journalists get it right: a physicians's guide to improving health care reporting. J Gen Intern Med 2003 Feb;18(2):138-145 [FREE Full text] [doi: 10.1046/j.1525-1497.2003.20220.x] [Medline: 12542589]

Edited by A Mavragani; submitted 10.11.22; peer-reviewed by DMA Awal, N Hu, G Mehdi, AM Auvinen; comments to author 09.03.23; revised version received 12.04.23; accepted 03.05.23; published 10.07.23.

Please cite as:
Islam MA, Rabbani MG, Rahaman Z, Joarder T
Professional Relationship Between Physicians and Journalists in Bangladesh: Web-Based Cross-Sectional Study
Interact J Med Res 2023;12:e44116
URL: https://www.i-jmr.org/2023/1/e44116
doi:10.2196/44116
PMID:37428550

Original Paper

Animated Videos Based on Food Processing for Guidance of Brazilian Adults: Validation Study

Maria Fernanda Gomes da Silva[1*], MSc; Luciana Neri Nobre[2*], PhD; Edson da Silva[3*], PhD

[1]Graduate Program in Nutrition Sciences, Federal University of Jequitinhonha and Mucuri Valleys, Diamantina, Brazil

[2]Department of Nutrition, Federal University of Jequitinhonha and Mucuri Valleys, Diamantina, Brazil

[3]Department of Basic Sciences, Federal University of Jequitinhonha and Mucuri Valleys, Diamantina, Brazil

[*]all authors contributed equally

Corresponding Author:
Edson da Silva, PhD
Department of Basic Sciences
Federal University of Jequitinhonha and Mucuri Valleys
Edifício DCB-DCBIO, Campus JK
Rodovia MGT 367, Alto da Jacuba, 5000
Diamantina, 39100-000
Brazil
Phone: 55 383532 1200
Email: edson.silva@ufvjm.edu.br

Abstract

Background: Ultraprocessed foods (UPFs) contribute almost one-fifth of the calories consumed by the Brazilian population. This consumption has been favored by aspects such as the ease of acquisition and low cost of this food group. Initiatives focused on supporting and promoting healthy eating practices have been implemented. Among them, the availability of educational resources is an important strategy to maximize the effectiveness of these actions in the field of food and nutrition education (FNE).

Objective: This study aims to describe the development and validation process of animated videos based on the NOVA food classification for FNE actions aimed at Brazilian adults.

Methods: This methodological study was developed in the following 4 phases: planning, preproduction, production, and postproduction. In the planning phase, a literature review was con-ducted on the topic and to define the content to be covered. The design of the material was based on the cognitive theory of multimedia learning. In the preproduction phase, video scripts were developed and evaluated by 7 content specialists. In the production phase, videos were developed based on the assessed scripts and then assessed by 3 multimedia production specialists. In the postproduction phase, the videos were evaluated by 15 representatives of the target audience. All results obtained in the evaluation phases were analyzed using the content validity index (CVI).

Results: We developed 3 animated videos covering the following themes: food processing levels, food categories according to processing levels, and UPFs and their impact on health. In the evaluation by the content specialists, the scripts of videos 1, 2, and 3 obtained CVIs at the scale level and average method equal to 0.96, 0.98, and 0.98, respectively. When the animated videos were evaluated by multimedia production specialists and representatives of the target audience, these indexes were equal to 1.0. These results attest to the videos' adequacy and quality in communicating the addressed content.

Conclusions: The animated videos developed and validated in this study proved to be adequate for their purpose. Thus, it is expected that they will be an important instrument for FNE actions aimed at an adult audience and for disseminating the Dietary Guidelines for the Brazilian Population.

(Interact J Med Res 2023;12:e49092) doi:10.2196/49092

KEYWORDS

food, processed; dietary guidelines; nutrition policy; instructional films and videos; validation study; food classification; validation; educational videos; nutrition; Brazil

Introduction

Eating patterns are changing rapidly in the vast majority of countries, particularly in middle-income countries. The main changes in this nutritional transition involve the replacement of culinary preparations based on in natura or minimally processed foods with ultraprocessed foods (UPFs) [1].

In Brazil, according to data from the Surveillance of Risk and Protective Factors for Chronic Diseases by Telephone Inquiry in 2019 [2], UPF already contributed almost one-fifth of the calories consumed, especially from salty cookies and packaged snacks, industrialized bread, sweet cookies, and cold cuts and sausages. A survey carried out by Monteiro and collaborators [3] pointed out that, in middle-income countries, the average growth in sales of these products reached approximately 10%.

This pattern of consumption has led to an imbalance in the supply of nutrients and excessive intake of calories [1]. A recent study showed that, in 2019, UPF consumption was responsible for approximately 57,000 deaths in Brazil, corresponding to 10.5% of all premature deaths and 21.8% of all preventable deaths from chronic noncommunicable diseases (NCDs) in adults aged 30 years to 69 years [4].

Aspects such as the greater ease of purchasing UPFs and their low cost and high energy density have contributed to the increase in consumption of these foods, as well as to the current scenario of NCDs like obesity, diabetes, and hypertension, which in turn are associated with a large economic burden for the Unified Health System in Brazil [3-6]. Another concern is associated with equity, given that UPF consumption has increased more rapidly among low-income households [4].

However, despite the understanding that a balanced diet is an important factor for maintaining health, healthy eating practices encompass not only the foods that are commonly consumed but also cultural factors, ways of life, food availability, income, health literacy, marketing, and media, among other aspects [7]. Thus, food and nutrition education (FNE) has been recognized as a strategy for the prevention and control of food and nutrition problems that incorporates sociocultural, biological, and environmental dimensions in a transdisciplinary, intersectoral, and multiprofessional field [7].

In this practice, the Dietary Guidelines for the Brazilian Population (DGBP) [1] is the official Brazilian document that brings together recommendations for healthy eating practices and is configured as an instrument to support FNE actions. The golden rule, "always prefer *in natura* or minimally processed foods and culinary preparations to ultraprocessed foods," is the main recommendation of the DGBP and is based on the NOVA classification system, which groups foods according to the extent and purpose of the processing they have undergone into the following 4 groups: (1) unprocessed and minimally processed foods, (2) processed culinary ingredients, (3) processed foods, and (4) UPFs [1,3].

NOVA not only influenced the foundation of the DGBP but also guided its implementation in public nutrition and health policies in Brazil and has been increasingly favorable for the epidemiological and food culture fields [8]. Strategies including development of protocols and instructions have played an important role in qualifying health service professionals for this purpose [6,8]. However, educational actions demand educational resources that can promote equitable access and engagement that meet the different needs and levels of health literacy of individuals, families, and communities [9].

The rise of online communication has expanded possibilities to make health communication more attractive [10]. In addition, educational animations, when properly produced, have been seen as a positive strategy to communicate complex health messages [11], and the adoption of digital resources has been associated with the strengthening of the interface between communication, science, and society as a consequence of aspects related to motivation, dynamics, transversality, and the absence of geographic barriers [12,13].

The elaboration of educational materials is part of a planned strategy for the implementation of the DGBP. Developing and validating educational videos for application in FNE actions can contribute to communication effectiveness and dissemination of knowledge among professionals and society. However, until now, the creation of videos has not been described in published, structured studies. Therefore, the purpose of this study was to develop and validate a series of 3 animated videos to support educational actions aimed at Brazilian adults.

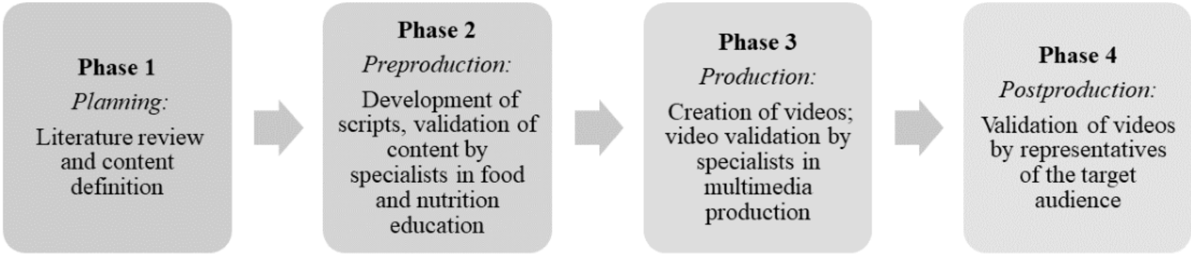
Methods

Study Types

This is a descriptive study of the development and validation of 3 animated videos based on the DGBP aimed at Brazilian adults.

The development of the videos was based on the stages of animation production proposed by Winder and Dowlatabadi [14] and Wright [15] and occurred in the following 4 phases: planning, preproduction, production, and postproduction (Figure 1).

Figure 1. Animated video development and validation process.



Video Development and Validation Process

Phase 1: Planning

A search was conducted in the scientific databases PubMed, Literatura Latino-Americana e do Caribe em Ciências da Saúde (LILACS), and Scientific Electronic Library Online (Scielo) with the objective of identifying studies and audiovisual educational materials aimed at the Brazilian adult population (18-59 years), based on food processing and DGBP, analyzing methodologies, limitations and possible advances for this proposal, and YouTube videos provided online. The searches in the databases were performed for articles published from 2014 to 2021, by combining the following descriptors (and their corresponding terms in Portuguese): "Food Guides" AND "Educational Film and Video" OR "Educational Technology." On YouTube, this search was made using the terms "Food Guide" and "NOVA Classification."

Next, the team that conducted the study applied the technique of brainstorming to collect and select ideas related to the topic and suitable for the target audience that the material would address. On this basis, we decided to draw on the content of Chapter 2 of the DGBP, "The choice of food," and the works published by Monteiro and collaborators [3,16], publications that are more recent than the NOVA classification.

This material was divided into 3 animated videos and planned to be used in FNE actions in a segmented way. The design was based on the principles and guidelines of the Framework of Reference for Food and Nutrition Education [7] and the applications of the cognitive theory of multimedia learning (CTML) [17].

The decision to develop and validate videos was based on an understanding of the 3 assumptions of CTML: (1) 2 channels (visual and auditory) for receiving and processing information, (2) limited capacity of information that can be processed in each channel, and (3) learning as an active process that involves organizing and integrating previous information into the most recent [17].

According to Mayer [17], the design of multimedia environments must be compatible with the way people learn. Therefore, the 12 didactic principles were considered in the design of the developed material: Coherence Principle (exclusion of unnecessary words, symbols, images, and sounds), Signaling Principle (highlighting key elements of the content covered), Redundancy Principle (using animation and narration instead of animation, narration, and text on the screen), Spatial Contiguity Principle (using pictograms and texts in spatial proximity), Temporal Contiguity Principle (simultaneously presenting pictograms and narration), Segmentation Principle (presenting the content in segments), Pre-training Principle (general presentation of the content before detailing it), Modality Principle (use of pictograms and narration instead of pictograms and text), Personalization Principle (use of conversational style words instead of formal style), Voice Principle (use of the narration in human voice), and Image Principle (the image of the narrator is dispensable) [17].

Phase 2: Preproduction

The central ideas for each video were planned to be approached in a sequential order seeking to deepen the theme of the previous media. Thus, a narrative text was developed for each material, taking this proposal into account. One of the premises for the material development was the method of approaching the content. To this end, the language of the material was planned and revised with the aim of minimizing the cognitive load of learning and allowing the target audience to better understand the topic. Then, each narrative text was divided into scenes, which were then described and outlined.

In the subsequent step, *storyboards* were developed, then a script for each video was structured in the following 3 parts: narration, scene description, and storyboard.

After the scripts were completed, the content was validated by specialists. At this stage, nutrition professionals were invited and selected for convenience. For the selection of these judges, knowledge and level of expertise were considered [18], adopting criteria from an existing model commonly used in validation studies in Brazil [19] that assigns a score for the judge's experience in the topic of interest. In this study, experience in FNE and DGBP was the considered criterion, as evidenced by participation in research, analysis of publications on the subject, and teaching in disciplines in the area. Having a master's degree was a minimum criterion. The identification of these specialists occurred through publications and searches in research programs of Brazilian universities. Professionals who obtained a minimum score of 5 points through curriculum analysis were considered able to evaluate the scripts [19].

After the selection of potential judges, the invitation to participate in the study was sent by email, which included a link to access a form developed in Google Forms with the informed consent form, a questionnaire to characterize the specialists' profiles, and the Educational Content Validation Instrument in Health (ECVIH; Multimedia Appendix 1) [20], as well as a blank space for suggestions to improve the material. The recruitment of judges was completed when the evaluation of 7 specialists was complete, with the panel defined by the recommendation that evaluation by at least 5 judges can provide a sufficient level of control for casual agreement [21]. This evaluation process took place from April 2022 to May 2022.

Phase 3: Production of Animated Vídeos

After the script was evaluated by the content specialists, the requested suggestions and adjustments were analyzed, including adding or replacing examples and rewording excerpts from the narratives to make them more understandable.

Once the adjustments were completed, the narratives for each video were recorded. Based on the storyboards, the scenes were developed on the Canva platform. For this purpose, a mood board was first created with videos, photos, and pictograms. Based on this, each scene was created following the flipbook concept of the platform resources, gradually changing the position of the photos and pictograms if necessary. The entire sequence of images created was transferred to Movavi Video Editor and sequenced to create the impression of motion.

With the first version of each video, an evaluation of the appropriateness of the material was carried out by multimedia production specialists. The identification of these experts occurred via LinkedIn, and the selection took place by analyzing the curriculum. To be considered able to evaluate the videos, we selected those who obtained a minimum score of 5 points in aspects that encompassed knowledge and experience in the production of animations [19].

These specialists were contacted, and the proposal was presented. For those who agreed to participate in the study, an email was sent with the same structure as in the previous step, but we used the Suitability Assessment of Materials (SAM) [22] tool, translated and adapted for use in Brazil [23] (Multimedia Appendix 2). The objective of the SAM is to evaluate the suitability of educational material. The recruitment was completed when the assessment of 3 professionals was complete, with the panel defined by the recommended minimum number [21]. The evaluation process took place between August 2022 and September 2022.

Phase 4: Postproduction

After the evaluation of the videos was completed, the dissonances between the sound and text were corrected, and the pictograms were changed according to the suggestions from the experts. There was also a proposal to unify the visual language of the video; however, this proposal could not be fully implemented because there are few pictograms with the same language in the graphic elements database of the Canva platform. Alternative platforms with public domain images were researched, but no substitute was found that fulfilled the same communication purpose.

With the adjustments made, the animations were evaluated by representatives of the target audience. The sample size was calculated using the formula for a finite population: n=Za2.P(1-P)/e2, where *Za* corresponds to the confidence level, *P* is the expected proportion of judges who agree with the item, and *e* is the expected difference. The following values were assumed for this purpose: Za=95%, *P*=90%, and e=15%. This resulted in 15 participants [24].

Through snowball sampling [25], the research team selected and invited participants through their social networks. Each guest could indicate 1 to 2 people from their social network, and the latter could indicate at least one other participant. The inclusion criteria were adults of both sexes aged between 18 years and 59 years who did not present with impairments in cognitive status, vision, and hearing. Recruitment was complete when 15 representatives of the target group were identified.

Each participant had access to an informed consent form provided on a form created with Google Forms and, after confirming his or her consent, was directed to a questionnaire collecting sociodemographic information (age, sex, race/color, education) and a self-report of diseases related to cognitive status, vision, and hearing adopted from population-based surveys in Brazil [2,26]. The assessment tool used in this phase was the SAM [23] (Multimedia Appendix 2). The evaluation process took place from September 2022 to October 2022.

Data Analysis

The ratings by content specialists, multimedia production professionals, and representatives of the target audience were analyzed using the content validity index (CVI), which has the objective of measuring the percentage of evaluators who agree on the aspects analyzed in the instrument [27].

This index was calculated in the following 2 ways: (1) using the item-level content validity index (I-CVI), obtained by the sum of the judges' agreement on each item of the ECVIH and SAM instruments divided by the total number of responses, and (2) using the scale-level content validity index, average calculation method (S-CVI/Ave) to assess the average agreement of the judges, obtained by the sum of the I-IVC of each item divided by the total number of items of each instrument [27].

For general content validation, an index of 0.90 or higher was considered desirable for the S-CVI/Ave [28]. For the I-CVI, values no lower than 0.78 were considered [21]. Indexes that had lower values than those described were revised according to the experts' suggestions. All analyses were performed using Excel.

Ethics Approval

The study was conducted in accordance with Resolution 466/2012 [29] of the Brazilian National Health Council. The study was approved by the Research Ethics Committee of the Federal University of Vales do Jequitinhonha e Mucuri (CAAE: 47624721.9.0000.5108).

Results

In this study, 3 educational videos were developed based on the NOVA classification of foods [1,3,16]: food processing levels, food categories by processing levels, and UPFs and their effects on health.

Animated video 1 is 3 minutes and 45 seconds long (Multimedia Appendix 3) and focuses on addressing the extent and purpose of food processing, introducing concepts, and situating the public on the subject addressed. Animated video 2 is 3 minutes and 25 seconds long (Multimedia Appendix 4) and provides deeper content on the main differences between the food groups in the NOVA classification system. Animated video 3 is 4 minutes and 12 seconds long (Multimedia Appendix 5) and highlights the group of UPFs only, describes the characteristics of the foods in this group, and reaffirms the importance of avoiding daily consumption of this group. Figure 2 shows scenes from the 3 videos. To arrive at the final result, each animated video was evaluated for its content and suitability as an educational resource for the target audience; the following sections describe how this entire process unfolded.

Figure 2. Scenes from (A) animated video 1, exemplifying the levels of processing; (B) animated video 2, showing food groups from the NOVA classification; (C) animated video 3, which cites some of the characteristics of ultra-processed foods ("high in calories, sugar, sodium, fat, and additives"); as well as (D) recommendations from the Food Guide for the Brazilian Population.

The stage involving validation of the script content involved 7 experts in FNE, of whom 57% (n=4) were doctors and 43% (n=3) had a master's degree. They were residents of 4 federative units and 3 different regions of the country. In terms of professional experience, 71% (5/7) had teaching experience in FNE, and 57% (4/7) reported having at least 10 years of practical experience in the area. All participants had research or work experience related to the NOVA and DGBP classification systems. Regarding scientific production, 86% (6/7) had published a scientific article in an indexed journal in the area of FNE.

The ECVIH [20], which was used to evaluate the scripts, has 18 items. From the analyzed results, the evaluation of the script referring to animated video 1 (S1) showed a mean overall agreement (S-CVI/Ave) of 0.96. For the script for animated video 2 (S2), this agreement was equal to 0.98, and for the script for animated video 3 (S3), it was 0.98 (Table 1). According to the criteria by Lynn [21], the values obtained for S-CVI/Ave indicate excellent content validity.

Table 1. FNE experts' agreement regarding the content of the scripts (n=7).

Item	I-CVI[a]		
	S1[b]	S2[c]	S3[d]
Objectives: purposes, goals, or targets			
Contemplates the proposed theme	1	0.86	1
Suits the teaching-learning process	1	1	1
Clarifies doubts on the addressed theme	1	1	1
Provides reflection on the theme	0.86	1	0.86
Encourages behavior change	0.71	0.86	0.86
Structure and presentation: organization, structure, strategy, consistency, and sufficiency			
Language appropriate to the target audience	1	1	1
Language appropriate to the educational material	1	1	1
Interactive language, enabling active involvement in the educational process	0.86	1	1
Correct information	1	1	1
Objective information	1	0.86	0.86
Enlightening information	1	1	1
Necessary informations	1	1	1
Logical sequence of ideas	1	1	1
Current theme	1	1	1
Appropriate text size	0.86	1	1
Relevance: significance, impact, motivation, and interest			
Encourages learning	1	1	1
Contributes to knowledge in the area	1	1	1
Arouses interest in the theme	1	1	1
S-CVI/Ave[e]	0.96	0.98	0.98

[a]I-CVI: item-level content validity index.

[b]S1: script 1.

[c]S2: script 2.

[d]S3: script 3.

[e]S-CVI/Ave: scale-level content validity index, averaging method.

When analyzing the I-CVI results obtained for each script (Table 1), for S1, 14 items presented an I-CVI of 1.00, and 3 had an index equal to 0.86, results that prove the good quality of the produced content. Relevant items for the development of the material, especially regarding the language and content, obtained an excellent evaluation. The item "Encourages behavior change" referring to the "Objectives" domain, however, obtained a lower score (I-CVI=0.71), being judged as inadequate by 2 of the 7 judges. Following the observation made by Lynn [21], when 6 or more experts carry out an evaluation, if one or more experts disagree with the others, what must be evaluated is content validity.

A suggestion for this script was the inclusion of activities for reflection, a phase already foreseen in the research but not described in this work. In addition, adjustments were made to the beginning and end of the script to generate greater interest and engagement among the target audience as an incentive for change.

In S2 and S3, 15 items had an I-CVI of 1.00, and 3 had an index of 0.86. For S2, one juror suggested that the title of the video be adjusted, justifying that interest in the topic might decrease if terms included in the name of the video were not familiar. To this end, the title "Food categories in the NOVA classification" was changed to "Food classification by processing levels." Regarding S3, suggestions for adjustments in certain parts of the text were adopted to avoid confusion of terms. Further comments have been inserted in Table 2.

Table 2. Experts' main comments and suggestions regarding general aspects of the scripts.

Script number	Comments and suggestions
Script 1	• When the script addresses ultraprocessing: ”I suggest valuing that processing is not possible at home, that it alters the fresh food so much that it is not possible to recognize it.“ (Judge 3) • ”I suggest including some activity proposals to stimulate reflection, for example: ‘Try to classify the food you have at home/How do you evaluate your diet after this check?’“ (Judge 6) • ”I found that some texts were a little long, but I don't see any major problems since it will be narrated.“ (Judge 7)
Script 2	• ”The recommendations of the guide were cited in all groups but the ultraprocessed ones: 'Avoid ultraprocessed foods.’ Include to standardize and emphasize.” (Similar comments between Judges 2 and 5) • “I suggest inserting a scene explaining the label and including some questions that lead the audience to look for a label and reflect on the food.“ (Judge 6)
Script 3	• ”Perhaps it is important to say, after mentioning ‘Cosmetic Additives,’ something like: 'that serve to beautify food.' It may give the impression that cosmetics are used (foundation, powder, mascara), since this term is better known for this class of products.“ (Judge 1) • ”I think it would be equally important to address the impacts of UPF on the environment and culture.“ (Judge 3) • On the characteristics of ultraprocessed foods: ”I thought saying that 'it allows them to be consumed in a practical way at any time of the year' sounded very positive. I think something like: '(...) allows them to be consumed without having to pay much attention, while walking, driving, watching TV, etc.’ Then, in the illustration, instead of showing all the characters eating ultraprocessed foods on the plate, include people eating cereal bars while driving, instant noodles while on the computer, etc.“ (Judge 4) • ”Insert something that leads the audience to review food and its accessibility.“ (Judge 6)

Based on the validated scripts for the animated videos, the content approach was divided into the following segments:

1. Introduction to the concept of food processing levels before they are purchased, prepared, and consumed and how these levels influence the quality of the food, whether in a positive way, making it more durable or facilitating preparation, or in a negative way, making the food nutritionally unbalanced and high in calories
2. Deepening the categories and food identification according to the NOVA classification system
3. An approach to UPFs and reports of some of their impacts on health

The videos were structured considering multimedia learning instructional principles. Thus, this production complied with the Coherence Principle, using only pictograms, videos, and narration to fulfill the communication purpose. To meet the Signaling Principle, chart arrows, symbols, and colors were used to highlight elements and topics during the videos. The use of animation and narration only, without adding text or subtitles, respected the Redundancy Principle. The Spatial Contiguity Principle and Temporal Contiguity Principle were respected when relating pictograms and texts in the same scene and when narrating and presenting the pictograms simultaneously, respectively.

The Segmentation Principle and Pre-training Principle were also respected, since the distribution of content was divided into topics and, with each video, the theme was deepened, and terms and concepts were reinforced. The Multimedia Principle was met by opting for the use of words and images. The Modality Principle was not fully met, since, despite the pictograms and videos being associated with the narration, at certain times, the use of technical terms was necessary; therefore, it was decided to add words to the scenes. To meet the Personalization Principle, Voice Principle, and Image Principle, simple and informal language was adopted, narrated by a human voice, without the narrator's exposure. The principles used in each animated video can be accessed in Multimedia Appendix 6.

Once the videos were created, the validation process began. This stage involved 3 specialists in multimedia production, aged between 22 years and 52 years, with an average experience of 6.3 (SD 3.2) years. The results referring to this evaluation are presented in Table 3. The SAM, an instrument used by these specialists, evaluates educational materials and aspects related to its adequacy and effectiveness of communication [23].

Table 3. Multimedia production specialists' agreement regarding the adequacy of animated videos (n=3).

Adequacy of material[a]	Animated video 1 E[b] (%)	Animated video 1 S[c] (%)	Animated video 1 %A[d]	Animated video 2 E (%)	Animated video 2 S (%)	Animated video 2 %A	Animated video 3 E (%)	Animated video 3 S (%)	Animated video 3 %A
Content	100	0	1	100	0	1	91.7	8.3	1
Language	80	20	1	93.3	6.7	1	73.3	26.7	1
Illustrations	100	0	1	100	0	1	100	0	1
Layout and presentation	77.8	22.2	1	77.8	22.8	1	66.7	33.3	1
Stimulation and motivation to learn	88.9	11.1	1	77.8	22.8	1	88.9	11.1	1
Cultural adequacy	100	0	1	83.3	16.7	1	83.3	16.7	1
S-CVI/Ave[e]	—[f]	—	1	—	—	1	—	—	1

[a]Adequacy of the material determined using the average of the items distributed in the 6 domains of the instrument.

[b]E: excellent.

[c]S: suitable.

[d]%A: percentage of agreement.

[e]S-CVI/Ave: scale-level content validity index, averaging method.

[f]Not applicable.

As documented in Table 3, the I-CVI and S-CVI/Ave for the 3 animated videos were equal to 1.00. These results attest to the adequacy and quality of the 3 animated videos in communicating the addressed content.

After being evaluated by specialists in multimedia production, the videos underwent adjustments and finalization and, subsequently, were evaluated by adult representatives of the target audience, proceeding to the last stage of validation.

This stage involved 15 Brazilian adults (6 men and 9 women) with an average age of 37.1 (SD 11.2) years who were residents of 2 states and regions of the country. Regarding race, 1 (7%) participant declared him/herself to be Black, 7 (47%) declared themselves to be Brown, and 7 (47%) were White. With regard to education, 1 (7%) participant declared not having completed basic education, 3 (20%) declared having completed primary education, 6 (40%) had completed secondary education, and 5 (33%) graduated. All of them evaluated the material on a smartphone.

The results of the evaluation of the adequacy of the videos by the target audience are presented in Table 4. The 3 animated videos were considered adequate in terms of content, language, illustrations, layout and presentation, stimulation and motivation to learn, and cultural adequacy. All items had an I-CVI of 1.00, and therefore, the S-CVI/Ave was equal to 1.0. Additional information on the results presented in Tables 1, 3, and 4 has been inserted in Multimedia Appendix 7.

Table 4. Assessment of material suitability by representatives of the target audience (n=15).

Adequacy of material[a]	Animated video 1 E[b] (%)	Animated video 1 S[c] (%)	Animated video 1 %A[d]	Animated video 2 E (%)	Animated video 2 S (%)	Animated video 2 %A	Animated video 3 E (%)	Animated video 3 S (%)	Animated video 3 %A
Content	96.7	3.3	1	98.3	1.7	1	98.3	1.7	1
Language	94.7	5.3	1	96	4	1	94.6	5.4	1
Illustrations	100	0	1	97.8	2.2	1	95.6	4.4	1
Layout and presentation	91.1	8.9	1	97.7	2.3	1	97.8	2.2	1
Stimulation and motivation to learn	97.8	2.2	1	93.3	6.7	1	97.8	2.2	1
Cultural adequacy	100	0	1	100	0	1	100	0	1
S-CVI/Ave[e]	—[f]	—	1	—	—	1	—	—	1

[a]Adequacy of the material as determined using the average of the items distributed in the 6 domains of the instrument.

[b]E: excellent.

[c]S: suitable.

[d]%A: percentage of agreement.

[e]S-CVI/Ave: scale-level content validity index, averaging method.

[f]Not applicable.

This evaluation stage showed that the videos met their purpose of communicating the subject and that they were suitable for what they set out to do, as described in the following participant quotes:

> *The material was very well prepared.* [Participant 14]
>
> *I thought it was very well explained and done.* [Participant 4]
>
> *(...) I learned a lot, a lot of things I did not imagine (...) I thought the idea was fantastic.* [Participant 11]

In addition to giving their opinion on the material itself, participants were asked to comment on aspects that could improve the strategy for communicating the content or improve the final quality of the material:

> *Addressing some diseases triggered by the frequent use of these ultraprocessed foods, but in the form of curiosity.* [Participant 5 about animated video 2]
>
> *Extra material explaining where xanthan gum and other ingredients used to modify flavor, texture, etc. come from.* [Participant 9 about animated video 3]
>
> *Make more videos.* [Participant 1 about animated video 3]

As far as Participant 5's comment is concerned, this approach is taken in animated video 3, in which the focus is on the characteristics of UPFs and their implications for health. The comments made by Participants 1 and 9, on the other hand, point to a demand that was not covered in the proposal for this work, but which opens up space for the development and dissemination of new educational materials on this subject.

An interesting note in the evaluation of the videos concerns the stimulation or excitement of watching them and reinforces the importance of using this material in contextualized actions.

> *Excitement/stimulation is also relative. The viewer of the material may not be having a good day or may not be interested in the subject.* [Participant 9]

Once this last stage of validation was completed, the animated videos were considered finished.

Discussion

Principal Findings

Based on intersectoral actions, the Promotion of Adequate and Healthy Eating (abbreviated in Portuguese as PAAS) aims to support individuals and communities to adopt appropriate dietary practices for biological, sociocultural, and environmental aspects [30]. Thus, this study presents the production and validation process of a series of videos that aim to contribute to the expansion of these actions by providing adequate educational resources for FNE actions.

The content validation by the specialists assessed the representativeness of the material in adequately addressing what is proposed, as well as its quality, and increases the likelihood of successful and effective understanding of the content addressed [20]. The evaluations obtained in this stage, even with suggestions for improvement of the message approach, considered the content adequate (according to the CVI), meeting the objective of verifying the cohesion and coherence of the developed scripts.

With regard to the transmission of the educational message, a video has a greater ability to capture the user's attention, in addition to being more dynamic than folders and booklets, for example [11]. However, to arouse interest and engagement, the educational video needs an effective design to facilitate the communication of its content and promote the audience's understanding. In this sense, during the material creation, special care was taken with the main terms and titles, as well as the exclusion of unnecessary elements, resulting in videos, according to experts, that were didactic and that served the purpose of communicating about the subject.

An educational resource needs to establish interactivity with its audience; therefore, the evaluation by representatives of the target audience was essential to increase the probability of success and effectiveness during educational actions.

It is noteworthy that the use of videos for educational activities has shown promising results in different areas of health [31-33]. According to Adam and collaborators [34], due to growing investment, the production of video content is a means for disseminating educational health content that will probably remain perennial for several years. Besides, this type of resource can influence a new generation of interventions that are more aligned with the needs and contexts of the target audience.

Meppelink and collaborators [11] concluded that animated visual information combined with narration is the best way to communicate complex health messages to people with low health literacy while also being suitable for people with higher levels of health literacy.

The planning stage of this study was important to analyze key aspects of the content approach related to the NOVA classification and DGBP developed so far [1]. In developing the instructional materials, care was taken to use language that communicated the content in a simple and clear manner to make knowledge acquisition effective. Aware that this effectiveness is related to the structure and way of approaching the content, recommendations from the CTML were taken into account in the development of this series of animated videos, applying principles to minimize the cognitive demands on the target audience in processing the topic covered.

Online communication and the use of multimedia educational resources have gained importance in recent years, especially since the COVID-19 pandemic led to new demands in the field of health promotion and surveillance, changing and expanding the strategies of community and individual communication [35]. Given that aspects of this approach to health persist and tend to expand, this type of educational resource is a promising strategy for FNE.

In this context, this study contributes to Promotion of Adequate and Healthy Eating by providing validated animated videos that address the NOVA classification of foods in simplified language while also contributing to the dissemination of the DGBP message by suggesting contextualization of the use of this material that incorporates other recommendations of this guideline [1,36-38].

To date, we have found no publications on studies in which animated educational videos based on NOVA have been developed and validated for use in FNE, so this study appears to be the first with this intention.

A limitation of this study was the lack of a health literacy assessment of the representatives of the target audience to determine the level of understanding of the content. Nevertheless, they were evaluated by 15 individuals with varying levels of education, and all received excellent ratings for language, illustrations, layout, presentation, cultural appropriateness, and motivation to learn. Another limitation concerns the representativeness of the target audience. Although suggestions from the content specialists were accepted to incorporate and expand examples that represented different regions of the country, the animated videos were evaluated by residents of 2 Brazilian regions. Furthermore, adaptations to suit the country's different cultures and eating habits are encouraged.

Finally, the material produced and validated in this study can be used in analytical studies and clinical trials, as well as in actions in primary health care, in person or via mobile health, through platforms or social media. Suggestions for use in workshops and educational actions can be found in a manual developed by the team, available free online [36]. The videos can also be accessed on YouTube by any user of the platform searching for and interested in the topic.

Conclusion

Animated videos based on the NOVA classification were considered valid in terms of content and suitability for Brazilian adults. It is hoped that they can contribute to the quality of FNE actions, they meet the educational requirements on the subject, and they influence healthy eating practices. We believe that the methodology used in this study to develop these multimedia programs can be explored and used to create new resources related to FNE and in health promotion. It is recommended that additional multimedia resources be developed based on the other chapters and recommendations of the DGBP.

Acknowledgments

We deeply thank all participants of the study. We thank the *Fundação de Amparo à Pesquisa do Estado de Minas Gerais* for the master's scholarship and the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*.

We declare that we did not use generative AI to write the manuscript. However, we did use it at times to spell-check the writing.

Data Availability

The data sets analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

The study was designed by MFGS, LNN, and ES. The content of the animated videos was developed by MFGS, and LNN contributed the final version of the content present in the scripts. MFGS developed the animated videos. MFGS and LNN performed the statistical analyses and interpreted the data. MFGS wrote the manuscript, and LNN and ES contributed to the final draft of the manuscript. All authors approved the final article.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Educational Content Validation Instrument in Health.
[DOCX File , 13 KB - ijmr_v12i1e49092_app1.docx]

Multimedia Appendix 2
Instrument Suitability Assessment of Materials.
[DOCX File , 14 KB - ijmr_v12i1e49092_app2.docx]

Multimedia Appendix 3
Animated video 1.
[MP4 File (MP4 Video), 23834 KB - ijmr_v12i1e49092_app3.mp4]

Multimedia Appendix 4
Animated video 2.
[MP4 File (MP4 Video), 16873 KB - ijmr_v12i1e49092_app4.mp4]

Multimedia Appendix 5
Animated video 3.
[MP4 File (MP4 Video), 23036 KB - ijmr_v12i1e49092_app5.mp4]

Multimedia Appendix 6
Instructional principles used in the animated videos.
[DOCX File , 21 KB - ijmr_v12i1e49092_app6.docx]

Multimedia Appendix 7
Agreement between content and multimedia specialists and representatives of the target audience.
[DOCX File , 23 KB - ijmr_v12i1e49092_app7.docx]

References

1. Dietary Guidelines for the Brazilian Population. Ministry of Health of Brazil. 2015. URL: https://bvsms.saude.gov.br/bvs/publicacoes/dietary_guidelines_brazilian_population.pdf [accessed 2023-02-20]
2. Ministério da Saúde, Secretaria de Vigilância em Saúde, Departamento de Análise em Saúde e Vigilância de Doenças não Transmissíveis. Vigitel Brasil 2019: vigilância de fatores de risco e proteção para doenças crônicas por inquérito telefônico : estimativas sobre frequência e distribuição sociodemográfica de fatores de risco e proteção para doenças crônicas nas capitais dos 26 estados brasileiros e no Distrito Federal em 2019. Ministério da Saúde. 2020. URL: https://bvsms.saude.gov.br/bvs/publicacoes/vigitel_brasil_2019_vigilancia_fatores_risco.pdf [accessed 2021-07-08]
3. Monteiro CA, Cannon G, Levy RB, Moubarac J, Louzada ML, Rauber F, et al. Ultra-processed foods: what they are and how to identify them. Public Health Nutr 2019 Apr;22(5):936-941 [FREE Full text] [doi: 10.1017/S1368980018003762] [Medline: 30744710]
4. Nilson EA, Ferrari G, Louzada MLC, Levy RB, Monteiro CA, Rezende LF. Premature deaths attributable to the consumption of ultraprocessed foods in Brazil. Am J Prev Med 2023 Jan;64(1):129-136. [doi: 10.1016/j.amepre.2022.08.013] [Medline: 36528353]
5. Sgambato MR, Lignani JDB, Pires CA, Ribeiro ECDSA, Domingos TB, Ferreira AA, et al. Inequalities in food acquisition according to the social profiles of the head of households in Brazil. Cien Saude Colet 2022 Nov;27(11):4303-4314 [FREE Full text] [doi: 10.1590/1413-812320222711.01582022] [Medline: 36259850]
6. Nilson E, Andrade R, de Brito DA, de Oliveira ML. Costs attributable to obesity, hypertension, and diabetes in the Unified Health System, Brazil, 2018. Rev Panam Salud Publica 2020;44:e32 [FREE Full text] [doi: 10.26633/RPSP.2020.32] [Medline: 32284708]
7. Ministério do Desenvolvimento Social e Combate à Fome – Secretaria Nacional de Segurança Alimentar e Nutricional. Marco de Referência de Educação Alimentar e Nutricional para as Políticas Pública. CREN. 2012. URL: https://www.cren.org.br/blog/portfolio_page/marco-educacao-alimentar-e-nutricional/ [accessed 2023-02-20]
8. Gabe KT, Tramontt CR, Jaime PC. Implementation of food-based dietary guidelines: conceptual framework and analysis of the Brazilian case. Public Health Nutr 2021 Aug 16;24(18):6521-6533. [doi: 10.1017/s1368980021003475]
9. Sørensen K, Levin-Zamir D, Duong T, Okan O, Brasil V, Nutbeam D. Building health literacy system capacity: a framework for health literate systems. Health Promot Int 2021 Dec 13;36(Supplement_1):i13-i23 [FREE Full text] [doi: 10.1093/heapro/daab153] [Medline: 34897445]
10. Stahl M, Cheung J, Post K, Valin JP, Jacobs I. Accelerating virtual health implementation following the COVID-19 pandemic: questionnaire study. JMIR Form Res 2022 May 16;6(5):e32819 [FREE Full text] [doi: 10.2196/32819] [Medline: 35323115]
11. Meppelink CS, van Weert JCM, Haven CJ, Smit EG. The effectiveness of health animations in audiences with different health literacy levels: an experimental study. J Med Internet Res 2015 Jan 13;17(1):e11 [FREE Full text] [doi: 10.2196/jmir.3979] [Medline: 25586711]
12. França T, Rabello ET, Magnago C. As mídias e as plataformas digitais no campo da Educação Permanente em Saúde: debates e propostas. Saúde debate 2019 Aug;43(spe1):106-115. [doi: 10.1590/0103-11042019s109]
13. Barbosa da Silva L, Mara C, de Melo Tavares M. Tecnologia digital de informação e comunicação como agente otimizador na relação ensino-aprendizagem na saúde. Revista Pró-UniverSUS 2019 Dec 11;10(2):108-111 [FREE Full text] [doi: 10.21727/rpu.v10i2.2068]
14. Winder C, Dowlatabadi Z. In: Miller-Zarneke T, editor. Producing animation, third edition. Boca Raton, FL: CRC Press; 2019.
15. Wright J. Animation Writing and Development: From Script Development to Pitch. Abingdon, United Kingdom: Routledge; 2005.
16. Monteiro CA, Cannon G, Moubarac J, Levy RB, Louzada MLC, Jaime PC. The UN Decade of Nutrition, the NOVA food classification and the trouble with ultra-processing. Public Health Nutr 2018 Jan;21(1):5-17 [FREE Full text] [doi: 10.1017/S1368980017000234] [Medline: 28322183]

17. Mayer RE. Multimedia learning, second edition. Cambridge, MA: Cambridge University Press; 2009.
18. Grant JS, Davis LL. Selection and use of content experts for instrument development. Res. Nurs. Health 1997 Jun;20(3):269-274. [doi: 10.1002/(sici)1098-240x(199706)20:3<269::aid-nur9>3.0.co;2-g]
19. Fehring R. Methods to validate nursing diagnoses. Heart Lung 1987 Nov;16(6 Pt 1):625-629. [Medline: 3679856]
20. Leite SDS, Áfio ACE, Carvalho LVD, Silva JMD, Almeida PCD, Pagliuca LMF. Construction and validation of an Educational Content Validation Instrument in Health. Rev Bras Enferm 2018;71(suppl 4):1635-1641 [FREE Full text] [doi: 10.1590/0034-7167-2017-0648] [Medline: 30088634]
21. Lynn MR. Determination and Quantification Of Content Validity. Nursing Research 1986;35(6):382-386. [doi: 10.1097/00006199-198611000-00017]
22. Doak CC, Doak LG, Root JH. In: Morton PG, editor. Teaching patients with low literacy skills. Philadelphia, PA: Lippincott Williams & Wilkins; 1996:49-59.
23. Sousa CS, Turrini RNT, Poveda VB. Tradução e adaptação do instrumento “Suitability Assessment of Materials" (SAM) para o Português. Revista de Enfermagem 2015;9(5):7854-7861 [FREE Full text]
24. Lopes M, Silva V, Araujo T. Methods for establishing the accuracy of clinical indicators in predicting nursing diagnoses. Int J Nurs Knowl 2012 Oct;23(3):134-139. [doi: 10.1111/j.2047-3095.2012.01213.x] [Medline: 23043652]
25. Polit DF, Beck CT. Fundamentos de Pesquisa em Enfermagem: Avaliação de Evidências para a Prática da Enfermagem. Porto Alegre, Brazil: Artmed Editora; 2011.
26. Questionnaire ConVid: behavior research. Fiocruz. URL: https://convid.fiocruz.br/arquivos/QuestionarioConVidPesquisaDeComportamentos.pdf [accessed 2021-07-08]
27. Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Res Nurs Health 2007 Aug;30(4):459-467. [doi: 10.1002/nur.20199] [Medline: 17654487]
28. Polit DF, Beck CT. The content validity index: are you sure you know what's being reported? Critique and recommendations. Res Nurs Health 2006 Oct;29(5):489-497. [doi: 10.1002/nur.20147] [Medline: 16977646]
29. Resolução No 466, de 12 de dezembro de 2012. Conselho Nacional de Saúde. 2012. URL: https://conselho.saude.gov.br/resolucoes/2012/Reso466.pdf [accessed 2023-08-10]
30. National Food and Nutrition Policy. Ministry of Health of Brazil. 2013. URL: http://189.28.128.100/dab/docs/portaldab/publicacoes/national_food_nutrition_policy.pdf [accessed 2022-10-16]
31. Denny MC, Vahidy F, Vu KYT, Sharrief AZ, Savitz SI. Video-based educational intervention associated with improved stroke literacy, self-efficacy, and patient satisfaction. PLoS One 2017;12(3):e0171952 [FREE Full text] [doi: 10.1371/journal.pone.0171952] [Medline: 28333925]
32. Kim S, Ju MK, Son S, Jun S, Lee SY, Han CS. Development of video-based educational materials for kidney-transplant patients. PLoS One 2020 Aug 3;15(8):e0236750 [FREE Full text] [doi: 10.1371/journal.pone.0236750] [Medline: 32745145]
33. Tuong W, Larsen ER, Armstrong AW. Videos to influence: a systematic review of effectiveness of video-based education in modifying health behaviors. J Behav Med 2014 Apr;37(2):218-233. [doi: 10.1007/s10865-012-9480-7] [Medline: 23188480]
34. Adam M, McMahon SA, Prober C, Bärnighausen T. Human-centered design of video-based health education: an iterative, collaborative, community-based approach. J Med Internet Res 2019 Jan 30;21(1):e12128 [FREE Full text] [doi: 10.2196/12128] [Medline: 30698531]
35. Medina MG, Giovanella L, Bousquat A, de Mendonça MHM, Aquino R, Comitê Gestor da Rede de Pesquisa em Atenção Primária à Saúde da Abrasco. Primary healthcare in times of COVID-19: what to do? Cad Saude Publica 2020;36(8):e00149720 [FREE Full text] [doi: 10.1590/0102-311x00149720] [Medline: 32813791]
36. Silva MFG, Nobre LN, Silva E. Adopting multimedia educational materials in Food and Nutrition Education actions. Diamantina: Universidade Federal dos Vales do Jequitinhonha e Mucuri; 2023.
37. How is your diet? Ministry of Health of Brazil. 2018. URL: http://189.28.128.100/dab/docs/portaldab/publicacoes/guiadebolso_folder.pdf [accessed 2023-08-26]
38. Food safety and quality: Online Edition: "Combined Compendium of Food Additive Specifications". Food and Agriculture Organization of the United Nations. URL: https://www.fao.org/food/food-safety-quality/scientific-advice/jecfa/jecfa-additives/en/ [accessed 2023-02-20]

Abbreviations

CTML: cognitive theory of multimedia learning
CVI: content validity index
DGBP: Dietary Guidelines for the Brazilian Population
ECVIH: Educational Content Validation Instrument in Health
FNE: food and nutrition education
I-CVI: item-level content validity index
LILACS: Literatura Latino-Americana e do Caribe em Ciências da Saúde
NCD: noncommunicable disease

S1: script for animated video 1
S2: script for animated video 2
S3: script for animated video 3
SAM: Suitability Assessment of Materials
S-CVI/Ave: scale-level content validity index, average calculation method
UPF: ultraprocessed foods

Edited by T de Azevedo Cardoso; submitted 17.05.23; peer-reviewed by N Noreen, A Elnaggar; comments to author 29.07.23; revised version received 18.08.23; accepted 22.08.23; published 11.09.23.

Please cite as:
Silva MFGD, Nobre LN, Silva ED
Animated Videos Based on Food Processing for Guidance of Brazilian Adults: Validation Study
Interact J Med Res 2023;12:e49092
URL: https://www.i-jmr.org/2023/1/e49092
doi:10.2196/49092
PMID:37695656

Original Paper

Identifying Targets for Innovation in Amazon Reviews of Bedwetting Alarms: Thematic Analysis

Astha Sahoo[1], BSc; Savannah Leah Starr[1], MD; Vadim Osadchiy[1,2], MD; Sophia Desai[3], BSc; Neha Iyer[3], BSc; Marie Luff[1], BS; Grace E Sollender[2], MD; Renea Sturm[1,2,4], MD

[1]David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, United States

[2]Department of Urology, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, United States

[3]Department of Bioengineering, Samueli School of Engineering, University of California, Los Angeles, Los Angeles, CA, United States

[4]University of California, Los Angeles Mattel Children's Hospital, Los Angeles, CA, United States

Corresponding Author:
Renea Sturm, MD
University of California, Los Angeles Mattel Children's Hospital
757 Westwood Plaza
Los Angeles, CA, 90095
United States
Phone: 1 (310) 825 9111
Email: rsturm@mednet.ucla.edu

Abstract

Background: Nocturnal enuresis (NE) is a frequent diagnosis in pediatric and adolescent populations with an estimated prevalence of around 15% at the age of 6 years. NE can have a substantial impact on multiple health domains. Bedwetting alarms, which typically consist of a sensor and moisture-activated alarm, are a common treatment.

Objective: This study aimed to determine areas of satisfaction versus dissatisfaction reported by the parents and caregivers of children using current bedwetting alarms.

Methods: Using the search term "bedwetting alarms" on the Amazon marketplace, products with >300 reviews were included. For each product, the 5 reviews ranked the "most helpful" for each star category were selected for analysis. Meaning extraction method was applied to identify major themes and subthemes. A percent skew was calculated by summing the total number of mentions of each subtheme,+1 for a positive mention, 0 for a neutral mention, and –1 for a negative mention, and dividing this total by the number of reviews in which that particular subtheme was observed. Subanalyses were performed for age and gender.

Results: Of 136 products identified, 10 were evaluated based on the selection criteria. The main themes identified across products were long-term concerns, marketing, alarm systems, and device mechanics and features. The subthemes identified as future targets for innovation included alarm accuracy, volume variability, durability, user-friendliness, and adaptability to girls. In general, durability, alarm accuracy, and comfort were the most negatively skewed subthemes (with a negative skew of –23.6%, –20.0%, and –12.4% respectively), which are indicative of potential areas for improvement. Effectiveness was the only substantially positively skewed subtheme (16.8%). Alarm sound and device features were positively skewed for older children, whereas ease of use had a negative skew for younger children. Girls and their caretakers reported negative experiences with devices that featured cords, arm bands, and sensor pads.

Conclusions: This analysis provides an innovation roadmap for future device design to improve patient and caregiver satisfaction and compliance with bedwetting alarms. Our results highlight the need for additional options in alarm sound features, as children of different ages have divergent preferences in this domain. Additionally, girls and their parents and caretakers provided more negative overall reviews regarding the range of current device features compared to boys, indicating a potential focus area for future development. The percent skew showed that subthemes were often more negatively skewed toward girls, with the ease of use being –10.7% skewed for boys versus –20.5% for girls, and comfort being –7.1% skewed for boys versus –29.4% for girls. Put together, this review highlights multiple device features that are targets for innovation to ensure translational efficacy regardless of age, gender, or specific family needs.

(Interact J Med Res 2023;12:e43194) doi:10.2196/43194

KEYWORDS

Amazon; bedwetting alarms; nocturnal enuresis; online reviews; early customer discovery; online; diagnosis; pediatric; teen; adolescent; sensor; treatment; age; sex; device; user; adaptability; efficacy; child health; bedwet; enuresis; Urology; pediatric urology; alarm; alert; notification; sleeping practice; sleep practice; sleep disorder; polysomnography; thematic analysis; natural language processing; NLP

Introduction

Nocturnal enuresis (NE) is defined as nighttime urinary incontinence in adults and children older than 5 years [1,2]. It is a common condition, affecting nearly 10% of school-age children [3]. NE can have negative consequences on pediatric mental health and academic performance [3]. Furthermore, having a child with NE can also have a negative impact on parental anxiety and depression [4].

Bedwetting alarms are the most common first-line treatment for NE. These alarms function by using moisture sensors, which provide audio or vibrating signals when detecting liquid, thereby conditioning the child to awaken when they need to urinate. Despite bedwetting alarms being the most common treatment for NE, they have several limitations. In a systematic meta-analysis comparing bedwetting alarms to medication-based therapy (desmopressin), bedwetting alarms had a higher likelihood of successfully treating NE with a lower relapse rate and better sustained response rate if patients completed the respective therapies [5]. However, nearly half of the participants treated with bedwetting alarms discontinued their use before the prescribed completion of treatment. Common reasons cited for early discontinuation included alarm discomfort, failure to awaken from sleep, the lack of efficacy, and false alarms [5]. Furthermore, an intention-to-treat analysis demonstrated that alarms did not outperform medication in achieving successful treatment [5]. This finding highlights the critical need to determine which factors limit parents and children from successfully continuing and ultimately completing treatment when using bedwetting alarms. Currently, there is insufficient data regarding product design needs that meet the real-world requirements of parents and children with NE to inform device innovation.

By analyzing consumer perspectives regarding commonly used over-the-counter health care products, top priorities for future prototypes can be identified, highlighting the importance of early customer discovery in health care innovation. "Customer discovery" is a common method used by product developers to determine whether actual customers for a product exist and what those customers desire before product development [6]; however, this method is not commonly applied to health care [6]. With the growing use of the internet to review many aspects of health care, from products to providers, this information presents a unique data source that remains underexplored. In this study, we applied early customer discovery to expand our understanding of the weaknesses and strengths of bedwetting alarms.

Using websites and social media for health care is an emerging field due to its efficiency in the advertisement and collection of data. Although internet sources such as Amazon Mechanical Turk, a crowdsourcing resource to distribute surveys and collect data [7], have been used for clinical research, there are many additional web-based sources of user-derived data that are relatively untapped in health care. For example, a recent study aimed to better understand the experiences of those struggling with male infertility by evaluating the popular discussion platform Reddit, analyzing responses on the topic using data analytics tools; the information gleaned has been applied to educate health care practitioners on the current concerns of the male infertility community [8]. Specific to consumer-marketed health technology, web-based reviews and social media platforms foster competition that has diversified the range of products and product features available by catering to different user demographics beyond those commonly participating in clinical research trials [9].

This study arose from a recent partnership between a medical student innovation program and a capstone engineering course at a research university. First, a stepwise biomedical innovation needs–mapping process was performed as part of a formal medical innovations program, Sling Health [10]. During this program, perspectives were obtained from various stakeholders, including patients and their families as well as health care staff through clinical observation [10]. The need for more efficient treatments for NE was identified as the top unmet need in pediatric urology [10]. Based on this finding, a team with backgrounds in bioengineering and medicine was assembled to create a prototype to meet this need. Further analysis of the strengths and deficits in current products was required to expand our understanding of needs across a range of settings, parent-child dynamics, and characteristics.

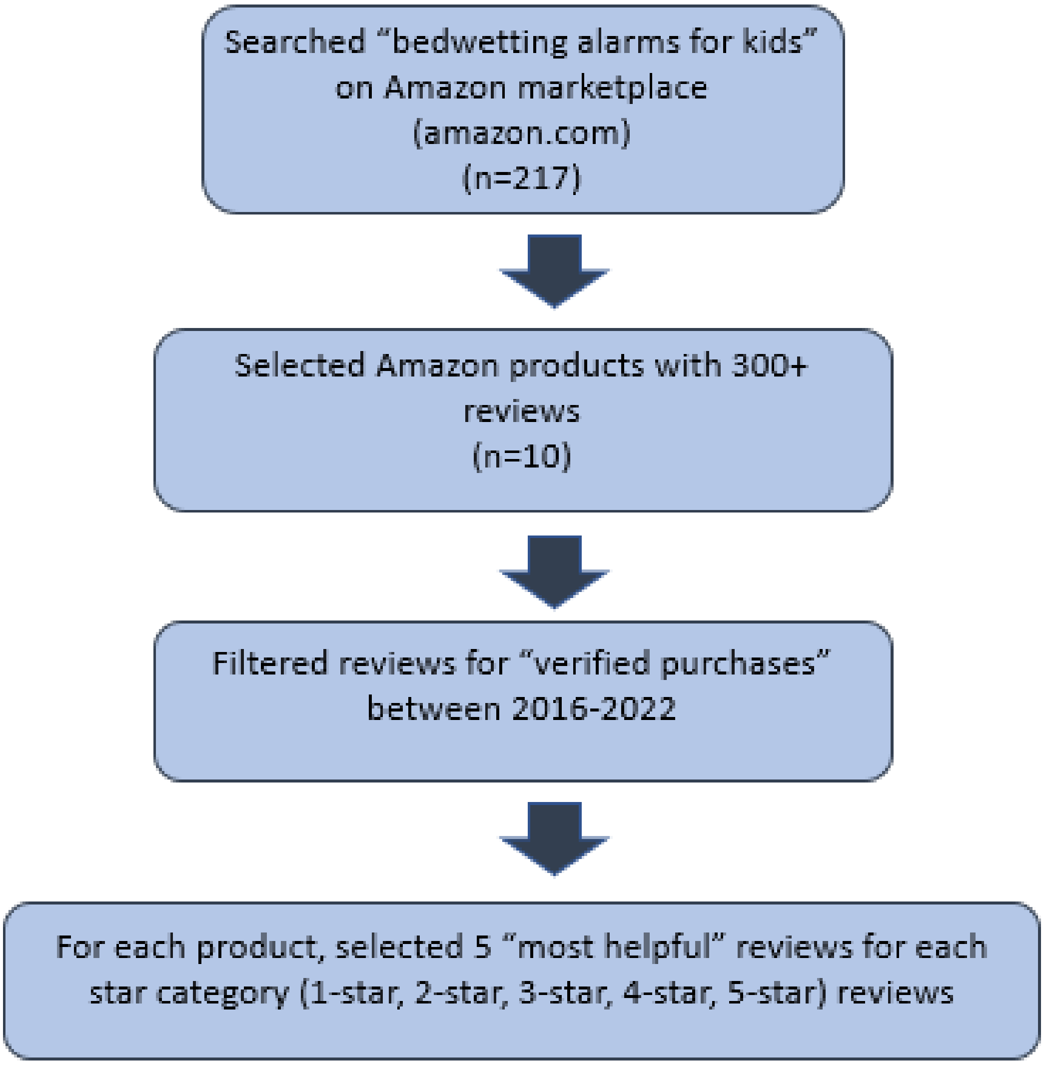
The primary aim of this study was to elucidate areas of satisfaction and dissatisfaction with the current alarm-based treatment of NE through a systematic evaluation of consumer reviews of bedwetting products, with a focus on defining specific targets for future innovation. Secondary analyses by gender and age were also performed to determine whether reported experiences differed between specific user groups.

Methods

Data Extraction

Using the search term "bedwetting alarms for kids," Amazon reviews from 2016 to 2022 were extracted from the Amazon marketplace [11] for all bedwetting products that had over 300 ratings in February and March 2022. Data was collected manually by 4 evaluators. Reviews were extracted if they met three criteria: (1) the review was posted within the target date range; (2) the review length was >5 words; and (3) Amazon verified that the reviewer had purchased the product. Amazon user product ratings occur on a 1- to 5-star scale; thus, reviews are categorized into 5 "star categories": 1-star, 2-star, 3-star, 4-star, or 5-star reviews. To further target a range of experiences, up to 5 reviews from each star category, for a maximum of 25 reviews per product, were analyzed. In the Amazon marketplace, users can mark a post as "helpful" if the review helped them decide whether to purchase the product or if the review helped them fix a problem they were having with the product [12]. To collect data that were in accordance with most users, the 5 reviews with the greatest number of "helpful" marks were collected for each star category (Figure 1). For each review that met these criteria, the following data were analyzed further: the date of publication, review title, and review content.

Figure 1. Method for extracting Amazon reviews of bedwetting alarms.



Ethical Considerations

All data were publicly available and did not involve human subjects. Therefore, this study was exempt from review by the Institutional Review Board of the University of California, Los Angeles.

Qualitative Thematic Analysis

A free-text and qualitative thematic analysis was completed through manual annotation of selected reviews. Meaning extraction method, a technique that forms simple overarching themes from text by extracting content words (ie, nouns, verbs, and adjectives) and removing connective words (ie, articles, prepositions, and pronouns), was applied to identify major themes and subthemes for each review [13]. Themes that related to unique categories of strength and deficits faced by users of bedwetting alarms were identified, and subthemes defined these further. A minimum of 2 investigators, including the research mentor, reviewed posts during data collection to identify a preliminary set of common themes. These themes were then discussed in depth among all 4 investigators to achieve consensus. The team of investigators then formed subthemes during the process of evaluation. After an iterative process evaluating all reviews, themes and subthemes were discussed and finalized. For each analyzed review, investigators rated the discussion of each subtheme on a 3-point scale, where –1 indicated that the reviewer had a negative experience regarding that subtheme, 0 indicated that the experience was neutral or not discussed, and 1 indicated that the reviewer had a positive experience with regard to that subtheme. After discussion, a minimum of 2 reviewers agreed upon each rating. All subthemes were ranked from most negative (weakness) to most positive (strength) in current NE products using the percent skew, which was calculated by taking the sum of all investigator ratings and dividing by the total number of reviews for each product.

In addition to major themes, relevant information, such as price and demographics of the child for whom the device was purchased, and the follow-up review were collected. Specifically, the gender and age of the child using the device were included in data collection and analysis if provided in the review. Subanalyses by gender (divided into boy and girl cohorts) and age (divided into 2 cohorts aged 3-8 and 9-12 years) for those reviews that provided this information were completed using percent skew. Of note, there were no ages mentioned outside the age ranges in these cohorts.

Consideration of Researcher Characteristics, Reflexivity, and Mitigation of Biases

Four investigators, with at least 2 investigators evaluating each review, were involved in the Amazon review evaluation process. One reviewer was a third-year medical student, one reviewer was a second-year neuroscience undergraduate, and 2 reviewers were bioengineering undergraduate students. The inherent potential for bias between bioengineering students was recognized. To prevent this bias and provide varied perspectives in the analysis of each review, the bioengineering students were paired with either the medical or neuroscience student whose clinical experience allowed the completion of the initial thematic mapping. When discrepancies arose, consensus was reached through a wider complete team review with the research mentor, who is a pediatric urologist.

Results

Of the 136 bedwetting alarm products identified in total, 10 were selected for further analysis based on the provided selection criteria. Products had an average of 1693 (range 329-4380) posted reviews with a mean overall rating of 3.9 (range 3.2-4.5) out of 5. The median price of products was US $68.5 (range US $33.99-US $299). A total of 250 reviews (25 for each product as described previously) were analyzed.

The 4 most frequently identified themes were long-term concerns, marketing, alarm characteristics, and device mechanics and features. Within these overarching themes, distinct subthemes emerged. For long-term concerns, reviewers discussed overall effectiveness, proper use, and durability. In terms of marketing, parents remarked on the cost-to-value ratio and customer service support available for the product. Alarm features and accuracy were common subthemes described regarding alarm characteristics. For device mechanics and features (excluding the alarm or sound), parents commented on specific device features, comfort, safety, ease of use, and effectiveness of reinforcement tools such as reward charts that accompanied certain bedwetting alarms (Table 1).

The total number of mentions for each subtheme was evaluated. The most-mentioned theme was long-term concerns (468 total mentions), and the most-mentioned subtheme was effectiveness (164 [mentions]/250 [total reviews], 65.6%). Marketing was the least-mentioned theme (78 total mentions), and proper use was the least-mentioned subtheme (6/250, 2.4%; Table 2). Representative quotes for each theme and subtheme are illustrated in Table 2. Although most subthemes had positive and negative mentions, there were 2 subthemes, safety and durability, that only had negative mentions (6/250, 2.4% and 67/250, 26.8%, respectively). Therefore, for safety and durability, no positive representative quotes were available (Table 2).

Figure 2 demonstrates the overall skew of each identified subtheme. The percent skew was calculated by summing the total number of mentions of each subtheme, +1 for a positive mention and –1 for a negative mention, and dividing this sum by the total number of reviews. Device subthemes that were positively skewed included effectiveness (42 [total composite score]/250 [total reviews], 16.8%) and the ability to customize a device (4/250, 1.6%). The most negatively skewed subtheme was the durability of the alarm (–59/250, –23.6%), followed by alarm accuracy (–50/250, –20%), comfort (–31/250, –12.4%), and alarm sound features (–31/250, –12.4%), thereby indicating that reviewers more commonly had negative experiences with these features.

Table 1. Identified themes and subthemes through thematic analysis of Amazon reviews of bedwetting alarms.

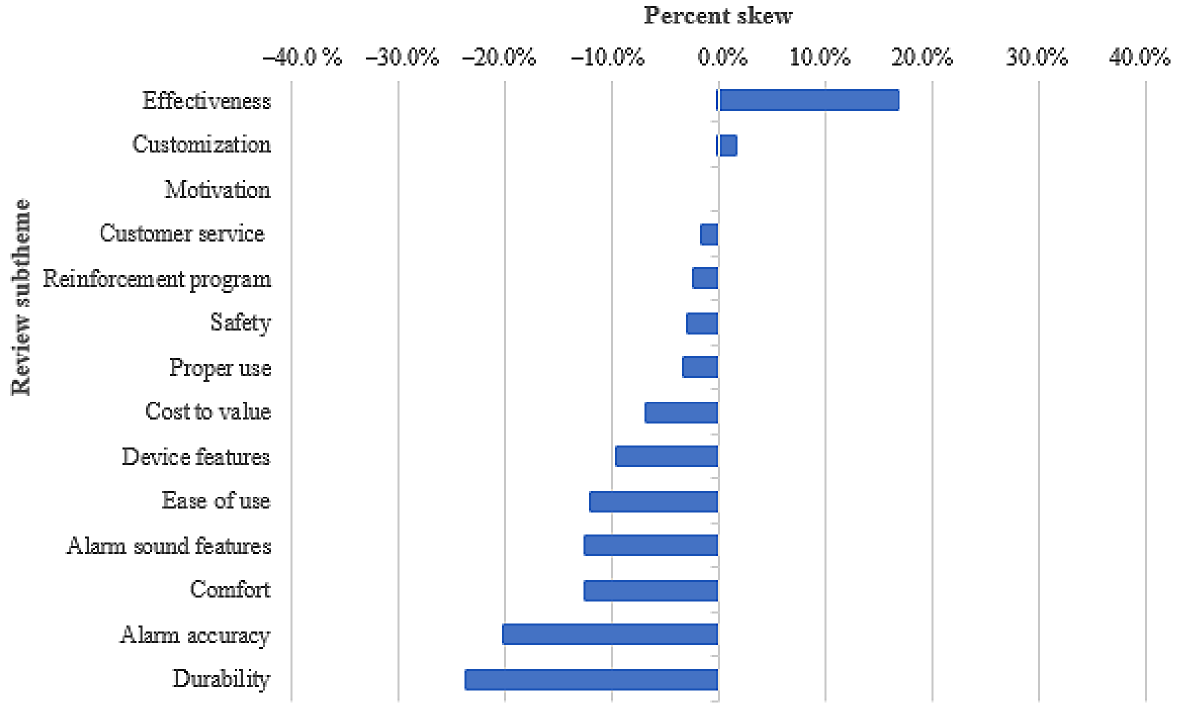
Theme	Subtheme
Long-term concerns	• Proper use • Durability • Effectiveness
Marketing	• Cost to value • Customer service
Alarm characteristics	• Alarm accuracy • Alarm sound features
Device mechanics and features	• Device features • Safety • Comfort • Reinforcement • Ease of use

Table 2. All identified themes and subthemes with the number of mentions and representative quotes.

Theme and subtheme	Mention (N=250 reviews), n (%)	Positive quote	Negative quote
Long-term concerns			
Effectiveness	164 (65.6)	"We are 10 weeks into the program and he has not had any accidents for 7 weeks! We also sit down and watch the program and check in videos together. The star system was also very motivating for our son. I would give this program 5 stars for effectiveness. It is a process but again, for our family, totally worth it for the quick results!"	"We have tried this alarm for 3 years now and it does nothing for our son. I am exhausted at trying to find things to stop his night time enuresis. I thought this would be the game changer, but the only thing it did was wake us up"
Durability	67 (26.8)	No positive quotes available	"Worked for about a month and all alarms stopped working. Ordered another probe because probes are first to die. Changed batteries also. No alarm (vibrate/audible) with new probe and new batteries"
Proper use	6 (2.4)	"For any review that says their child won't wake up with the alarm and they won't get out of bed, you have to sleep in their room to make sure they get up. That's half the battle. If the alarm isn't working, it's most likely user error. If you do it right, we are proof that it'll work"	"The first night, I realized the cord wasn't in the alarm properly (it requires an extra step where you turn it 90 degrees). The 2nd night, we used vibration only and my son doesn't think the alarm went off even though he wet (or maybe it did then stopped without waking him? no idea). The 3rd night, we tried both music and vibration, and again it either didn't go off or didn't wake him when he wet. Trying again tonight with it on full volume."
Effectiveness	164 (65.6)	"This literally was worth every penny"	"We paid several hundred dollars for this product and all we got for our money was months of sleepless nights and a shorted out non working pad"
Durability	67 (26.8)	"If you want to help your child with nighttime dryness, purchase this product and you'll get a team who also cares about your child to help you."	"I tried calling the number for the company several times but only got voicemail, and messages left on it weren't replied to. I'm buying a similar device from a different company."
Marketing			
Cost to value	64 (25.6)	"I like the fact that it has multiple sounds. That way if your child has become familiar with a sound you can change it. The volume is great and I can hear it in my room across the hall."	"Shocked at how loud this alarm is, with no volume control. It woke up the entire house and my 9-year-old was hysterical!"
Customer service	14 (5.6)	"We attached it to the front of his underwear and just the slightest drop of liquid sets it off. It's loud and won't quit till he unclips it which is good."	"Tried it and it does not go off. On occasion, it would go off, but way too late."
Alarm characteristics			
Alarm sound features	88 (35.2)	"He liked this one since it was a mat under the sheet and not something attached to his clothing."	"My 8 year old wore this for two nights, but refused to wear it anymore. Says it's not comfortable and itchy."
Alarm accuracy	79 (31.6)	"We really like that the sensor and alarm are separate pieces and that the piece that wakes you up is not an arm band."	"We had this product for 30 days before it broke. The wiring pulled out from the alarm part."
Device mechanics and features			
Comfort	39 (15.6)	No positive quotes available	"In the middle of the night my son screamed loudly, we immediately ran to his room and disconnected the alarm. He told us that the alarm was giving him shocks and noticed that his entire area was red. We applied ice packs on him with little to no results we had to visit our doctor in the morning."
Device features	30 (12)	"Pros...It comes with stickers and log book."	"…[T]he reward stickers are insensitive and humiliating. Stickers of a baby in diapers!!! How insulting! And if they pee the bed, they're supposed to use a sticker of a crying baby in a diaper."

Theme and sub-theme	Mention (N=250 reviews), n (%)	Positive quote	Negative quote
Safety	9 (3.6)	No positive quotes available	"Worked for about a month and all alarms stopped working. Ordered another probe because probes are first to die. Changed batteries also. No alarm (vibrate/audible) with new probe and new batteries"
Reinforcement	7 (2.8)	"For any review that says their child won't wake up with the alarm and they won't get out of bed, you have to sleep in their room to make sure they get up. That's half the battle. If the alarm isn't working, it's most likely user error. If you do it right, we are proof that it'll work"	"The first night, I realized the cord wasn't in the alarm properly (it requires an extra step where you turn it 90 degrees). The 2nd night, we used vibration only and my son doesn't think the alarm went off even though he wet (or maybe it did then stopped without waking him? no idea). The 3rd night, we tried both music and vibration, and again it either didn't go off or didn't wake him when he wet. Trying again tonight with it on full volume."

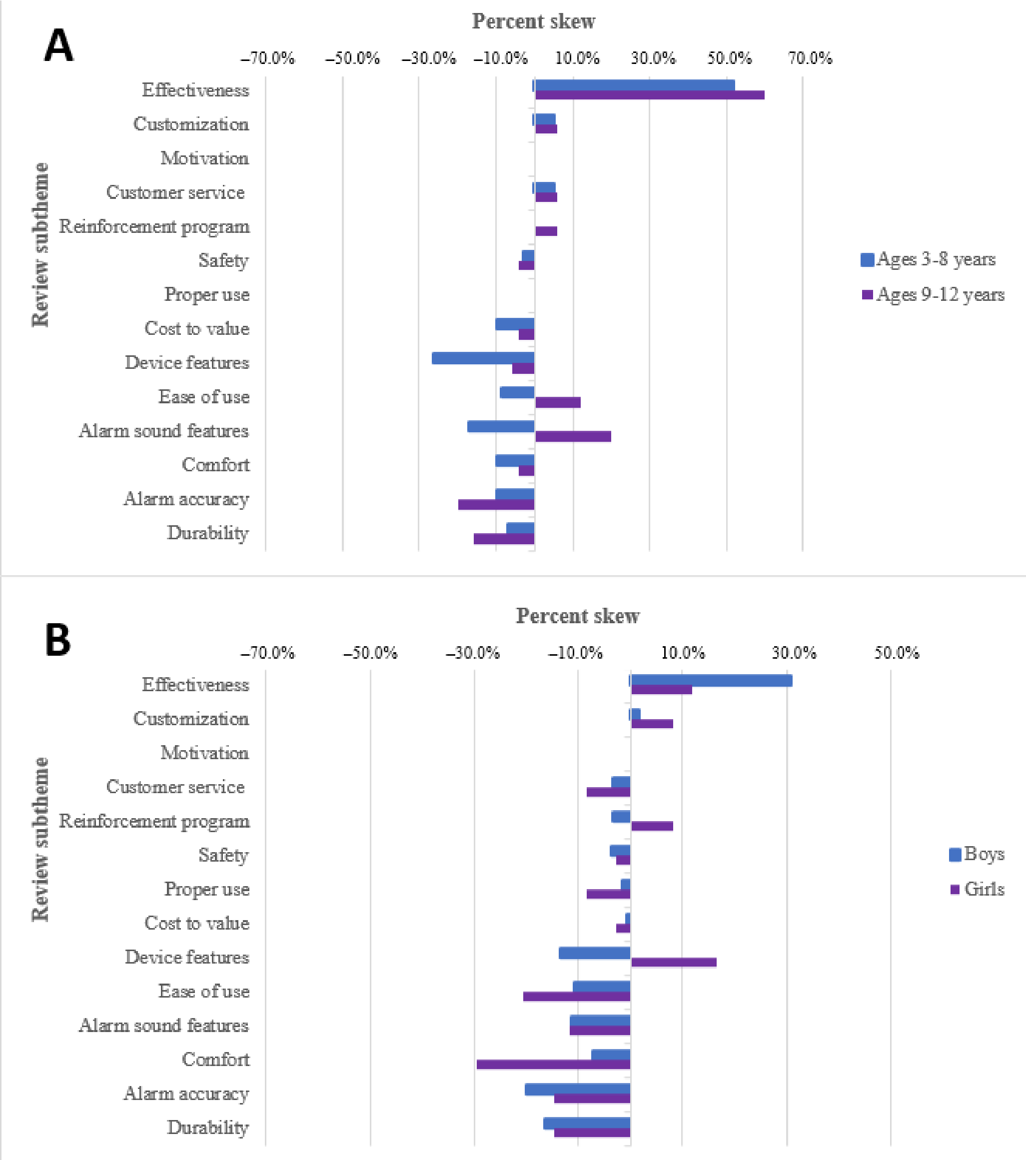
Figure 2. Percent skew of subthemes identified in Amazon reviews of bedwetting alarms.



Additional data analysis revealed 2 specific subpopulations with more negative experiences compared with the overall reviews of current products: younger age and identification as a girl. Of the 250 total reviews, 95 mentioned the age of the child (70 with children aged 3-8 years and 25 with children aged 9-12 years). After analyzing for age, the subtheme alarm sound features had a positive skew for older children (5/25, 20%) compared to a negative skew for younger children (–12/70, –16.9%). Additionally, the subtheme ease of use had a negative skew for younger children (–670, –8.5%) compared with a positive skew for older children (5/25, 20%). Finally, the subtheme device features more negatively affected younger children (–18/70, –26.3%) than older children (–1/25, –4%; Figure 3A). A total of 175 reviews mentioned the gender of the child (141 boys and 34 girls). In the gender subanalysis, specific subthemes emerged as more negatively skewed for girls compared with boys. Specifically, effectiveness (43/141, 30.7% for boys and 4/34, 11.8% for girls), ease of use (15/141, –10.7% for boys and –7/34, –20.5% for girls), and comfort (–10/141, –7.1% for boys and –10/34, –29.4% for girls) demonstrated differences between boys and girls (Figure 3B).

The evaluation of consumer reviews also revealed different opinions stratified by gender regarding specific device features, including alarms with long cords, wireless alarms with a sensor in the underwear, alarms with arm bands, and sensor pads. In the evaluation of these specific device features by gender, alarms with long cords (–1/4, –25% for girls and –1/10, –10.0% for boys), arm bands (–1/3, –33% for girls and 1/30, 3.3% for boys), and sensor pads (–1/12, –8.3% for girls and 0/15, 0% for boys) negatively impacted girls more than boys. Perceptions of wireless alarms had no skew for either gender (0/17, 0% for girls and 1/63, 1.6% for boys; Multimedia Appendix 1).

Figure 3. Percent skew subanalysis by age and gender: (A) ages 3-8 years versus 9-12 years and (B) boys versus girls.



Discussion

Principal Findings

To improve satisfaction and compliance with bedwetting alarms, our study evaluated consumer perceptions of currently available products. The evaluation of bedwetting alarm reviews on Amazon provided the identification of specific targets that need to be addressed. Our study was developed following a recent qualitative study performed by our group. Through 6 weeks of clinical immersion, which involved both clinical observation and interview-based insight extraction across all members of the treatment and patient care team in the pediatric urology clinic, we identified challenges with current bedwetting alarms as the top unmet need [10]. Consistent with our single-center needs mapping, the cause of bedwetting alarm inefficacy and dissatisfaction across a range of products was dissatisfaction with and challenges presented by certain device features as identified in Amazon reviews. From this evaluation of consumer experiences, we have identified 3 prominent targets for innovation: the availability of a wider range of alarm sounds, including more volume and tone options; improved user-friendliness of design; and the incorporation of preferred features across genders. Findings from this study add granularity to the challenges experienced across current solutions. In doing so, our findings provide innovators a roadmap for developing subsequent prototypes for clinical testing. Likewise, our findings will inform clinicians of potential challenges to address with families when prescribing bedwetting alarms to maximize device acceptance and use and will provide children and their families an overview of key features that may impact effective use.

Experience with alarm sound features was negatively skewed in our study, suggesting that more diverse sound options are needed for alarms to accommodate different children and family needs while minimizing alarm fatigue. In many reviews for one product in particular, there was a mixture of complaints regarding volume. For example, one reviewer stated "[the] alarm does not even remotely stir my son," whereas another reviewer for the same product complained that the alarm was "obnoxiously and blaringly loud." Reviewers frequently expressed a desire for more options, including different volume options, vibration options, and various song and nonsong alarm tones. Although we found that parents are seeking more sound options, innovators in this space should also account for the evidence that particular sounds may be more effective at awakening children [14]. A recent study that evaluated sound options for residential fire alarms found that low-frequency tones, the voice of a female stranger, and the voice of the child's mother were significantly more effective at waking children than high-frequency tones [14]. This is consistent with the findings of a systemic review on the effectiveness of different alarm types on sleep inertia [15]. Based on our findings and prior evidence, bedwetting alarms require more alarm tone and volume options, and it is recommended that device developers ensure that low-frequency and voice tones are incorporated to maximize their effectiveness.

Negative experiences with comfort and the ease of use were also identified, indicating a need for improved user-friendliness for bedwetting alarms. This has been previously described as being essential for successful treatment [5,16]. Although certain complex device features may have been added primarily to drive sales volume, such features may in fact impede routine use and may increase the likelihood of early discontinuation, thereby limiting clinical effectiveness [16]. Specific device features that were discussed in reviews included the negative impact of devices with multiple versus fewer components. In particular, products that required multiple steps to turn off and set the alarm received less favorable reviews. Based on these findings, it is recommended that the development of easy-to-use, essential options be a primary focus while keeping the number of separate pieces of equipment to a minimum.

Although more diverse sound options for alarms and improved user-friendliness were themes that have been previously discussed in the literature [5,16], we describe the importance of the incorporation of girl-friendly features, which was frequently noted as a unique target area. Across all bedwetting devices analyzed, girls specifically had more negative skew in many of the major themes identified than boys, particularly with the ease of use and comfort. One reviewer stated, "directions are geared toward boys, so we had to improvise the placement of the sensor," and another reviewer for a different product stated, "[the product] really needs a redesign for usability by young children, especially girls." In multiple reviews, parents stated that the placement of the sensor was especially geared toward boys, and "large, stiff" cords and alarms between the legs of young girls made the use of the product uncomfortable. From our study, we demonstrated that girls were more negatively impacted by device features such as long cords, arm bands, and sensor pads. A potential indirect cause of this variant experience between genders may be the lack of equal gender representation in prior clinical studies of bedwetting alarms. In a recent systemic review on alarm use for the treatment of NE, of the 5026 participants in the 74 studies evaluated, only 33% of participants were girls [16]. Moreover, little is known about gender variance in prototype testing. The lack of female representation in these studies and their inclusion of predominantly male patients could be a potential contributor to the perceived lack of preferred features by girls currently present in bedwetting alarms. These findings highlight the importance of including equal gender representation in future prototype testing and clinical studies. By ensuring the inclusion of girls in future studies and prototype testing from our engineering team, girl-friendly features will be readily incorporated into developing bedwetting alarms.

Through the process of early customer discovery used in this study, we have identified key targets for improvement in future bedwetting alarm prototypes. Analyzing Amazon reviews enabled the research team to assemble a wide range of patient opinions to drive innovation. In the past, Amazon reviews have also been applied in health care to understand marketing myths [17] and patient experiences of uncommonly used treatment modalities [18] and to evaluate product efficacy and safety [19]. However, to our knowledge, this is the first such study for bedwetting alarms. This underutilized data source can enrich our understanding of health care by improving our understanding of the lived patient experience associated with the home use of commercially available health care products.

Limitations

Limitations of this study include the use of a limited data set, potential for implicit bias, and product selection. First, we were limited by solely analyzing the reviews of customers who voluntarily wrote a review. These reviews may have been skewed by those who had a more negative experience, since individuals are more likely to highlight and express their negative views on an experience rather than their positive views [20]. Furthermore, implicit bias could have affected the reviewer team, who were aware of the star rating of the product when extracting language for analysis. However, regardless of the rating, reviewers often had negative and positive perceptions about different aspects of the product. By evaluating each product feature independently, bias and subjectivity from the star rating were mitigated. Finally, analyses were limited to the device and patient features that authors chose to include in each review, and a limited number of products were analyzed. Out of 136 products, 10 common products were reviewed, representing 7% of bedwetting alarm products on the Amazon marketplace. The selection criteria for these devices could have contributed to selection bias for those most commonly purchased. Furthermore, our data set was limited by only selecting bedwetting alarm products on Amazon rather than alternative marketplaces. Put together, this study represents products with a range of common features. Future studies are needed to include a wider range of reviews and devices, potentially aided by automated analysis such as natural language processing and targeted feature analysis to further inform future device development.

Conclusions

This study analyzed Amazon reviews, both validating and expanding upon our qualitative findings in the clinical setting regarding an unmet need for feature innovation in bedwetting alarm systems. To improve satisfaction and compliance with bedwetting alarms, our results provide an innovator roadmap for subsequent device innovation. Our results suggest that more variance in alarm sound features would be beneficial to children and families to improve device compliance, as children of different ages have different preferences in terms of volume and the type of alarm. Less complexity in devices is suggested, as added device features such as arm bands and lengthy cords resulted in devices that limited consistent, effective, and prolonged use. Additionally, current device features may impact girls more negatively overall than boys, indicating that devices were not specifically optimized across genders. Increased representation of girls in device testing is needed. Subthemes such as durability and comfort further suggest that the optimal device should focus on child-friendly features, allowing for wear and tear while ensuring patient comfort. Future prototype evaluation should prioritize these identified device features to optimize translatability and clinical effectiveness. In addition to pinpointing areas necessary for optimization and improvement, this study has highlighted the value of marketplace reviews and the information they can provide to researchers analyzing clinical products developed for home use.

Data Availability

The data sets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Percent skew analysis by gender for specific device features of bedwetting alarms identified in Amazon reviews.

[DOCX File , 41 KB - ijmr_v12i1e43194_app1.docx]

References

1. Bakhtiar K, Pournia Y, Ebrahimzadeh F, Farhadi A, Shafizadeh F, Hosseinabadi R. Prevalence of nocturnal enuresis and its associated factors in primary school and preschool children of khorramabad in 2013. Int J Pediatr 2014;2014:120686-120687 [FREE Full text] [doi: 10.1155/2014/120686] [Medline: 25374608]
2. Welcome to My GOGO Band. GOGO Band. URL: https://mygogoband.com/ [accessed 2022-08-30]
3. Sarici H, Telli O, Ozgur BC, Demirbas A, Ozgur S, Karagoz MA. Prevalence of nocturnal enuresis and its influence on quality of life in school-aged children. J Pediatr Urol 2016 Jun;12(3):159.e1-159.e6. [doi: 10.1016/j.jpurol.2015.11.011] [Medline: 26778419]
4. Meydan EA, Civilibal M, Elevli M, Duru NS, Civilibal N. The quality of life of mothers of children with monosymptomatic enuresis nocturna. Int Urol Nephrol 2012 Jun;44(3):655-659. [doi: 10.1007/s11255-011-0087-7] [Medline: 22081411]
5. Peng CCH, Yang SSD, Austin PF, Chang SJ. Systematic review and meta-analysis of alarm versus desmopressin therapy for pediatric monosymptomatic enuresis. Sci Rep 2018 Nov 13;8(1):16755 [FREE Full text] [doi: 10.1038/s41598-018-34935-1] [Medline: 30425276]
6. Thamjamrassri P, Song Y, Tak J, Kang H, Kong H, Hong J. Customer discovery as the first essential step for successful health information technology system development. Healthc Inform Res 2018 Jan;24(1):79-85 [FREE Full text] [doi: 10.4258/hir.2018.24.1.79] [Medline: 29503756]
7. Buhrmester M, Kwang T, Gosling SD. Amazon's Mechanical Turk: a new source of inexpensive, yet high-quality, data? Perspect Psychol Sci 2011 Jan;6(1):3-5. [doi: 10.1177/1745691610393980] [Medline: 26162106]
8. Osadchiy V, Mills JN, Eleswarapu SV. Understanding patient anxieties in the social media era: qualitative analysis and natural language processing of an online male infertility community. J Med Internet Res 2020 Mar 10;22(3):e16728 [FREE Full text] [doi: 10.2196/16728] [Medline: 32154785]
9. Farsi D. Social media and health care, part i: literature review of social media use by health care providers. J Med Internet Res 2021 Apr 05;23(4):e23205 [FREE Full text] [doi: 10.2196/23205] [Medline: 33664014]
10. Luff MK, Zarrin DA, Zhou L, Sahoo A, Desai S, Iyer N, et al. Clinical needs discovery in pediatric urology: utilizing the biodesign Process. Front Urol 2022 Jun 30;2:895057. [doi: 10.3389/fruro.2022.895057]
11. Amazon.com. URL: https://www.amazon.com/ [accessed 2023-06-19]
12. Wan Y, Nakayama M. Are Amazon.com online review helpfulness ratings biased or not? 2012 Presented at: WEB 2011: E-Life: Web-Enabled Convergence of Commerce, Work, and Social Life; December 4, 2011; Shanghai, China p. 46-54. [doi: 10.1007/978-3-642-29873-8_5]
13. Jiang T, Osadchiy V, Mills J, Eleswarapu S. Is it all in my head? self-reported psychogenic erectile dysfunction and depression are common among young men seeking advice on social media. Urology 2020 Aug;142:133-140. [doi: 10.1016/j.urology.2020.04.100] [Medline: 32437776]
14. Smith GA, Chounthirath T, Splaingard M. Do sleeping children respond better to a smoke alarm that uses their mother's voice? Acad Pediatr 2020 Apr;20(3):319-326. [doi: 10.1016/j.acap.2019.06.016] [Medline: 31276840]

15. McFarlane SJ, Garcia JE, Verhagen DS, Dyer AG. Alarm tones, voice warnings, and musical treatments: a systematic review of auditory countermeasures for sleep inertia in abrupt and casual awakenings. Clocks Sleep 2020 Oct 20;2(4):416-433 [FREE Full text] [doi: 10.3390/clockssleep2040031] [Medline: 33118526]
16. Caldwell PC, Codarini M, Stewart F, Hahn D, Sureshkumar P. Alarm interventions for nocturnal enuresis in children. Cochrane Database Syst Rev 2020 May 04;5(5):CD002911 [FREE Full text] [doi: 10.1002/14651858.CD002911.pub3] [Medline: 32364251]
17. Balasubramanian AT, Thirumavalavan N, Srivatsav A, Yu J, Lipshultz LI, Pastuszak AW. Testosterone imposters: an analysis of popular online testosterone boosting supplements. J Sex Med 2019 Feb;16(2):203-212 [FREE Full text] [doi: 10.1016/j.jsxm.2018.12.008] [Medline: 30770069]
18. Lee A, Mmonu NA, Thomas H, Rios N, Enriquez A, Breyer BN. Qualitative analysis of Amazon customer reviews of penile clamps for male urinary incontinence. Neurourol Urodyn 2021 Jan 09;40(1):384-390. [doi: 10.1002/nau.24572] [Medline: 33165983]
19. Balasubramanian AT, Thirumavalavan N, Srivatsav A, Yu J, Hotaling JM, Lipshultz LI, et al. An analysis of popular online erectile dysfunction supplements. J Sex Med 2019 Jun;16(6):843-852 [FREE Full text] [doi: 10.1016/j.jsxm.2019.03.269] [Medline: 31036522]
20. Vaish A, Grossmann T, Woodward A. Not all emotions are created equal: the negativity bias in social-emotional development. Psychol Bull 2008 May;134(3):383-403 [FREE Full text] [doi: 10.1037/0033-2909.134.3.383] [Medline: 18444702]

Abbreviations

NE: nocturnal enuresis

Edited by T Leung; submitted 27.10.22; peer-reviewed by G Ash, ME Chatzimina , Z Huang; comments to author 16.03.23; revised version received 06.04.23; accepted 26.05.23; published 06.07.23.

Please cite as:
Sahoo A, Starr SL, Osadchiy V, Desai S, Iyer N, Luff M, Sollender GE, Sturm R
Identifying Targets for Innovation in Amazon Reviews of Bedwetting Alarms: Thematic Analysis
Interact J Med Res 2023;12:e43194
URL: https://www.i-jmr.org/2023/1/e43194
doi:10.2196/43194
PMID:37410523

Original Paper

Triggering Weight Management Using Digital Avatars: Prospective Cohort Study

Moyez Jiwa[1], MA, MD; Tafadzwa Nyanhanda[1], BSc, PhD; Michael Dodson[2], BSc, BMedSci, MBBS, MPH, MHM, PhD

[1]School of Medicine, Melbourne Clinical School, The University of Notre Dame Australia, Werribee, Australia
[2]Medical Education Unit, Werribee Mercy Hospital, Mercy Health, Werribee, Australia

Corresponding Author:
Moyez Jiwa, MA, MD
School of Medicine, Melbourne Clinical School
The University of Notre Dame Australia
300 Princes Highway
Werribee, 3030
Australia
Phone: 61 282044259
Email: moyez.jiwa@nd.edu.au

Abstract

Background: There is evidence that showing motivated people with a less-than-ideal BMI (>25 kg/m^2) digital and personalized images of their future selves with reduced body weight will likely trigger them to achieve that new body weight.

Objective: The purpose of this study is to assess whether digital avatars can trigger weight management action and identify some of the measurable factors that distinguish those who may be triggered.

Methods: A prospective cohort study followed participants for 12 weeks through 5 recorded interviews. Participants were screened for suitability for the study using the Cosmetic Procedure Screening Questionnaire as a measure of body dysmorphia. At interview 1, participants were shown 10 images from a "Food-pics" database and invited to estimate their calorie value. The intervention, the FutureMe app, delivered at interview 2, provided each participant an opportunity to see and take away a soft copy of an avatar of themselves as they might appear in the future depending on their calorie consumption and exercise regimen. Participants completed the readiness for change (S-Weight) survey based on Prochaska Stages of Change Model and the processes of change (P-Weight) survey. Any changes in diet, exercise, or weight were self-reported.

Results: A total of 87 participants were recruited, and 42 participants completed the study (48% of recruited participants). Body dysmorphia was a rare but possible risk to participation. The majority (88.5%) of the participants were female and older than 40 years. The average BMI was 34.1 (SD 4.8). Most people wanted to reduce to a BMI of 30 kg/m^2 or lose on average 10.5 kg within 13 weeks (–0.8 kg per week). Most participants stated that they would achieve these results by limiting their calorie intake to 1500 calories per day and taking the equivalent of 1 hour of bicycling per day. At interview 1, more participants were in the preparation stage of behavior change than in subsequent interviews. By interview 5, most of the participants were at the maintenance stage. Participants who overestimated the recommended number of calories were more likely to be in the contemplation stage (P=.03).

Conclusions: Volunteers who participated in the study were mainly women older than 40 years and beyond the contemplation stage of change for weight management, and those who took weight management action were demonstrated to have a more accurate idea of the calorie content of different foods. Most participants set ambitious targets for weight loss, but few, if any, achieve these goals. However, most people who completed this study were actively taking action to manage their weight.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12619001481167; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=378055&isReview=true

(Interact J Med Res 2023;12:e42001) doi:10.2196/42001

KEYWORDS

weight management; digital avatar, behavior change, calorie awareness, obesity, health promotion, motivation, processes of change; stages of change; BMI; weight; body dysmorphia; diet; exercise; calorie; tool; weight management; digital

Introduction

Having a BMI over 25 kg/m^2 is recognized as one of the major risk factors for chronic and life-limiting illnesses [1]. Having overweight or obesity is associated with the consumption of more calories than the body requires for physiological functioning. As a consequence, the body stores the excess calories in the form of adipose tissue. This has a negative effect on health and impacts the perception of a person's physical appearance [2]. Dissatisfaction with physical appearance is one of the most potent drivers for weight management efforts [3]. Among the factors that stimulate an individual to make different lifestyle choices is the desire to achieve a different body shape [4]. There are many diet and exercise programs that an individual can select to achieve that outcome. Individuals will find the means to achieve their goals if they are motivated, feel able to achieve the results, and are triggered to change [5].

This study builds on prior evidence that motivated people with a BMI of >25 kg/m^2 who are are shown digital and personalized images of themselves with a lower body weight may take steps to achieve that new body shape [6]. The intervention in the previous study did not incorporate a tailored program to achieve the desired results [6]. Rather, participants found their own means to achieve their goals or, alternatively, revisited their goals. This study included participants recruited by a hospital employer. There is evidence that a health services employer is well-placed to promote healthy lifestyles [7].

This prospective cohort study involved convenience sampling of participants who self-reported that they wished to address weight management (BMI>25 kg/m^2), obese (BMI>30 kg/m^2), or wanted to maintain their current weight. This study also screened and excluded participants with body dysmorphia. Validated and reliable measures were adopted to follow up on participants' stages of and processes of change. Any actions taken to achieve a lower BMI were self-reported.

Methods

Ethics Approval

The trial was approved by the Mercy Health Human Research Ethics Committee (HR 2019-016) and registered with the Australian New Zealand Clinical Trials Registry (ANZCTR): ACTRN12619001481167. The study adheres to the principles of the Helsinki Declaration. Written informed consent was obtained at inclusion.

Participants and Recruitment

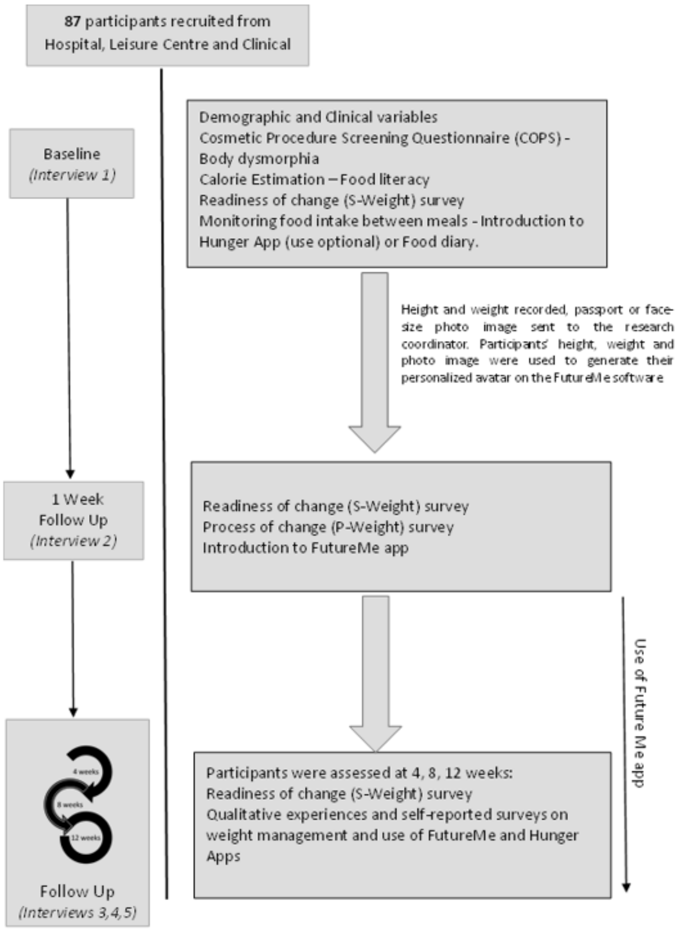
A convenience sample of participants was recruited from hospital staff, a leisure center, and a clinical trial participant registry. Participants were recruited over 13 months, from November 2020 to December 2021. A total of 634 individuals applied to participate and were screened using eligibility criteria, resulting in 87 final participants. Eligibility criteria required participants to be at least 16 years old, have access to a smartphone, be able to give informed consent, and be interested in or consider weight management (regardless of their current BMI). Individuals with body dysmorphia (a mental illness characterized by constant worrying over a perceived or slight defect in one's appearance) and those who were pregnant or breastfeeding were not eligible to participate. Hospital staff participants were recruited through the hospital's website, emails, and flyers.

Study Design

Overview

The study team followed up with participants for 12 weeks and included telephone and video-recorded interviews. As we conducted the study during the COVID-19 pandemic and in different states in Australia, it was not possible to see participants in person. Validated tools were deployed to collect data. The intervention delivered at interview 2 was a previously tested digital tool to trigger efforts at weight management [6]. The study flowchart is shown in Figure 1.

Figure 1. Study flowchart.



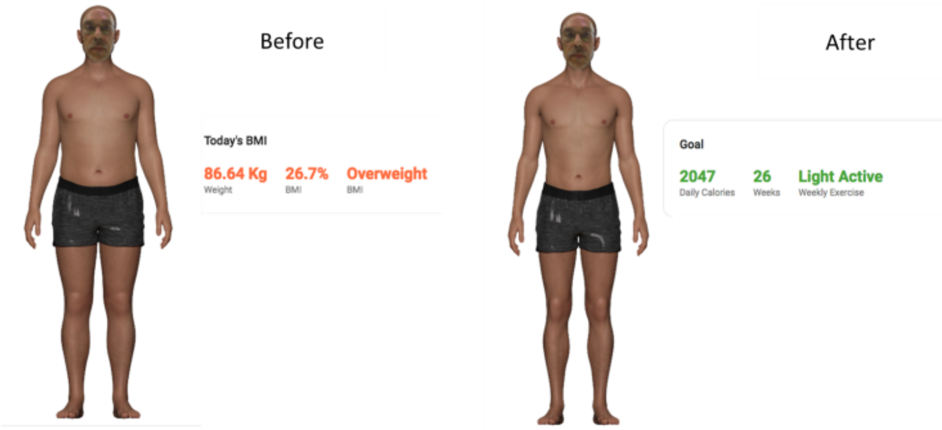
Interview 1

Participants were screened for suitability for the study using the Cosmetic Procedure Screening Questionnaire as a measure of body dysmorphia [8]. Potential participants with suspected body dysmorphia were not eligible to participate and were advised to consult their general medical practitioner for other weight management options.

As a proxy measure of food literacy, participants were shown pictures of up to 10 common foods and asked to estimate their calorie value [9]. Participants completed the 6-item readiness for change (S-Weight) survey, which is based on Prochaska Stages of Change Model [10]. Participants were also invited to monitor their food consumption using an electronic or paper-based food diary. A freely available food diary app was suggested as an option. We did not collect data on the use of any food diaries.

Each participant was invited to attend a review and a second interview 1 week later. Prior to interview 2, participants were reminded to record their height and weight and to send a passport-size or face-size photo image to the research coordinator. A participant's height, weight, and photo image were used to generate their personalized avatar on the FutureMe software (Archetype Health and Continuum Digital). An example of the avatars and the factors that determined the changes in the before and after images and avatars are shown in Figure 2.

Figure 2. Personalized avatar depicting current weight (before) and goal weight (after) images.



Interview 2

Interview 2 was conducted and video-recorded via Zoom. At the interview, participants discussed their food diary with the researcher and undertook the readiness for change (S-Weight) survey. Participants indicated whether they wanted to modify their diet or carry out more exercise (Motivation) and whether they had a strategy to address their food consumption and build more exercise into their routines (Ability). They were then invited to complete a validated process of change questionnaire (P-Weight) about weight management [11]. Following completion of this questionnaire, those who wished to modify their diet or carry out more exercise were referred to freely available resources to assist their efforts and were advised to see their general practitioner or primary care physician or seek a referral for further advice if necessary. All participants were offered web-based resources as well as a screenshot of their "before and after avatars," from the FutureMe app. The participants chose an avatar that reflected how they would like to look at a time of their choosing in the future (12 or 26 weeks), how many daily calories they would consume, and how many weekly calories they would expend in exercise to achieve that goal. The participants were followed up again 2 weeks later.

Interviews 3, 4, and 5

Interviews were conducted at weeks 4, 8, and 12 and video-recorded via Zoom (Zoom Video Communications). Participants repeated the readiness for change (S-Weight) 6-item survey. They were also asked about their experience following the FutureMe app's recommendations with respect to calorie consumption and exercise. The researcher also recorded any steps they had taken to achieve the goals they set for themselves.

Instruments

Participants completed several self-report surveys 3 times in the study (weeks 4, 8, and 12), including sociodemographic characteristics and the English versions of the validated P-Weight and S-Weight surveys [11,12].

- Sociodemographic data: information was gathered about age, sex, and BMI.
- Stages of change for weight management (S-Weight): S-Weight is a questionnaire that consists of 5 mutually exclusive items that aim to allocate participants to one of the 5 stages of change for weight management proposed by the transtheoretical model (precontemplation, contemplation, preparation, action, and maintenance) [10].
- Process of change for weight management (P-Weight): the P-Weight questionnaire aimed to determine the processes (attitudes and behaviors to control weight) involved in the change and included 33 items within 4 change processes: emotional reevaluation, weight management actions (WMAs), supporting relationships, and weight consequences evaluation [11]. The participants answered questions on a 5-point Likert scale ranging from 1 (strong disagreement) to 5 (strong agreement). All scores were obtained and then used to calculate and identify the individuals' process of changing status.
- Calorie estimation: at interview 1, as a proxy measure of food literacy, participants were shown 10 images of common foods obtained from the widely used and validated "Food-pics" database and asked to estimate their calorie value [9]. Participants were shown a visual image of a common snack food and asked to estimate the number of calories in the food by choosing the best answer from multiple options. Participants were also shown 2 images of a man and a woman and asked to estimate the number of calories required for them to maintain their weight, choosing the best answer from multiple options.
- FutureMe app: the FutureMe app was used to provide each participant with the opportunity to see an avatar of themselves as they might appear in the future, depending on their diet and exercise. The participant chose their physical appearance in the future as demonstrated in a full-body avatar, which included their own face and skin color, how many daily calories they should consume, and how many weekly calories they must expend in exercise to achieve that goal at various possible dates in the future, from 12, 26, and 52 weeks. The choice of their "best" look in the future was entirely at the discretion of the participant.

Statistical Analysis

For sample size calculation, we estimated that the proportion of participants likely to be triggered by the Future Me app to make weight loss attempts from our previous RCT would be 20% [6]. Therefore, we aimed to recruit 100 participants to detect a similar proportion triggered within a 5% margin of error at the 99% confidence level. The estimated size of the sample pool of those eligible to participate was 3951 potential participants (n=600 [Hospital]+1800 [Leisure Centre]+1551 [Clinical Trials Registry]).

All surveys were web-based and developed using the Qualtrics web-based platform (Provo) [13]. The surveys were exported from Qualtrics to Excel (Microsoft Corp). Statistical analyses were performed using SPSS Statistics for Windows (version 27.0; IBM Corp). Descriptive statistics (frequencies and percentages for categorical variables, means, and SDs for variables measured on a continuous scale) were used to summarize the participants' demographic data, stages of change at each interview, processes of change at interview 2, and survey responses.

For the determination of the stages of change across all interviews, for dropouts or withdrawals, a "Last Observation Carried Forward" strategy was used to estimate all missing measurements [6], whereby each missing value was replaced by the stage of change at the previous interview.

After completion of the P-Weight questionnaire, a raw process score was generated by tallying up the responses to questions. The raw score (with a range of 5-54) was then converted to a 100-point scale to allow for comparisons across processes. We also categorized participants in the action and maintenance stages of change as being in an active phase of change and those in the precontemplation, contemplation, and preparation stages as being in a nonactive phase of change. We described mean scores with SDs for participants with active and nonactive phases of change and compared means using a 2-tailed independent t test with a significance level set at P=.05. The stages of change (S-Weight) data for the participants were recorded as categorical data and compared to P-Weight scores (numeric data). Comparisons between groups were performed using ANOVA and the Student t test for continuous variables and the chi-square test for categorical variables. The significance of the results was considered with P< .05.

For calorie estimation, responses were grouped as follows: underestimation (any response selected with a caloric value lower than the accurate value), accurate estimation (correct estimation of calories), and overestimation (any response selected with a caloric value higher than the accurate value foods). Foods with more than 150 calories were categorized as high-calorie foods and those below 150 as low-calorie foods. Descriptive statistics were conducted, and the chi-square test was used for association tests. Open-ended survey responses from participants' experiences of using the FutureMe app and weight management were manually coded inductively by emerging themes.

Results

Overview

At interview 1, two potential recruits to the study were found to not be eligible on the basis of potential body dysmorphia. A total of 87 participants were recruited, and 42 participants completed the study (48% of recruited participants). Demographic data are presented in Table 1. The majority of participants were female, and all but 2 participants were over 25 years. The average BMI was 34.1 (SD 4.8). Most people wanted to reduce to a BMI of 30 or lose on average 10.5 kg within 13 weeks (–0.8 kg per week). Most participants stated that they would achieve these results by taking a moderate amount of exercise, equivalent to using 580 calories per day or 1 hour of cycling per day, and eating no more than 1500 calories per day on average.

Table 1. Study participants' demographics.

	Interview 1 (N=67)	Interview 2 (N=75)	Interview 3 (N=63)	Interview 4 (N=47)	Interview 5 (N=42)
Gender, n (%)					
Male	9 (10.3)	9 (12)	8 (12.7)	8 (17)	8 (19.1)
Female	77 (88.5)	66 (88)	55 (87.3)	39 (83)	34 (80.9)
Nonbinary	1 (1.2)	N/A[a]	N/A	N/A	N/A
Age (years), n (%)					
18-25	2 (2.3)	N/A	N/A	N/A	N/A
26-40	21 (24.1)	19 (25.3)	15 (23.8)	11 (23.4)	10 (23.8)
41-50	19 (21.8)	14 (18.7)	12 (19.1)	10 (21.3)	9 (21.4)
51-60	30 (34.5)	28 (37.3)	24 (38.1)	15 (31.9)	13 (31)
<60	15 (17.3)	14 (18.7)	12 (19)	11 (23.4)	10 (23.8)

[a]N/A: not applicable.

Estimation of Daily Caloric Requirements

At interview 2, one in 3 participants accurately determined the recommended number of calories required by men and women. A significant percentage of participants (32.3%) overestimated the requirement, while 26.4% and 8% of them underestimated and did not know, respectively.

Calorie Estimation of Foods

Overall, more participants tended to overestimate the calorie content of low-calorie foods compared to high-calorie foods. A higher percentage accurately determined the number of calories in high-calorie foods compared to low-calorie foods. Participants older than 40 years were more likely to overestimate the calorie content of calorie-dense foods, whereas participants aged between 26 and 40 years were more likely to give an accurate estimation (P=.05). Participants who overestimated the recommended number of calories were more likely to be in the contemplation stage (P=.03). There were no significant associations between calorie estimation and stages of behavior change.

Stages of Change

A higher number of participants at interviews 3 and 4 were in the action stage compared to earlier interviews, and there were fewer participants at interview 5 in the preparation stage compared to interview 1 (P=.03). By interview 5, most of the participants were in the action and maintenance stages.

Processes of Change

At interview 2, emotional reevaluation was the most common change process used.

We categorized participants in the action and maintenance stages of change as being in an active phase of change and those in the precontemplation, contemplation, and preparation stages as being in a nonactive phase of change. The mean processes of change scores with SD are provided in Table 2 below.

Participants in the active phase of change (n=59) had significantly higher mean WMA scores (active phase mean WMA score 53.8, SD 11.9 vs nonactive phase mean WMA score 45.6, SD 16, P=.02) than those in the nonactive phase of change (n=16).

Table 2. Active and nonactive phases—processes of change mean scores.

	Emotional reevaluation (EmR)	Weight consequences evaluation (WCE)	Supporting relationships (SRs)	Weight management actions (WMAs)
Nonactive phase, mean (SD; n)	75.3 (10.9; 16)	59.6 (18.4; 16)	58.4 (16.4; 16)	45.6 (12.3; 16)
Active phase, mean (SD; n)	76.8 (9.7; 59)	54.3 (14.5; 59)	57.5 (15.2; 59)	53.8 (11.9; 59)
Total, mean (SD; n)	76.5 (9.9; 75)	55.4 (15.4; 75)	57.7 (15.4; 75)	52.1 (12.4; 75)

Self-reported Attempts at Weight Management

Approximately half of the participants adopted positive weight management actions in the 4 weeks prior to interview 5. For those participants who completed interview 5 (21.4%) reported reducing their calorie intake in the 4 weeks prior to the interview, and 14.3% of them had not changed their diet but claimed to be exercising regularly.

More than half the participants (54.8%) reported following the FutureMe recommendations at least 4 times a week or more.

Many participants (45%) self-reported that it was difficult to limit their calorie intake or to exercise more.

Despite some participants finding it difficult to consistently eat less and exercise more, others described experiencing positive results. For example, a few reported losing weight and changing their exercise and eating habits. At the end of the study, 3 male participants and 3 female participants who chose to reply to the open-ended questions reported positive results. Some described how participating in the research and using the FutureMe app held them accountable and motivated them (Textbox 1).

Textbox 1. Participants' description of participating in the research.

- "Lost 3kg so far" [Participant 75, 41-50 years, Male]
- "Lost 11.3kg since starting, changing my target to 95, rather than 100 kg initial goal" [Participant 78, >60 years, Male]
- "Hit goal weight from sticking to calorie goal, now trying to maintain" [Participant 24, >60 years, Female]
- "It has been fairly easily once gotten used to reduced calories and calorie counting in diary keeps you on track" [Participant 59, >60 years, Female]
- "General well-being has improved and managed to lose weight from following recommendations and exercising more" [Participant 21, 41-50 years, Male]

Discussion

Principal Results

Upon deploying FutureMe, a similar proportion of the participants as reported previously may have been triggered to take weight management actions [6]. Shifts in the movement of participants between the different stages of behavioral change from interviews 1 to 5 were evident. At interview 1, more participants were in the preparation stage. By interview 5, most of the participants were at the maintenance stage. In this active stage of change, the data suggest that participants were more likely to take weight management action. Therefore, it is plausible and consistent with the self-reported fact that most participants who completed the study were actively working to manage their weight.

It was also evident from the data, as well as from the literature, that most of the participant's knowledge about the calorie content of food was not accurate [14]. The data suggest that those with a more accurate understanding of the calorie content of food were more likely to be triggered to take weight management action. We also acknowledge that our participants' ability to take the necessary action may have been limited by other factors that were not explored in this observational study but have been the focus of study by others [15].

The data add to the literature that motivated people can be triggered to make attempts to enhance their lifestyle by personalized avatars of their future selves [16]. Most of the data so far relate to virtual reality images and not to the types of avatars used in this study. The targets for weight loss set by the participants in this study were challenging. A weight loss of 10 kg in 13 weeks starting with a BMI of 34.1, well into the obese range, would require major lifestyle changes. Participants were invited to reconsider these targets at interview 2. At that interview, they were shown how they would have to restrict their diet and increase their exercise to achieve the desired appearance. Knowledge of the significant cost of diet restriction and commitment to regular exercise, as suggested by FutureMe, did not change participants' weight loss goals. Although these goals may be ambitious, there is some evidence in the literature that higher goals motivate weight loss more than undermine effort [17].

Comparison With Prior Work

Prior work using the FutureMe app intervention focused on documented weight loss and the timing of showing subjects their future avatars from the point of recruitment [6]. In this study, we focused on the participants' knowledge of calorie values in food, the stage of change, and the process of change associated with weight management. This may help health professionals identify those who might be triggered as well as when to introduce the avatars. We also note that in our studies so far, the actual weight loss has been much lower than the target of 0.8 kg per week set by participants. Both the ability to achieve a self-selected target and the initial onboarding with reference to calorie awareness may be important for those who attempt weight management.

Though limited research has been published, the inclusion of avatar technology in weight loss interventions triggers weight maintenance [6,18-20]. Avatar personalization, with the person's actual face, skin color, hairstyle, and personal choices of diet and exercise, seems to be important in triggering weight management [6,21].

Limitations

One limitation of our study is that we did not achieve the target of 100 participants. Most volunteer participants were older females; this may reflect the appeal of the intervention to that demographic, but we cannot report on its value in other groups. The study was also conducted at a time of pandemic-related lockdown across state borders in Australia, so weight loss was not confirmed by in-person weight measurements. Therefore, we were unable to verify any weight loss reported. We were also advised to adopt self-reports, as those who might fail might be distressed if they were to reflect on these data. We, therefore, acknowledge that social desirability bias cannot be ruled out. There was also a high attrition rate in the study. However, we had no indication that those who dropped out of the study had specific characteristics. Additionally, the observed attrition rate of 52% is on par with most weight-loss trials, where attrition is commonly at least between 20% and 50% [22].

Conclusions

In this study, participants in the FutureMe intervention mainly included women older than 40 years. The most promising results were for those who could more accurately estimate the calorie content of food and were beyond the contemplation stage of change. The participants were offered a trigger to manage their weight, primarily in the form of an avatar depicting their future selves after a period of calorie restriction and increased exercise. As in the previous trial, this may have been effective for many participants [6]. Most of those who were followed up at 12 weeks reported actively working to manage their weight.

However, most participants set very ambitious goals for calorie restriction and exercise regimens that could not be achieved in the time they had nominated to achieve those results. It appears that most people who opted to participate were unlikely to succeed based on their early stage of change, their inaccurate estimates of the calorie content of food, and the ambitious goals they set for calorie restriction and exercise regimens. Motivation and being triggered are not sufficient to achieve behavior change if the person is not able to achieve their goals. Our data suggest that some aspects of knowledge that are necessary in regard to weight management were lacking in the volunteers in this study.

Acknowledgments

The authors would like to thank Kaitlyn Smith, Anwyn Hocking, Dr Catherine Krejany, and Epi Kanjo for performing the initial study design and data collection. The study was funded by the Digital Health Collaborative Research Center.

Authors' Contributions

MJ developed the study design, interpreted data analysis, drafted the manuscript, and reviewed the final manuscript draft. TN analyzed the raw data, interpreted the data analysis, drafted the manuscript, and reviewed the final manuscript draft. MD developed the study design, interpreted data analysis, drafted the manuscript, and reviewed the final manuscript draft.

Conflicts of Interest

None declared.

References

1. Bray GA. Medical consequences of obesity. J Clin Endocrinol Metab 2004;89(6):2583-2589 [FREE Full text] [doi: 10.1210/jc.2004-0535] [Medline: 15181027]

2. Makara-Studzińska M, Zaborska A. Obesity and body image. Psychiatr Pol 2009;43(1):109-114 [FREE Full text] [Medline: 19694405]
3. Rosen JC. Improving body image in obesity. In: Thompson JK, editor. Body Image, Eating Disorders, and Obesity: An Integrative Guide for Assessment and Treatment. Washington, DC, US: American Psychological Association; 2001:425-440.
4. Munt AE, Partridge SR, Allman-Farinelli M. The barriers and enablers of healthy eating among young adults: a missing piece of the obesity puzzle: a scoping review. Obes Rev 2016;18(1):1-17 [FREE Full text] [doi: 10.1111/obr.12472] [Medline: 27764897]
5. Jiwa M, Burford O, Parsons R. Preliminary findings of how visual demonstrations of changes to physical appearance may enhance weight loss attempts. Eur J Public Health 2015;25(2):283-285 [FREE Full text] [doi: 10.1093/eurpub/cku249] [Medline: 25644139]
6. Ossolinski G, Jiwa M, McManus A, Parsons R. Do images of a personalised future body shape help with weight loss? A randomised controlled study. Trials 2017;18(1):180 [FREE Full text] [doi: 10.1186/s13063-017-1907-6] [Medline: 28420417]
7. Heinen L, Darling H. Addressing obesity in the workplace: the role of employers. Milbank Q 2009;87(1):101-122 [FREE Full text] [doi: 10.1111/j.1468-0009.2009.00549.x] [Medline: 19298417]
8. Veale D, Ellison N, Werner TG, Dodhia R, Serfaty MA, Clarke A. Development of a cosmetic procedure screening questionnaire (COPS) for body dysmorphic disorder. J Plast Reconstr Aesthet Surg 2012;65(4):530-532 [FREE Full text] [doi: 10.1016/j.bjps.2011.09.007] [Medline: 22000332]
9. Blechert J, Meule A, Busch NA, Ohla K. Food-pics: an image database for experimental research on eating and appetite. Front Psychol 2014;5:617 [FREE Full text] [doi: 10.3389/fpsyg.2014.00617] [Medline: 25009514]
10. Prochaska JO, Velicer WF, Rossi JS, Goldstein MG, Marcus BH, Rakowski W, et al. Stages of change and decisional balance for 12 problem behaviors. Health Psychol 1994;13(1):39-46 [FREE Full text] [doi: 10.1037//0278-6133.13.1.39] [Medline: 8168470]
11. Andrés A, Saldaña C, Gómez-Benito J. Establishing the stages and processes of change for weight loss by consensus of experts. Obesity (Silver Spring) 2009;17(9):1717-1723 [FREE Full text] [doi: 10.1038/oby.2009.100] [Medline: 19360014]
12. Andrés A, Saldaña C, Beeken RJ. Assessment of processes of change for weight management in a UK sample. Obes Facts 2015;8(1):43-53 [FREE Full text] [doi: 10.1159/000373900] [Medline: 25765163]
13. Qualtrics. URL: https://www.qualtrics.com [accessed 2023-05-03]
14. Larkin D, Martin CR. Caloric estimation of healthy and unhealthy foods in normal-weight, overweight and obese participants. Eat Behav 2016;23:91-96 [FREE Full text] [doi: 10.1016/j.eatbeh.2016.08.004] [Medline: 27591830]
15. Albright CA, Pratt KJ, Martin SB, Hulshult H, Brown CL, Lewis KH, et al. Family members' experiences with adult participation in weight management programs: triadic perspectives from patients, partners and children. Clin Obes 2020;10(2):e12354 [FREE Full text] [doi: 10.1111/cob.12354] [Medline: 31965733]
16. Taylor L, Ranaldi H, Amirova A, Zhang L, Ahmed AA, Dibb B. Using virtual representations in mHealth application interventions for health-related behaviour change: a systematic review. Cogent Psychol 2022;9(1):2069906 [FREE Full text] [doi: 10.1080/23311908.2022.2069906]
17. Linde JA, Jeffery RW, Levy RL, Pronk NP, Boyle RG. Weight loss goals and treatment outcomes among overweight men and women enrolled in a weight loss trial. Int J Obes (Lond) 2005;29(8):1002-1005 [FREE Full text] [doi: 10.1038/sj.ijo.0802990] [Medline: 15917847]
18. Behm-Morawitz E, Lewallen J, Choi G. A second chance at health: how a 3D virtual world can improve health self-efficacy for weight loss management among adults. Cyberpsychol Behav Soc Netw 2016;19(2):74-79 [FREE Full text] [doi: 10.1089/cyber.2015.0317] [Medline: 26882324]
19. Cesa GL, Manzoni GM, Bacchetta M, Castelnuovo G, Conti S, Gaggioli A, et al. Virtual reality for enhancing the cognitive behavioral treatment of obesity with binge eating disorder: randomized controlled study with one-year follow-up. J Med Internet Res 2013;15(6):e113 [FREE Full text] [doi: 10.2196/jmir.2441] [Medline: 23759286]
20. Manzoni GM, Cesa GL, Bacchetta M, Castelnuovo G, Conti S, Gaggioli A, et al. Virtual reality-enhanced cognitive-behavioral therapy for morbid obesity: a randomized controlled study with 1 year follow-up. Cyberpsychol Behav Soc Netw 2016;19(2):134-140 [FREE Full text] [doi: 10.1089/cyber.2015.0208] [Medline: 26430819]
21. Napolitano MA, Hayes S, Russo G, Muresu D, Giordano A, Foster GD. Using avatars to model weight loss behaviors: participant attitudes and technology development. J Diabetes Sci Technol 2013;7(4):1057-1065 [FREE Full text] [doi: 10.1177/193229681300700430] [Medline: 23911189]
22. Moroshko I, Brennan L, O'Brien P. Predictors of dropout in weight loss interventions: a systematic review of the literature. Obes Rev 2011;12(11):912-934 [FREE Full text] [doi: 10.1111/j.1467-789X.2011.00915.x] [Medline: 21815990]

Abbreviations

WMA: weight management action

Edited by A Mavragani; submitted 18.08.22; peer-reviewed by N Farič, J Kvedar; comments to author 03.02.23; revised version received 22.02.23; accepted 09.03.23; published 29.05.23.

Please cite as:
Jiwa M, Nyanhanda T, Dodson M
Triggering Weight Management Using Digital Avatars: Prospective Cohort Study
Interact J Med Res 2023;12:e42001
URL: https://www.i-jmr.org/2023/1/e42001
doi:10.2196/42001
PMID:37247208

Original Paper

Validation of a Brief Internet-Based Self-Report Measure of Maladaptive Personality and Interpersonal Schema: Confirmatory Factor Analysis

Hyeonseong Kim[1*], MA; Seohyun Jeong[1*], MA; Inae Hwang[2], MA; Kiyoung Sung[2], MD; Woori Moon[2], MD, PhD; Min-Sup Shin[1,3], PhD

[1]Psychiatry Department, Seoul National University Hospital, Seoul, Republic of Korea

[2]40FY Inc, Seoul, Republic of Korea

[3]Seoul National University College of Medicine, Seoul, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Min-Sup Shin, PhD
Psychiatry Department
Seoul National University Hospital
101 Daehak-ro, Jongno-gu
6th Floor
Seoul, 03080
Republic of Korea
Phone: 82 0220720725
Email: shinms@snu.ac.kr

Abstract

Background: Existing digital mental health interventions mainly focus on the symptoms of specific mental disorders, but do not focus on Maladaptive Personalities and Interpersonal Schemas (MPISs). As an initial step toward considering personalities and schemas in intervention programs, there is a need for the development of tools for measuring core personality traits and interpersonal schemas known to cause psychological discomfort among potential users of digital mental health interventions. Thus, the MPIS was developed.

Objective: The objectives of this study are to validate the MPIS by comparing 2 models of the MPIS factor structure and to understand the characteristics of the MPIS by assessing its correlations with other measures.

Methods: Data were collected from 234 participants who were using web-based community sites in South Korea, including university students, graduate students, working professionals, and homemakers. All the data were gathered through web-based surveys. Confirmatory factor analysis was used to compare a single-factor model with a 5-factor model. Reliability and correlation analyses with other scales were performed.

Results: The results of confirmatory factor analysis indicated that the 5-factor model (χ^2_{550}=1278.1; Tucker-Lewis index=0.80; comparative fit index=0.81; and Root Mean Square Error of Approximation=0.07) was more suitable than the single-factor model (χ^2_{560}=2341.5; Tucker-Lewis index=0.52; comparative fit index=0.54; and Root Mean Square Error of Approximation=0.11) for measuring maladaptive personality traits and interpersonal relationship patterns. The internal consistency of each factor of the MPIS was good (Cronbach α=.71-.88), and the correlations with existing measures were statistically significant. The MPIS is a validated 35-item tool for measuring 5 essential personality traits and interpersonal schemas in adults aged 18-39 years.

Conclusions: This study introduced the MPIS, a concise and effective questionnaire capable of measuring maladaptive personality traits and interpersonal relationship schemas. Through analysis, the MPIS was shown to reliably assess these psychological constructs and validate them. Its web-based accessibility and reduced item count make it a valuable tool for mental health assessment. Future applications include its integration into digital mental health care services, allowing easy web-based administration and aiding in the classification of psychological therapy programs based on the obtained results.

Trial Registration: ClinicalTrials.gov NCT05952063; https://www.clinicaltrials.gov/study/NCT05952063

(Interact J Med Res 2023;12:e48425) doi:10.2196/48425

KEYWORDS

maladaptive schema; measure of schema; self-report measure; internet-based measure; digital mental health care; interpersonal schema

Introduction

Digital mental health services include mental health care programs provided via web or mobile platforms, thereby extending traditional face-to-face mental health care. These services use web-based platforms, offering advantages such as cost savings and improved accessibility. Digital mental health services have been rapidly growing in the recent years because of these benefits [1]. Many studies have reported their effectiveness in treating a range of mental disorders such as obsessive-compulsive disorder, depression, panic disorder, insomnia, and attention-deficit/hyperactivity disorder [2,3]. However, digital mental health services that are currently reported to be effective, mostly focus on the treatment of specific mental disorders or syndromes, thus making it difficult to expand their scope to enhance the mental health of the general population. Thus, meeting the needs of adults reporting a diverse range of psychological discomfort, ranging from subclinical symptoms to personality issues, is still a challenge.

Alternatively, interventions that focus on maladaptive personality traits and coping styles should be considered in the field of relevant research [4]. Psychological discomfort varies depending on how individuals identify and respond to stress [5]. Personality traits refer to an individual's tendencies to think and act consistently or similarly in different situations, and certain personality traits can become risk factors for stress in specific situations [6]. For instance, a trait that places significant importance on anxiety and exhibits avoidance behaviors can trigger stress in an individual more easily in situations like an examination or a presentation.

To provide interventions based on maladaptive personality traits within digital mental health services, it is necessary to conduct a selection process for representative types of maladaptive personality traits. In that regard, a review of prior research is essential to determine how maladaptive personality traits can be categorized. The Diagnostic and Statistical Manual of Mental Disorders, 5th edition (DSM-5) [7] identifies five broad trait domains: (1) negative affectivity, (2) detachment, (3) antagonism, (4) disinhibition, and (5) psychoticism, as maladaptive variants corresponding to the traits presented in the 5-factor model of personality. Each trait domain is further divided into 25 specific trait facets, including insecurity, hostility, withdrawal, and rigid perfectionism. Additionally, some researchers have considered early maladaptive schemas, a key concept in schema therapy, as maladaptive personality traits [8]. Early maladaptive schemas are defined as self-defeating patterns of emotion and thought that begin early in an individual's life and are repeated throughout their life [9]. It comprises five broad categories: (1) disconnection and rejection, (2) impaired autonomy and performance, (3) impaired limits, (4) other-directedness, and (5) overvigilance and inhibition, along with 18 psychological schemas under subcategories like emotional deprivation, dependence or incompetence, entitlement or grandiosity, self-sacrifice, and unrelenting standards or hypercriticalness [10].

Due to the diversity of maladaptive personality traits, digital mental health care services seeking to intervene must specify and evaluate the maladaptive personality traits that the intervention seeks to mitigate. It is important to note that intervention and evaluation must be conducted in a non–face-to-face environment. However, current personality traits and schema assessment scales may not be well-suited for on the internet use due to their length and large number of items, which makes them unsuitable for web-based implementation and thereby increases the entry barriers to intervention programs at the initial assessment phase. For example, the Personality Inventory for DSM-5 (PID-5), which assesses the personality traits of the alternative model of DSM-5 personality disorders, presented earlier, comprises 220 items [11], and the Young Schema Questionnaire-Long form-3 (YSQ-L3), which is mainly used to evaluate early maladaptive schemas, is composed of 232 items [9]. Furthermore, an intuitive tool is essential for web-based interventions, as users must independently recognize the need for intervention through an initial assessment and evaluate the perceived effectiveness of the program by checking the trend of self-reported scores before deciding whether to continue. However, it is difficult for potential users of digital mental health services to intuitively understand their psychological characteristics without an expert's assistance when using existing scales. For example, there are 25 trait facets in PID-5, and it is difficult to understand each trait facet without an expert explanation. Furthermore, it is difficult to intuitively understand various maladaptive personality patterns depending on whether the level of each facet is high or low, and a combination of it. While in schema therapy, psychological schemas may be explored through additional interviews instead of relying solely on results of the YSQ-L3 [9]. Therefore, considering the potential implications for digital mental health services, a valid assessment scale that can evaluate an individual's maladaptive psychological characteristics on the internet in a concise and intuitive manner is critical.

For this study, the Maladaptive Personality and Interpersonal Schema (MPIS; 40FY Inc, unpublished), which is a brief internet-based self-report measure, was developed as a way to measure maladaptive psychological characteristics and schemas of individuals. Input from 1 psychiatrist and 3 clinical psychologists, examination of various existing scales, and consultation with experts on item development, reliability, and validity were crucial steps in developing the scale. In the process of scale development, 1 psychiatrist and 3 clinical psychology experts generated 50 items. Subsequently, correlation analysis and exploratory factor analysis were conducted among the items, leading to the final selection of 35 items for the MPIS. Exploratory factor analyses were conducted, and five primary psychological schemas that have the potential to induce stress were identified as follows: (1) Shelly (social isolation), related to lack of belongingness to a group and social skills; (2) Flippy (anger), related to a lack of patience and hot-tempered coping style; (3) Riggy (perfectionism), related to strict standards; (4) Pleaser (low self-esteem), related to a tendency to excessively pursue affection and attention from others; and (5) Jumpy (anxiety), related to pessimistic and anxious tendencies. Through web-based survey responses, participants could identify which maladaptive psychological characteristics and schemas they aligned with and also obtain information about the severity of those characteristics and schemas. This study thus verified the reliability and validity of the MPIS.

Methods

Recruitment and Survey Procedure

This study recruited adult participants from large web-based community sites. The website used for participant recruitment was a community site open to a diverse range of occupations and age groups, including university students, graduate students, working professionals, and homemakers. The recruitment period was from July 26, 2022, to August 16, 2022.

This study used a cross-sectional design with a convenience sample. The inclusion criteria were as follows: (1) 18 to 39 years of age; (2) the ability to read and write in Korean without external assistance; and (3) the ability to access websites and respond to questions using a mobile phone or computer. Furthermore, individuals who refused to complete the survey without a specific reason or deliberately gave inappropriate responses repeatedly during their participation were excluded from this study. A 2-step method was used to detect careless responses [12]. First, the ratio of the number of identical responses to the total number of items for each participant was calculated, and the top 5% (12/234) of participants with the highest repetition of the identical responses were selected. Second, an analysis was conducted to determine whether these participants' responses appeared as straight-line patterns within their respective questionnaires. If straight linings were not detected, their responses were included in the final analysis.

Instructions on how to complete and submit this study's participation application on the internet were provided in the form of a post on the internet community site. Participants who completed this study's participation application received explanations about the participation and consent form completion process through phone calls or emails. Additionally, participants were sent an email containing a link to access the web-based survey. Once participants submitted their final responses, the answers were automatically saved on the researchers' computer server.

The target participation size for conducting factor analysis was determined to be at least 5 times the number of MPIS items, which amounted to 175 individuals [13]. In internet-based survey research, participant dropout rates have been observed to vary widely, typically ranging from 10% to 30% [14]. Additionally, the rate of careless responses varies from 0.5% to 14% [12]. Adopting a conservative perspective, the rate of dropout and careless responses was assumed to be 40% (70/175). As a result, the final target number of participants was set at 250 individuals.

Participants voluntarily completed a study consent form and then took part in a 30-minute web-based survey. Survey items were presented in a predetermined order to prevent response bias. Additionally, all surveys displayed only 2-3 questions per screen, and as participants completed their responses, the screen would automatically scroll to the next set of questions. Each page of the survey included a maximum of 45 items, and the survey consisted of a total of 15 pages. A completeness check was implemented for each item, requiring participants to answer all questions on a page before proceeding to the next page. Furthermore, a review step was provided to allow participants to review and modify their answers before submitting their responses. To identify unique visitors, only 1 response per participant was collected using an assigned URL for each participant. Prior to participant recruitment, technical tests were conducted to ensure the proper functioning of the web-based survey.

The informed consent form specified the survey duration, types of data collected, data retention period, investigators, research objectives, scope of personal information collection, and anonymization methods for ensuring personal information protection. As compensation for their participation, participants received a monetary reward of 10,000 KRW (equivalent to US $7 as of August 2023) and the MPIS results report.

Measures

Measure of MPIS

The MPIS comprises 35 items rated on a 5-point Likert scale, with higher scores indicating a greater presence of dysfunctional psychological personality traits. In the process of scale development, the internal consistency (Cronbach α) of the scale was .86 for Shelly, .86 for Flippy, .78 for Riggy, .74 for Pleaser, and .85 for Jumpy.

Perceived Stress Scale

The Perceived Stress Scale (PSS) [15,16] comprises 10 items rated on a 4-point Likert scale, with 6 items assessing negative perceptions and 4 assessing positive ones. The items reflecting positive perceptions are reverse scored with a higher overall score indicating a higher degree of perceived stress. The PSS assesses the extent to which individuals perceive life events as stressful and experience unpredictability, uncontrollability, and excessive demands in daily life. The negative and positive perceptions were found to be positively correlated with depression, anxiety, and negative emotions, and negatively correlated with positive emotions. In this study, the internal consistency was 0.87 for negative perceptions and 0.73 for positive perceptions.

Self-Efficacy Scale

The Self-Efficacy Scale (SES) [17,18] has 17 items related to achievement self-efficacy and 6 related to social self-efficacy, rated on a 5-point Likert scale. Higher scores indicate a higher level of self-efficacy. In this study, the internal consistency was 0.93 for achievement self-efficacy and 0.75 for social self-efficacy.

Center for Epidemiologic Studies Depression Scale

The Center for Epidemiologic Studies Depression Scale (CES-D) [19,20] comprises 20 items rated on a 4-point Likert scale, measuring depression experienced by individuals over the past week. In this study, the internal consistency of the scale was found to be 0.93.

State-Trait Anxiety Inventory

The State-Trait Anxiety Inventory (STAI) [21] comprises 20 items for state anxiety and 20 for trait anxiety on a 4-point Likert scale. In this study, 20 items related to trait anxiety were used. The internal consistency in this study was 0.92 for trait anxiety.

University of California-Los Angeles Loneliness Scale

The University of California-Los Angeles Loneliness Scale (UCLA-LS) [22,23] comprises 20 items rated on a 4-point Likert scale that asks about social relationships in a positive or negative direction. Items in the positive direction were reverse scored and summed with a higher score indicating a higher degree of loneliness. In this study, the internal consistency was 0.94.

Short Form of the Korean Inventory of Interpersonal Problems Circumplex Scale

The short form of the Korean Inventory of Interpersonal Problems Circumplex Scale (KIIP-SC) [24] comprises 40 items rated on a 5-point Likert scale, 18 of which address topics associated with having difficulties relating to others and 22 of which describe situations where it seems like things are "too much." Subfactors are Korean Inventory of Interpersonal Problems domineering or controlling, Korean Inventory of Interpersonal Problems vindicative or self-centered, Korean Inventory of Interpersonal Problems cold or distant, Korean Inventory of Interpersonal Problems socially inhibited or avoidant (KIIP-FG), Korean Inventory of Interpersonal Problems nonassertive (KIIP-HI), Korean Inventory of Interpersonal Problems overly accommodating or exploitable, Korean Inventory of Interpersonal Problems self-sacrificing or overly nurturant (KIIP-LM), and Korean Inventory of Interpersonal Problems intrusive or needy (KIIP-NO). In this study, the overall internal consistency was 0.93, and the internal consistency for each subfactor was 0.71 (Korean Inventory of Interpersonal Problems domineering or controlling), 0.81 (Korean Inventory of Interpersonal Problems vindicative or self-centered), 0.85 (Korean Inventory of Interpersonal Problems cold or distant), 0.88 (KIIP-FG), 0.88 (KIIP-HI), 0.80 (Korean Inventory of Interpersonal Problems overly accommodating or exploitable), 0.68 (KIIP-LM), and 0.66 (KIIP-NO).

Diagnostic Test for Personality Disorders-Dependent Subscale

The Diagnostic Test for Personality Disorders-Dependent subscale (DTPD-D) [25] is a 15-item instrument rated on a 4-point Likert scale that measures dependent personality disorder, as defined in the DSM-IV, and its predisposition. In this study, the internal consistency was 0.86.

State-Trait Anger Expression Inventory

The State-Trait Anger Expression Inventory (STAXI) [26,27] comprises 20 items for state and trait anger and 24 items for anger expression style (anger-in, anger-out, and anger-control) on a 4-point Likert scale. In this study, 24 items related to anger expression style were used. The internal consistency in this study was 0.82 for anger-in, 0.82 for anger-out, and 0.60 for anger-control.

Impulsive Behavior Scale

The Urgency-Premeditation-Perseverance-Sensation seeking -Positive urgency [28-30] is a 59-item measure, rated on a 4-point Likert scale with five subfactors: (1) negative urgency, (2) lack of premeditation, (3) lack of perseverance, (4) sensation seeking, and (5) positive urgency. Only items belonging to the negative urgency category were used for this study. In the present sample, the overall internal consistency was 0.95, and the internal consistency for the negative urgency factor was 0.89.

Hewitt Multidimensional Perfectionism Scale

The Hewitt Multidimensional Perfectionism Scale (HMPS) [18,31] comprises 45 items rated on a 7-point Likert scale with three subfactors: (1) self-oriented perfectionism, (2) other-oriented perfectionism, and (3) socially prescribed perfectionism. In this study, the internal consistency was 0.92 for self-oriented perfectionism, 0.82 for other-oriented perfectionism, and 0.85 for socially prescribed perfectionism.

State Self-Esteem Scale

The State Self-Esteem Scale (SSES) [32,33] is a 20-item measure, with 7 items on performance self-esteem, 6 on appearance self-esteem (referred to as "general self-esteem" in the Korean version), and 7 on social self-esteem, rated on a 5-point Likert scale. A higher score is associated with higher self-esteem. In this study, internal consistency was 0.87 for performance self-esteem, 0.81 for appearance self-esteem, and 0.86 for social self-esteem.

Rejection Sensitivity Questionnaire

The Rejection Sensitivity Questionnaire (RSQ) [34,35] comprises 18 situations in which an individual makes demands of a significant other, such as a parent, friend, or lover in daily life. In each situation, the anxiety associated with rejection of one's request (rejection anxiety) and the expectation that the other person will accept it (acceptance expectation) were evaluated on a 6-point Likert scale. Items corresponding to acceptance expectations were reverse scored and interpreted as rejection expectations. In this study, the internal consistency was 0.93.

Beck Anxiety Inventory

The Beck Anxiety Inventory (BAI) [36,37] is a 21-item measure for evaluating anxiety severity on a 4-point Likert scale. A total score of 0-7 corresponds to normal anxiety, 8-15 to mild anxiety, 16-25 to moderate anxiety, and 26-63 to high anxiety. The internal consistency in this study was 0.95.

Intolerance of Uncertainty Scale

The Intolerance of Uncertainty Scale (IUS) [38-40] is a 12-item measure rated on a 4-point Likert scale. In this study, the internal consistency was 0.87.

Penn State Worry Questionnaire

The Penn State Worry Questionnaire (PSWQ) [41,42] is a 16-item instrument rated on a 5-point Likert scale. It measures the frequency and intensity of chronic, uncontrollable worry. In this study, the internal consistency was 0.94.

Data Analysis

Data for this study were analyzed using the following processes. First, confirmatory factor analysis (CFA) was conducted using the Mplus (version 7; Muthén & Muthén) for Windows. A valid model was established by comparing the goodness of fit indices of the hypothesized 5-factor model with the alternative single-factor model. The estimation of model parameters was conducted using the maximum likelihood method. To evaluate the goodness of fit of the model, the chi-square, Tucker-Lewis index (TLI), comparative fit index (CFI), and root-mean-square error of approximation (RMSEA) values were reported. The cutoff used for the factor loading to remove any item from the MPIS was 0.4 [13]. Even when the factor loading of an item was below 0.4, the final decision on including the item in the MPIS was based on comparing model fit indices and information indices between models with and without the item, and considering the item's content. In addition, internal consistency was assessed using Cronbach α. Convergent and discriminant validity were then evaluated using the Pearson correlation coefficient.

Ethical Considerations

All study procedures were approved by the institutional review board of Seoul National University Hospital (IRB H-2203-108-1309).

Results

Characteristics of Participants

A total of 323 individuals applied to participate in this study, and after excluding those who withdrew from the survey, a total of 250 participants successfully completed the web-based survey. The completion rate was 77% (250/323). A total of 16 participants who were outside the age range were excluded. The analysis was conducted only on the completed questionnaires. There was no participant who met the criteria for careless response, and all 234 samples were included in the analysis. The participants had a mean age of 26.33 (SD 5.41) years. Table 1 presents the demographic information.

Table 1. Demographic information (N=234).

	Value, n (%)
Sex	
Male	77 (32.9)
Female	157 (67.1)
Age (years)	
18-29	160 (68.4)
30-39	74 (31.6)
Marital status	
Unmarried	195 (83.3)
Married	39 (16.7)
Education (years)	
12	97 (41.5)
14	4 (1.7)
16	133 (56.8)

Characteristics of Each MPIS Item

In Table 2, the mean range of each MPIS item was 2.07 to 4.01, and the SD range of each item was 0.87 to 1.39. For all items, the absolute values of skewness and kurtosis did not exceed 2 and 7, respectively. The result is provided in Multimedia Appendix 1. Thus, it was assumed that the data followed a normal distribution [43].

Table 2. Factor loading of the MPIS[a] (N=234).

Item number	Shelly (factor 1)	Flippy (factor 2)	Riggy (factor 3)	Pleaser (factor 4)	Jumpy (factor 5)	Mean (SD)
29	0.78	—[b]	—	—	—	2.52 (1.18)
35	0.77	—	—	—	—	2.43 (1.26)
19	0.74	—	—	—	—	2.66 (1.30)
25	0.71	—	—	—	—	2.91 (1.27)
10	0.70	—	—	—	—	2.21 (1.07)
17	0.70	—	—	—	—	2.71 (1.16)
1	0.66	—	—	—	—	2.49 (1.09)
2	0.64	—	—	—	—	2.81 (1.21)
24	0.54	—	—	—	—	2.64 (1.33)
11	0.36	—	—	—	—	2.94 (1.17)
16	—	0.86	—	—	—	2.65 (1.24)
9	—	0.79	—	—	—	2.59 (1.14)
23	—	0.79	—	—	—	2.53 (1.11)
5	—	0.72	—	—	—	2.20 (1.17)
28	—	0.65	—	—	—	2.46 (1.19)
31	—	0.59	—	—	—	2.07 (1.12)
33	—	—	0.78	—	—	3.29 (1.08)
27	—	—	0.68	—	—	3.67 (1.13)
14	—	—	0.67	—	—	2.94 (1.16)
7	—	—	0.67	—	—	4.01 (0.87)
22	—	—	0.66	—	—	3.60 (1.11)
21	—	—	—	0.80	—	2.53 (1.16)
26	—	—	—	0.64	—	2.85 (1.22)
32	—	—	—	0.64	—	3.19 (1.15)
13	—	—	—	0.55	—	2.43 (1.31)
18	—	—	—	0.54	—	2.41 (1.19)
6	—	—	—	0.04	—	3.41 (1.04)
15	—	—	—	—	0.73	3.26 (1.27)
4	—	—	—	—	0.72	2.50 (1.19)
8	—	—	—	—	0.71	3.18 (1.17)
34	—	—	—	—	0.71	3.44 (1.14)
12	—	—	—	—	0.70	3.12 (1.39)
3	—	—	—	—	0.61	2.50 (1.16)
30	—	—	—	—	0.57	3.07 (1.21)
20	—	—	—	—	0.53	2.51 (1.32)

[a]MPIS: Maladaptive Personality and Interpersonal Schema.

[b]—: not available.

Confirmatory Factor Analysis

To validate the MPIS, which was predefined, a CFA was conducted. The proposed model consisted of 5 factors. Meanwhile, an alternative model was a single-factor structure.

Regarding the results of the CFA for the single-factor model, the model fit indices are unacceptable (χ^2_{560}=2341.5; TLI=0.52; CFI=0.54; and RMSEA=0.11). However, for the results of the CFA for the 5-factor model, the model fit indices were acceptable, or slightly less than the good fit values

(χ^2_{550}=1278.1; TLI=0.80; CFI=0.81; and RMSEA=0.07). Therefore, the conclusion drawn was that the 5-factor model was more suitable than the single-factor model.

In Table 2, the factor loading for item 6 was 0.04, which is below 0.40. There is a need for consideration regarding the exclusion of item 6. With the exception of item 6, the revised 5-factor model displayed model fit indices that are acceptable, or slightly less than the criteria for a good fit (χ^2_{517}=1196.5; TLI=0.81; CFI=0.82; and RMSEA=0.07). This is not significantly different from the proposed model that includes item 6.

Additionally, information criteria were examined when comparing the proposed model and the revised 5-factor model. The proposed model yielded Akaike Information Criterion=22,890.95 and Bayesian Information Criterion=23,288.31, while the revised 5-factor model showed Akaike Information Criterion=22,201.73 and Bayesian Information Criterion=22,588.73, indicating minimal discrepancy between the 2 models. Given that item 6 encompassed essential content for the composition of the MPIS (a question regarding perceived interpersonal schemas in individuals with low self-esteem), it was retained as part of the final item set for the scale. The path model of 5-factor model is described in Multimedia Appendix 2.

Reliability

First, Cronbach α was .88 for factor 1, .87 for factor 2, .82 for factor 3, .71 for factor 4 (but increased to .77 when item 6 was excluded), and .86 for factor 5. Looking at the range of correlation coefficients between the subitems and the total scores by factor, it was 0.48-0.77 for factor 1, 0.70-0.85 for factor 2, 0.73-0.80 for factor 3, 0.59-0.78 for factor 4 (correlation coefficient between item 6 and factor 4 total score was 0.28), and 0.63-0.75 for factor 5.

Convergent and Discriminant Validity

In this study, PSS, SES, CES-D, and State-Trait Anxiety Inventory were used because the dysfunctional schema is associated with high stress levels, low self-efficacy, and high negative emotions such as depression and anxiety. Furthermore, the UCLA-LS, KIIP-SC, DTPD-D, STAXI, Urgency-Premeditation-Perseverance-Sensation seeking-Positive urgency Impulsive Behavior Scale-Negative Urgency, HMPS, SSES, RSQ, BAI, IUS, and PSWQ were used to verify the validity of the measure’s subfactors.

Tables 3 and 4 present the results of the analyses. First, the correlation between the measure’s subfactors was in the range of 0.17-0.63 (median 0.38; *P*<.001), indicating that the subfactors were appropriately differentiated. The total score of the MPIS was negatively correlated with SES, and positively correlated with PSS, CES-D, and State-Trait Anxiety Inventory.

Consequent to the correlation analysis of the subfactors and reference scale, Shelly was positively correlated with UCLA-LS, KIIP-FG, and DTPD-D. Flippy was positively correlated with STAXI anger-in, anger-out, anger-control, and Urgency-Premeditation-Perseverance-Sensation seeking-Positive urgency Impulsive Behavior Scale-Negative Urgency. Riggy was positively correlated with HMPS self-oriented perfectionism, other-oriented perfectionism, and socially prescribed perfectionism. Pleaser showed a negative correlation with SSES-social and a positive correlation with KIIP-HI, KIIP-LM, KIIP-NO, RSQ-rejection anxiety, and RSQ-rejection expectation. Jumpy showed a positive correlation with BAI, IUS, and PSWQ. These results demonstrate that the MPIS has good convergent and discriminant validity.

Table 3. Mean, SD, and Pearson correlation coefficients of factor-wise total scores for the MPIS[a] (N=234).

	Mean (SD)	Pearson correlation coefficient				
		Shelly (factor 1)	Flippy (factor 2)	Riggy (factor 3)	Pleaser (factor 4)	Jumpy (factor 5)
Shelly	26.54 (8.39)	—[b]	—	—	—	—
Flippy	14.49 (5.51)	0.38	—	—	—	—
Riggy	17.52 (4.12)	0.37	0.17	—	—	—
Pleaser	16.82 (4.56)	0.45	0.27	0.34	—	—
Jumpy	23.59 (7.07)	0.63	0.39	0.51	0.51	—

[a]MPIS: Maladaptive Personality and Interpersonal Schema.

[b]—: not available.

Table 4. Pearson correlation coefficients between the MPIS[a] and related assessment (N=234). A high score for SES[b] means high self-efficacy. A high score for SSES[c] means a higher state of self-esteem.

MPIS and related assessment	Pearson correlation coefficient
Total composite score	
PSS[d]	0.65
SES	–0.52
CES-D[e]	0.64
STAI[f]	0.74
Shelly	
UCLA-LS[g]	0.73
KIIP-FG[h]	0.65
DTPD-D[i]	0.46
Flippy	
STAXI[j] anger-in	0.41
STAXI anger-out	0.60
STAXI anger-control	0.45
UPPS-NU[k]	0.63
Riggy	
HMPS[l] self-oriented	0.60
HMPS other-oriented	0.21
HMPS socially prescribed	0.30
Pleaser	
SSES-social	–0.47
KIIP-HI[m]	0.52
KIIP-LM[n]	0.57
KIIP-NO[o]	0.31
RSQ[p]-rejection anxiety	0.36
RSQ-rejection expectation	0.18
Jumpy	
BAI[q]	0.55
IUS[r]	0.60
PSWQ[s]	0.79

[a]MPIS: Maladaptive Personality and Interpersonal Schema.
[b]SES: Self-Efficacy Scale.
[c]SSES: State Self-Esteem Scale.
[d]PSS: Perceived Stress Scale.
[e]CES-D: Center for Epidemiologic Studies Depression scale.
[f]STAI: State-Trait Anxiety Inventory.
[g]UCLA-LS: University of California - Los Angeles Loneliness Scale.
[h]KIIP-FG: Korea Inventory of Interpersonal Problems circumplex scale-socially inhibited or avoidant.
[i]DTPD-D: Diagnostic Test for Personality Disorders-Dependent.
[j]STAXI: State-Trait Anger Expression Inventory.
[k]UPPS-NU: Urgency-Premeditation-Perseverance-Sensation seeking-Positive urgency impulsive behavior scale-Negative Urgency.

[l]HMPS: Hewitt Multidimensional Perfectionism Scale.
[m]KIIP-HI: Korean Inventory of Interpersonal Problems nonassertive.
[n]KIIP-LM: Korean Inventory of Interpersonal Problems self-sacrificing or overly nurturant.
[o]KIIP-NO: Korean Inventory of Interpersonal Problems intrusive or needy.
[p]RSQ: Rejection Sensitivity Questionnaire.
[q]BAI: Beck Anxiety Inventory.
[r]IUS: Intolerance of Uncertainty Scale.
[s]PSWQ: Penn State Worry Questionnaire.

Discussion

Principal Findings

This study verified the reliability and validity of an internet-based self-report measure developed to assess MPIS. First, the result of the CFA indicating that the 5-factor model is more suitable than the single-factor model signifies that understanding maladaptive personality traits and interpersonal relationship patterns measured by the MPIS in 5 distinct types is justified. Particularly, the significance of these findings lies in the prospect of using the MPIS as a screening and classification measure for future digital mental health care services. This is enabled by deriving scores for each factor through the MPIS and thereby offering type-specific psychological therapy programs to individuals with high severity scores within each factor.

Next, item 6 had a low factor loading. While this might suggest a possibility of low validity of the item, it cannot be excluded that the interaction between item content and respondent characteristics also played a role. Those who excessively attribute importance to the interest or approval of others tend to adapt their behavior to gain such attention or recognition. Item 6, respondents were questioned about their efforts to satisfy the needs of others. However, even individuals excessively valuing the attention or recognition of others might not be fully aware of their behavioral patterns or could psychologically deny them. These possible causes might have contributed to the low factor loading observed for item 6.

Finally, based on the analysis of convergent and discriminant validity, the MPIS factors were all found to be appropriately distinct. Additionally, it was indicated that each factor adequately measures the targeted maladaptive personality and interpersonal relationship schema.

The implications of this study are as follows. First, a reliable and valid measure that can effectively identify an individual’s maladaptive personality traits was obtained. Previous measures like the PID-5 and YSQ-L3 identified 25 and 18 personality traits or psychological schemas, respectively, using more than 200 items [9,11]. Contrastingly, the MPIS uses only 35 items derived from the 5 core personality traits associated with individual maladjustment, which is more suited to digital mental health interventions as it allows participants to complete the questionnaire in a short time. Another benefit of the measure is that it is available on the internet. Currently, psychological measures conducted offline have not been easy for respondents to access; therefore, they are often underused to meet the needs of people who want to be accurately evaluated for their schema or who seek to establish therapeutic interventions based on the respondent's schema. However, on the internet, availability of these measures can improve usability and motivation to participate by increasing public accessibility. Finally, this measure may be useful for preventing the exacerbation of psychological problems.

Limitations

This study has some limitations. First, the majority of the sample comprised undergraduate and graduate students (133/234, 56.8%). Recruitment was conducted through community sites for office workers, stay-at-home spouses, and so on to diversify the participants; however, in the analysis, the student group still comprised the majority. Due to the nature of web-based surveys administered through platforms like mobile apps and computers, it seems that the higher accessibility for younger student age groups could be attributed to their ample personal usage time for smartphones or computers. In future research, validation could be conducted by considering the characteristics of the actual population group and implementing sampling to ensure diversity within the sample.

Second, the measures used may have underestimated some statistical results such as reliability due to an insufficient number of items corresponding to each maladaptive personality trait. In particular, some personality traits were measured using only 5 items, so the number of items per factor in the measure was disproportionate. It is difficult to detect maladaptive personality traits which are abstract psychological concepts using a small number of items. Therefore, effort is needed to add items without compromising the simplicity of the measure.

Third, item 6, which showed a low factor loading in the factor analysis, was determined to be appropriate and was not excluded from the measure, but future research should consider excluding item 6 or replacing it with another new item, thus considering the correlation of items by factor.

Fourth, individual personality traits were evaluated using only self-reported measures. There is no experimental approach to observe the actual behavioral patterns of individuals to analyze how much they correspond to the results of the maladaptive personality traits tests. In subsequent studies, it may be possible to construct scenarios or experimental situations related to anxiety, perfectionism, anger, low self-esteem, and isolation. This could help analyze whether maladaptive personality traits and interpersonal relationship schemas measured by the MPIS are correlated with descriptive responses or actual behaviors.

Fifth, in this study, only the single-factor model and the 5-factor model were compared using CFA, and a broader range of models was not incorporated into the analysis. Subsequent research could encompass a variety of models based on theoretical foundations beyond just the single-factor and 5-factor models.

Conclusions

This study has resulted in the development of a questionnaire capable of measuring 5 types of maladaptive personality traits and interpersonal relationship schemas. Through analysis, it was established that the 5 factors effectively measure psychological traits. The noteworthy significance of the MPIS lies in its web-based administration, providing excellent accessibility, cost-effectiveness, and featuring a reduced number of items compared to traditional psychological scales.

In the future, the MPIS could prove instrumental in offering digital mental health care services. Its web-based administration facilitates easy access, allowing for the convenient implementation of the MPIS and enabling the categorization of psychological therapy program content based on the results.

Acknowledgments

We would like to thank the researchers at 40FY Inc who participated in the development of the Maladaptive Personality and Interpersonal Schema (MPIS) scale used in this study. This study is funded by small- and medium-sized enterprises and Startups of Korean Government (S3200094).

Data Availability

The data sets analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

HK, SJ, and MSS contributed to conception and design of this study. SJ organized the database. HK performed the statistical analysis. SJ and HK wrote the first draft of this paper. IH, KS, and WM developed the Maladaptive Personality and Interpersonal Schema of this study. MSS contributed to this paper’s revision, and approved the submitted version.

Conflicts of Interest

IH, KS, and WM work for the company that developed the Maladaptive Personality and Interpersonal Schema.

Multimedia Appendix 1
Skewness and kurtosis of the Maladaptive Personality and Interpersonal Schema.
[DOCX File , 19 KB - ijmr_v12i1e48425_app1.docx]

Multimedia Appendix 2
The path model of the five-factor model.
[PDF File (Adobe PDF File), 116 KB - ijmr_v12i1e48425_app2.pdf]

References

1. Mohr DC, Tomasino KN, Lattie EG, Palac HL, Kwasny MJ, Weingardt K, et al. IntelliCare: an eclectic, skills-based app suite for the treatment of depression and anxiety. J Med Internet Res 2017;19(1):e10 [FREE Full text] [doi: 10.2196/jmir.6645] [Medline: 28057609]
2. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018;47(1):1-18 [FREE Full text] [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
3. Chen X, Zhang X, Zhu X, Wang G. Efficacy of an internet-based intervention for subclinical depression (MoodBox) in China: study protocol for a randomized controlled trial. Front Psychiatry 2020;11:585920 [FREE Full text] [doi: 10.3389/fpsyt.2020.585920] [Medline: 33510658]
4. Allemand M, Flückiger C. Personality change through digital-coaching interventions. Curr Dir Psychol Sci 2022;31(1):41-48 [FREE Full text] [doi: 10.1177/09637214211067782]
5. Dumitru VM, Cozman D. The relationship between stress and personality factors. Hum Vet Med 2012;4(1):34-39 [FREE Full text]
6. Afshar H, Roohafza HR, Keshteli AH, Mazaheri M, Feizi A, Adibi P. The association of personality traits and coping styles according to stress level. J Res Med Sci 2015;20(4):353-358 [FREE Full text] [Medline: 26109990]
7. American Psychiatric Association, DSM-5 Task Force. Diagnostic and Statistical Manual of Mental Disorders. 5th Edition. Washington, DC: American Psychiatric Association; 2013.
8. Bach B, Bernstein DP. Schema therapy conceptualization of personality functioning and traits in ICD-11 and DSM-5. Curr Opin Psychiatry 2019;32(1):38-49. [doi: 10.1097/YCO.0000000000000464] [Medline: 30299307]
9. Young JE, Brown G. Young schema-questionnaire. In: Cognitive Therapy for Personality Disorders: A Schema-focused Approach. Sarasota, FL: Professional Resource Press; 1994:63-76.

10. Young JE, Klosko JS, Weishaar ME. Schema Therapy: A Practitioner's Guide. New York: The Guilford Press; 2003.
11. Krueger RF, Derringer J, Markon KE, Watson D, Skodol AE. Initial construction of a maladaptive personality trait model and inventory for DSM-5. Psychol Med 2012;42(9):1879-1890 [FREE Full text] [doi: 10.1017/S0033291711002674] [Medline: 22153017]
12. Park WW, Mah SH, Bae SH, Ji SY, Lee YW, Kim JY. Careless responding in surveys: detection methods and the impact of screening on research. Korean Manag Rev 2020;49(2):331-364. [doi: 10.17287/kmr.2020.49.2.331]
13. Tak JK. Psychological Testing: An Understanding of Development and Evaluation Method. Seoul: Hakjisa Publisher; 2007.
14. Hoerger M. Participant dropout as a function of survey length in internet-mediated university studies: implications for study design and voluntary participation in psychological research. Cyberpsychol Behav Soc Netw 2010;13(6):697-700 [FREE Full text] [doi: 10.1089/cyber.2009.0445] [Medline: 21142995]
15. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983;24(4):385-396. [Medline: 6668417]
16. Park JO, Seo YS. Validation of the Perceived Stress Scale (PSS) on samples of Korean university students. Korean J Psychol Gen 2010;29(3):611-629.
17. Sherer M, Maddux JE, Mercandante B, Prentice-Dunn S, Jacobs B, Rogers RW. The self-efficacy scale: construction and validation. Psychol Rep 1982;51(2):663-671. [doi: 10.2466/pr0.1982.51.2.663]
18. Hong HY. The relationship of perfectionism, self-efficacy and depression. Ewha Womans University. Korea; 1995. URL: https://www.ewha.ac.kr/ewhaen/index.do [accessed 2023-09-14]
19. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Meas 1977;1(3):385-401. [doi: 10.1177/014662167700100306]
20. Chon KK, Choi SC, Yang BC. Integrated adaptation of CES-D in Korea. Korean J Health Psychol 2001;6(1):59-76.
21. Kim J, Shin D. A Study Based on the Standardization of the STAI for Korea. Newest Medical J 1978;21(11):69-75 [FREE Full text]
22. Russell D, Peplau LA, Cutrona CE. The revised UCLA loneliness scale: concurrent and discriminant validity evidence. J Pers Soc Psychol 1980;39(3):472-480. [doi: 10.1037//0022-3514.39.3.472] [Medline: 7431205]
23. Lee S. The moderating effect of early maladaptive schema in the relation between loneliness and interpersonal problems. Sookmyung Women's University. Korea; 2012. URL: https://e.sookmyung.ac.kr/en/etc/search.do [accessed 2023-09-14]
24. Hong S, Park E, Kim Y, Kwon J, Cho Y, Kim Y. Short form of the Korean Inventory of Interpersonal Problems Circumplex Scales (KIIP-SC). Korean J Health Psychol 2002;21(4):923-940 [FREE Full text]
25. Seo JS, Hwang ST. Development and validation of the Korean personality disorders test. Korean J Clin Psychol 2006;25(1):273-288 [FREE Full text]
26. Spielberger CD. State-Trait Anger Expression Inventory: Professional Manual. Odessa, FL: Psychological Assessment Resources; 1988.
27. Han D, Lee J, Jeon G. The Korean version of State-Trait Anger Expression Inventory (STAXI-K). Korean J Health Psychol 1998;3(1):18-32.
28. Whiteside SP, Lynam DR. The five factor model and impulsivity: using a structural model of personality to understand impulsivity. Pers Individ Differ 2001;30(4):669-689. [doi: 10.1016/S0191-8869(00)00064-7]
29. Cyders MA, Smith GT, Spillane NS, Fischer S, Annus AM, Peterson C. Integration of impulsivity and positive mood to predict risky behavior: development and validation of a measure of positive urgency. Psychol Assess 2007;19(1):107-118. [doi: 10.1037/1040-3590.19.1.107] [Medline: 17371126]
30. Lim SY, Lee YH. A Korean validation of the UPPS-P impulsive behavior scale in college students. Korean J Clin Psychol 2014;33(1):51-71 [FREE Full text] [doi: 10.15842/kjcp.2014.33.1.004]
31. Hewitt PL, Flett GL. Perfectionism in the self and social contexts: conceptualization, assessment, and association with psychopathology. J Pers Soc Psychol 1991;60(3):456-470. [doi: 10.1037//0022-3514.60.3.456] [Medline: 2027080]
32. Heatherton TF, Polivy J. Development and validation of a scale for measuring state self-esteem. J Pers Soc Psychol 1991;60(6):895-910. [doi: 10.1037/0022-3514.60.6.895]
33. Park H, Lee J. A validation study of the Korean version of state self-esteem scale. Korean J Psychol Gen 2015;34(1):133-158.
34. Downey G, Feldman SI. Implications of rejection sensitivity for intimate relationships. J Pers Soc Psychol 1996;70(6):1327-1343. [doi: 10.1037//0022-3514.70.6.1327] [Medline: 8667172]
35. Lee B. Adult attachment and heterosexual relationship satisfaction: focusing on the mediating effect of rejection sensitivity and attribution style. Korea University. Korea; 2000. URL: https://www.korea.edu/mbshome/mbs/en/index.do [accessed 2023-09-14]
36. Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. J Consult Clin Psychol 1988;56(6):893-897. [doi: 10.1037//0022-006x.56.6.893] [Medline: 3204199]
37. Kwon S. Differential roles of dysfunctional attitudes and automatic thoughts in depression: an integrated cognitive model of depression. University of Queensland. Australia; 1992. URL: https://www.uq.edu.au/ [accessed 2023-09-14]
38. Freeston MH, Rhéaume J, Letarte H, Dugas MJ, Ladouceur R. Why do people worry? Pers Individ Differ 1994;17(6):791-802. [doi: 10.1016/0191-8869(94)90048-5]

39. Carleton RN, Norton MAPJ, Asmundson GJG. Fearing the unknown: a short version of the intolerance of uncertainty scale. J Anxiety Disord 2007;21(1):105-117. [doi: 10.1016/j.janxdis.2006.03.014] [Medline: 16647833]
40. Kim S. The relationship of fear of negative and positive evaluation, intolerance of uncertainty, and social anxiety. Ewha Womans University. Korea; 2010. URL: https://www.ewha.ac.kr/ewhaen/index.do [accessed 2023-09-14]
41. Meyer TJ, Miller ML, Metzger RL, Borkovec TD. Development and validation of the Penn State Worry Questionnaire. Behav Res Ther 1990;28(6):487-495. [doi: 10.1016/0005-7967(90)90135-6] [Medline: 2076086]
42. Kim J, Min B. Relationship between worry, intolerance of uncertainty, and problem-solving style. 1998 Presented at: the Annual Meeting for the Korean Psychological Association, Korea; October 1998; Seoul, Korea.
43. Curran PJ, West SG, Finch JF. The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. Psychol Methods 1996;1(1):16-29. [doi: 10.1037/1082-989x.1.1.16]

Abbreviations

BAI: Beck Anxiety Inventory
CES-D: Center for Epidemiologic Studies Depression Scale
CFA: confirmatory factor analysis
CFI: comparative fit index
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, 5th edition
DTPD-D: Diagnostic Test for Personality Disorders-Dependent
HMPS: Hewitt Multidimensional Perfectionism Scale
IUS: Intolerance of Uncertainty Scale
KIIP-FG: Korean Inventory of Interpersonal Problems socially inhibited or avoidant
KIIP-HI: Korean Inventory of Interpersonal Problems nonassertive
KIIP-LM: Korean Inventory of Interpersonal Problems self-sacrificing or overly nurturant
KIIP-NO: Korean Inventory of Interpersonal Problems intrusive or needy
KIIP-SC: Korean Inventory of Interpersonal Problems Circumplex Scale
MPIS: Maladaptive Personality and Interpersonal Schema
PID-5: Personality Inventory for Diagnostic and Statistical Manual of Mental Disorders, 5th edition
PSS: Perceived Stress Scale
PSWQ: Penn State Worry Questionnaire
RMSEA: Root Mean Square Error of Approximation
RSQ: Rejection Sensitivity Questionnaire
SES: Self-Efficacy Scale
SSES: State Self-Esteem Scale
STAXI: State-Trait Anger Expression Inventory
TLI: Tucker-Lewis index
UCLA-LS: University of California - Los Angeles Loneliness Scale
YSQ-L3: Young Schema Questionnaire-Long form-3

Edited by T de Azevedo Cardoso; submitted 23.04.23; peer-reviewed by M Freestone, R Meng; comments to author 04.08.23; revised version received 24.08.23; accepted 04.09.23; published 29.09.23.

Please cite as:
Kim H, Jeong S, Hwang I, Sung K, Moon W, Shin MS
Validation of a Brief Internet-Based Self-Report Measure of Maladaptive Personality and Interpersonal Schema: Confirmatory Factor Analysis
Interact J Med Res 2023;12:e48425
URL: https://www.i-jmr.org/2023/1/e48425
doi:10.2196/48425
PMID:37773606

Original Paper

The Role of Daily Steps in the Treatment of Major Depressive Disorder: Secondary Analysis of a Randomized Controlled Trial of a 6-Month Internet-Based, Mindfulness-Based Cognitive Behavioral Therapy Intervention for Youth

Kevin Dang[1], MSc; Paul Ritvo[1,2], PhD; Joel Katz[1,2], PhD; David Gratzer[3,4], MD; Yuliya Knyahnytska[3,4], MD, PhD; Abigail Ortiz[3,4], MD; Clarice Walters[3], BSc; Mohamed Attia[3,4], MD; Christina Gonzalez-Torres[3,4], MD; Andrew Lustig[3,4], MD; Zafiris Daskalakis[3,4,5], MD, PhD

[1]School of Kinesiology and Health Sciences, York University, Toronto, ON, Canada
[2]Department of Psychology, York University, Toronto, ON, Canada
[3]Centre for Addiction and Mental Health, Toronto, ON, Canada
[4]Department of Psychiatry, University of Toronto, Toronto, ON, Canada
[5]Department of Psychiatry, University of California San Diego, San Diego, CA, United States

Corresponding Author:
Paul Ritvo, PhD
School of Kinesiology and Health Sciences
York University
4700 Keele St
Toronto, ON, M3J 1P3
Canada
Phone: 1 416 580 8021
Email: pritvo@yorku.ca

Abstract

Background: Current evidence supports physical activity (PA) as an adjunctive treatment for major depressive disorder (MDD). Few studies, however, have examined the relationship between objectively measured PA and MDD treatment outcomes using prospective data.

Objective: This study is a secondary analysis of data from a 24-week internet-based, mindfulness-based cognitive behavioral therapy program for MDD. The purpose of this analysis was twofold: (1) to examine average daily step counts in relation to MDD symptom improvement, and whether pain moderated this relationship; and (2) to examine whether changes in step activity (ie, step trajectories) during treatment were associated with baseline symptoms and symptom improvement.

Methods: Patients from the Centre for Addiction and Mental Health were part of a randomized controlled trial evaluating the effects of internet-based, mindfulness-based cognitive behavioral therapy for young adults (aged 18-30 years old) with MDD. Data from 20 participants who had completed the intervention were analyzed. PA, in the form of objectively measured steps, was measured using the Fitbit-HR Charge 2 (Fitbit Inc), and self-reported depression severity was measured with the Beck Depression Inventory-II (BDI-II). Linear regression analysis was used to test PA's relationship with depression improvement and the moderating effect of pain severity and pain interference. Growth curve and multivariable regression models were used to test longitudinal associations.

Results: Participants walked an average of 8269 steps per day, and each additional +1000-step difference between participants was significantly associated with a 2.66-point greater improvement (reduction) in BDI-II, controlling for anxiety, pain interference, and adherence to Fitbit monitoring (P=.02). Pain severity appeared to moderate (reduce) the positive effect of average daily steps on BDI-II improvement (P=.03). Higher baseline depression and anxiety symptoms predicted less positive step trajectories throughout treatment (Ps≤.001), and more positive step trajectories early in the trial predicted greater MDD improvement at the end of the trial (Ps<.04). However, step trajectories across the full duration of the trial did not significantly predict MDD improvement (Ps=.40).

Conclusions: This study used objective measurements to demonstrate positive associations between PA and depression improvement in the context of cognitive behavioral treatment. Pain appeared to moderate this relationship, and baseline symptoms of anxiety and depression predicted PA trajectories. The findings inform future interventions for major depression. Future research with larger samples should consider additional moderators of PA-related treatment success and the extent to which outcomes are related to PA change in multimodal interventions.

Trial Registration: Clinical Trials.gov NCT03406052; https://www.clinicaltrials.gov/ct2/show/NCT03406052

International Registered Report Identifier (IRRID): RR2-10.2196/11591

(Interact J Med Res 2023;12:e46419) doi:10.2196/46419

KEYWORDS

accelerometer; anxiety; CBT; chronic pain; cognitive behavioral therapy; controlled trials; depression; depressive symptoms; digital health; eHealth; exercise; fitbit; intervention study; longitudinal study; major depressive disorder; mHealth; mindfulness; mindfulness-based CBT; objectively measured activity; online health; online intervention; online therapy; pain; physical activity; prospective study; randomized controlled trial; RCT; step; steps

Introduction

Background

Depression is now the leading cause of global disability, severely impacting daily functioning and quality of life for 322 million people worldwide [1]. A significant subset of patients do not respond to antidepressants [2,3] or psychotherapy [4], and about 40% of patients experience relapse following psychotherapeutic treatment [5]. Given the burden of depression, there is a growing emphasis on developing cost-effective, complementary interventions that improve treatment response and reduce relapse risk.

Current evidence supports increased physical activity (PA) and its exercise subsets [6,7] as an adjunctive treatment for major depressive disorder (MDD) or as a single-modality treatment [8-12] with small to moderate antidepressant effects [13,14] and protective benefits against future depression [15]. Despite promising findings, significant research gaps remain.

First, studies investigating the antidepressant effects of PA in patients with MDD have generally relied on self-reported measures [16], which are prone to cognitive biases [17,18] and discrepant results when compared to objectively measured PA [19,20]. Such biases are important to rule out in MDD samples, who often present cognitive impairments [21,22] that can limit the accuracy of self-reported data. Vancampfort et al [23], for example, suggested that individuals with severe mental illness may underestimate their levels of sedentary PA while overestimating vigorous activity, which is consistent with other meta-analytic findings focused on MDD samples [16]. Moreover, Choi et al [19] found that PA was predictive of lower MDD risk when objectively measured but not when subjectively measured. An additional limitation is that many studies that investigated the relationship between depression and objectively measured PA [24] used brief monitoring periods (often 7 days) and samples with specific health conditions that limit the generalizability of findings to patients with MDD.

Another research gap is the lack of data on whether PA's benefits in MDD samples are influenced by chronic pain. This empirical gap is surprising given the high comorbidity prevalence between pain and depression [25] and findings of pain's variable effects on PA-treatment outcomes (eg, depression and mental health improvement) [26]. Specifically, PA in pain-free populations often results in reduced pain sensitivity, whereas the relationship is variable in chronic pain populations, with unchanged pain levels in some patients and worsened pain sensitivity in others [27,28]. Thus, the question of whether pain modifies the relationship between PA and depression improvement requires further investigation. Indeed, the heterogeneity of depression has led to calls for more research on moderators of PA's antidepressant effects [29,30].

A further gap in the empirical literature is that prospective studies typically examine the effect of PA on depressive symptoms [31] without focusing on questions of reverse causality, such as whether baseline depression and anxiety levels predict longitudinal PA patterns [32]. Some studies have found that higher baseline symptoms predict reduced PA over time [33-38], while other studies have not supported this reverse relationship [39,40]. Furthermore, these studies were limited by a reliance on self-reported PA.

Although psychosocial interventions often improve depressive symptoms [41], remarkably little is known about whether positive changes in PA predict such treatment-related improvements [42]. A study found increased PA levels in patients with MDD who responded to repeated transcranial magnetic stimulation [43], and another study found similar associations in patients with cardiovascular disease treated with cognitive behavioral therapy (CBT) [44]. Yet, no previous study has investigated whether objectively measured increases in PA predict MDD symptom reductions in the context of a multimodal mindfulness-based cognitive behavioral therapy (MCBT) intervention for clinical depression. Moreover, it is unclear whether early changes in PA predict future outcomes, despite the clinical importance of identifying early predictors of depression treatment response [45,46], and several studies showing that early increases in behavioral activation predict later symptom improvement [47,48].

Purpose

This study examined whether daily step counts were associated with MDD symptom improvement over the course of a 24-week internet-based, mindfulness-based cognitive behavioral therapy (iMCBT) intervention for young adults (18-30 years old), controlling for potential confounders (adherence [49,50], baseline anxiety [51,52], and baseline pain [26,53] were selected as covariates given their established relationships with PA and depression improvement). We also examined the interaction effect between steps and baseline pain (severity and interference) on MDD symptom improvement.

A further aim was to examine whether baseline depression and anxiety symptoms were predictive of step trajectories across the trial, and whether step trajectories were associated with MDD symptom improvement. Understanding patterns and moderators of objectively measured PA in relation to treatment outcomes could improve the recommendations for future patients with MDD receiving internet-based cognitive behavioral treatment.

Hypotheses

Hypothesis 1: Higher average step levels (daily steps) are positively associated with MDD symptom improvement.

Hypothesis 2: Pain severity and pain interference moderate the positive relationship between daily steps and MDD symptom improvement.

Hypothesis 3: Baseline depression and anxiety levels predict week-to-week changes in PA (ie, step trajectories) over the course of iMCBT.

Hypothesis 4: Early step increases and step trajectories across the full length of the iMCBT intervention are positively associated with MDD symptom improvement.

Methods

Design and Participants

This secondary analysis uses participant data from the intervention arm of a parallel, 2-arm randomized controlled trial (RCT) comparing iMCBT plus standard psychiatric care (intervention) with standard psychiatric care alone (waitlist control). This analysis did not include the control arm because PA monitoring was part of the intervention arm only. The RCT evaluated the efficacy of an internet- and telephone-based iMCBT intervention for young adults (aged 18-30 years) diagnosed with major depressive disorder. The RCT design [54] and RCT results [55] have been described elsewhere.

All participants were diagnosed by a Centre for Addiction and Mental Health (CAMH) psychiatrist, with diagnoses confirmed through a Mini-International Neuropsychiatric Interview (MINI) administered at the screening visit. Participants were identified from CAMH service waitlists by research coordinators and in the prescreening of new clinic referrals. All self-report measures and clinical interviews were conducted at the same CAMH Ambulatory Service setting. The inclusion and exclusion criteria for this study are listed in Textbox 1.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- adults aged between 18 and 30 years
- a Beck Depression Inventory-II (BDI-II) score of at least mild severity with no upper limit (BDI-II score ≥14 [56])
- a Mini-International Neuropsychiatric Interview–confirmed psychiatric diagnosis of major depressive disorder [57]
- English fluency

Exclusion criteria

- currently receiving weekly structured psychotherapy
- meets the DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition) criteria for severe alcohol or substance use disorder in the past 3 months
- demonstrates clinically significant suicidal ideation, defined as imminent intent
- attempted suicide in the past 6 months
- diagnosed with borderline personality disorder, bipolar disorder, schizophrenia, or obsessive-compulsive disorder

Intervention

All RCT participants received standard psychiatric care, defined as monthly treatment-as-usual sessions with a CAMH psychiatrist focused primarily on medication adjustment. Participants in the iMCBT intervention also received a Fitbit-HR Charge 2 (Fitbit Inc) and access to NexJ Connected Wellness (NexJ Health Inc), a cloud-based digital health platform accessible through smartphones and internet-connected devices. Participants were instructed to wear their Fitbit monitors 24 hours daily. Fitbit-tracked daily step counts were automatically uploaded to the NexJ Connected Wellness (NexJ Health Inc) platform, allowing participants and Health Navigator-Coaches to review daily step activity. The purpose of step monitoring was to reinforce iMCBT concepts by providing participants real-time feedback about how the behaviors they modify link to cognitive-affective changes.

The platform also enabled text-message communications between participants and Health Navigator-Coaches, and access to iMCBT content delivered through 24 workbooks and 56 instructional videos reflecting CBT and mindfulness principles. Intervention participants additionally received weekly phone support from Health Navigator-Coaches to facilitate behavior change, the application of iMCBT content, and participant adherence. The workbook content was built on previous web-based MCBT RCTs [58,59] and included a spectrum of MDD-targeted topics such as Living by Your Truths, Overcoming Wired-ness and Tired-ness, Mindfulness and Relationships, Loss and Grief, Resilience, Befriending Ourselves, Befriending Your Body With Exercise, Body Image and Mindfulness, Intimacy, Forgiveness, Overcoming Procrastination, Dealing With Negative Moods, Stress Resilience, Overcoming Performance Anxiety, and Cultivating Inspiration.

Measures

Depression

Depression symptoms at baseline and at the end of the 24-week intervention were measured using the Beck Depression Inventory-II (BDI-II) [60]. The BDI-II is a 21-item self-report inventory for assessing depression symptom severity; each item is measured on a 4-point Likert scale (0 to 3). Higher scores indicate higher levels of depression. The maximum total score is 63. Generally accepted categorical classes are "minimal depression" (0-13), "mild depression" (14-19), "moderate depression" (20-28), and "severe depression" (29-63). The BDI-II has demonstrated acceptable test-retest reliability, convergent and divergent validity [61], and in this sample, acceptable internal consistency (Cronbach $\alpha_{baseline}$=.83, Cronbach $\alpha_{follow\text{-}up}$=.90).

Anxiety

Baseline anxiety symptoms were measured using the Beck Anxiety Inventory (BAI) [62]. The BAI is a 21-item self-report scale for assessing anxiety severity; each item is measured on a 4-point Likert scale (0 to 3). Higher scores indicate higher levels of anxiety. The maximum total score is 63. The clinical classifications of scored results are "minimal anxiety" (0-7), "mild anxiety" (8-15), "moderate anxiety" (16-25), and "severe anxiety" (26-63). The BAI has demonstrated acceptable test-retest reliability [63], convergent validity [63], and divergent validity [64], and in this sample, acceptable internal consistency (Cronbach α=.78).

Pain

Baseline pain dimensions (severity and interference) were measured using the Brief Pain Inventory (BPI) subscales [65]. The BPI is a self-reported measure that gives 2 composite scores representing the severity of pain (Brief Pain Inventory: Severity [BPI-sev]) and the degree to which pain interferes with daily activities (Brief Pain Inventory: Interference [BPI-intf]). The BPI-sev and BPI-intf subscales consist of 4 and 7 items, respectively, and are scored as the average of their respective items. Each item is rated from 0 to 10, with higher ratings indicating worse severity or interference. The maximum composite score for each subscale is 10. The BPI subscales have demonstrated acceptable test-retest reliability and construct validity [66], and in this sample, acceptable internal consistency (Cronbach $\alpha_{BPI\text{-}sev}$=.67, Cronbach $\alpha_{BPI\text{-}intf}$=.94).

Daily Step Count

Physical steps throughout the trial were measured and automatically synchronized with the NexJ Connected Wellness (NexJ Health Inc) platform using the Fitbit-HR Charge 2 (Fitbit Inc), which has demonstrated accurate step estimates in free-living conditions [67-69]. Step counts are a well-established measure of PA, with evidence generally indicating that more is better [70] up to around 8000–10,000 steps/day for adults aged 60 years or younger [71]. Day-to-day step records were obtained from the NexJ Connected Wellness (NexJ Health Inc) platform.

Statistical Analyses

BDI-II change was operationalized as BDI-II at baseline minus BDI-II at postintervention; higher BDI-II change scores represent greater reductions in depression symptoms (ie, improvement). Daily steps were operationalized as and adherence to behavioral monitoring was operationalized as .

Step records with fewer than 100 steps were assumed invalid and excluded from the analyses.

Linear regression assumptions were evaluated using graphical methods. Linearity and homoscedasticity were checked using residual plots, and normality was checked using QQ-plots and histograms of model residuals. Hat values, Studentized residuals, and Cook's distances were used to verify the lack of problematic outliers.

Hypothesis 1: Daily Steps Are Positively Associated With MDD Symptom Reduction

Linear regression models were used to examine the effect of daily steps (focal predictor) on BDI-II improvement (outcome), controlling for baseline adherence, anxiety, and pain (severity and interference).

Hypothesis 2: Pain Moderates the Relationship Between Steps and BDI-II Improvement

Given the potential for pain to alter the antidepressant effects of daily PA, as well as the observed bimodality in the distribution of BPI subscales (suggestive of 2 clinical subgroups; Figure S1 in Multimedia Appendix 1), we investigated whether pain and daily steps interact to influence BDI-II improvement.

Linear regression was used to evaluate the interaction effects of daily steps × pain interference and daily steps × pain severity on BDI-II improvement.

Hypothesis 3: Baseline Depression and Anxiety Levels Predict Week-to-Week Step Trajectories Over the Course of iMCBT

We used 2-level linear conditional growth curve models to test whether baseline characteristics (depression and anxiety) predicted week-by-week step trajectories. To facilitate longitudinal analyses for this hypothesis (and for Hypothesis 4), each participant's day-to-day steps were collapsed into weekly averages; for example:

Weekly averages (level 1) were nested within individuals (level 2), and baseline characteristics were treated as level 2 variables. Linear models were considered adequate based on inspection of individual trajectories. Slopes and intercepts were allowed to vary randomly across participants to model change trajectories at the individual level. A first-order autoregressive covariance structure was fit to account for the autocorrelation inherent in longitudinal data. Parameters were estimated using restricted maximum likelihood.

Hypothesis 4: Early Step Trajectories (Slopes) and Trajectories Across the Trial Are Positively Associated With Greater BDI-II Improvement

As "early change" has not been consensually operationalized in the psychotherapeutic literature, we defined the first 6 weeks (the first quarter of the iMCBT intervention) as the period of early PA change. To assess whether early step trajectories (slopes) predict later MDD symptom improvement, we first estimated each individual's slope ( ) and intercept (where the slope crosses the y-axis) using week-by-week step averages for the first 6 weeks; the intercept represents the estimated step level at week 0 (ie, at baseline). We then regressed BDI-II change (outcome) on estimated slopes (focal predictor), controlling for intercepts, anxiety, and pain (severity and interference). Intercepts (estimated baseline step levels) were included as a covariate, as past PA behavior is known to predict subsequent PA [72,73] and the development of depressive symptoms [15,31].

To assess whether positive trajectories across the trial were associated with greater MDD symptom reduction, we repeated this procedure using week-by-week step averages for the full duration of the trial. Participants with less than 50% adherence were excluded from the full-length trajectory analysis, as large quantities of missing data can distort slopes and misrepresent change trajectories.

Statistical Software

Data cleaning and preparation were carried out using Python (Python Software Foundation). R version 4.1.3 (R Core Team) along with the *nlme* package [74] were used to carry out statistical analyses and mixed-effects modeling. Hypotheses were tested using 2-tailed t tests. The significance threshold across analyses was P=.05.

Sample Size Considerations

This study is a secondary analysis of an RCT [55] that originally aimed to enroll 168 participants, with 50% of the participants from a First Nations background and the other 50% from all other ethnic backgrounds, stratified into 2 intervention groups and 2 waitlist control groups (ie, n=42 per group). However, participant enrollment was reduced due to the reluctance of individuals from First Nations backgrounds to participate, despite extensive recruitment efforts. Additionally, given that each recruited participant had to undergo an extensive psychiatric exam to establish a MDD diagnosis, we confronted a limit to the pace of psychiatric examinations that could be scheduled given the existing staff of psychiatrists.

As this is a secondary analysis, no formal power calculations were performed. This approach adheres to the International Council for Harmonisation E9 statistical principles for clinical trials, which state that sample size should be "determined by the primary objective of the trial" [75].

Ethical Considerations

Ethics approval was obtained from the Research and Ethics Boards of the Centre for Addiction and Mental Health (protocol 115/2016-01) and York University (certificate 2017–154) in Toronto, Ontario, Canada (ClinicalTrials.gov ID NCT03406052). All participants provided in-person written consent for the use of their data in primary and secondary analyses. Participant confidentiality was maintained throughout the study through careful deidentification of data.

Results

Overview

A total of 22 participants recruited from February 2018 to September 2018 were enrolled in the iMCBT intervention. Of the 22 iMCBT participants, a total of 2 withdrew due to stressful life events. This secondary analysis is based on the 20 participants who completed the iMCBT intervention. Multimedia Appendix 2 summarizes the flow of participants, and Table 1 shows the sample characteristics for the 20 participants who completed the iMCBT intervention.

Table 1. Demographic, psychological, and intervention variables (N=20).

Variable	Value
Age (years), mean (SD)	24.75 (3.48)
Gender, n (%)	
Male	9 (45)
Female	11 (55)
Ethnicity, n (%)	
Caucasian	12 (60)
Asian	6 (30)
Indo-Caribbean	1 (5)
Mixed	1 (5)
Education level, n (%)	
High school	2 (10)
Completing college	3 (15)
College	1 (5)
Completing university	5 (25)
University	8 (40)
Graduate or professional school	1 (5)
Work status, n (%)	
Employed	11 (55)
Unemployed	9 (45)
Depression duration, mean (SD)	
Age of depression onset (years)	17.20 (4.10)
Number of depressive episodes in lifetime (n=19)	5.16 (5.10)
Duration of current depressive episode (months)	8.88 (15.30)
Psychological variables at baseline, mean (SD)	
BDI-II[a]	29.20 (8.22)
BAI[b]	28.15 (8.22)
BPI-sev[c]	1.69 (1.46)
BPI-intf[d]	1.99 (2.55)
Outcome and intervention variables, mean (SD)	
BDI-II improvement[e]	15.60 (9.75)
Daily steps[f]	8263 (2842)
Total steps	1,196,806 (678,338)
Adherence[g] (percentage), median (IQR)	84.35 (70.08-98.36)

[a]BDI-II: Beck Depression Inventory-II.
[b]BAI: Beck Anxiety Inventory.
[c]BPI-sev: Brief Pain Inventory: Severity.
[d]BPI-intf: Brief Pain Inventory: Interference.
[e]Pretreatment BDI-II minus posttreatment BDI-II.
[f]Total steps recorded divided by number of Fitbit-tracked days.
[g]Number of Fitbit-tracked days divided by number of days in intervention period.

Hypothesis 1: Daily Steps Are Positively Associated With MDD Symptom Reduction

Multivariable regression was used to test the relationship between daily steps and BDI-II reduction, controlling for potential confounders. The daily steps variable was rescaled by a divisor of 1000 to improve the interpretability of parameter estimates. Results are presented in Table 2.

Per Model A, there was evidence of a nonsignificant positive relationship between daily steps and BDI-II improvement, controlling for adherence, pain severity, and anxiety (b=1.95; SE=0.98; t_{15}=1.99; P=.06). In Model B, where we controlled for pain interference instead of pain severity, there was a significant positive effect of daily steps on BDI-II improvement (b=2.66; SE=1.00; t_{15}=2.65; P=.02).

Table 2. Regression of Beck Depression Inventory-II (BDI-II) change on steps, adherence, and comorbidities (N=20).

Variable	*b*	SE	*t* (*df*)	*P* value	95% CI	R^2
Model A						.42[a]
(Intercept)	–44.08	21.23	–2.08 (15)	.06	–89.32 to 1.17	
Daily steps (per 1000)	1.95	0.98	1.99 (15)	.06	–0.13 to 4.03	
Adherence	0.25	0.10	2.60 (15)	.02	0.04 to 0.45	
BPI-sev[b]	3.94	1.68	2.35 (15)	.03	0.36 to 7.52	
BAI[c]	0.62	0.36	1.72 (15)	.11	–0.15 to 1.40	
Model B						.49[d]
(Intercept)	–43.76	19.51	–2.24 (15)	.04	–85.34 to –2.17	
Daily steps (per 1000)	2.66	1.00	2.65 (15)	.02	0.52 to 4.79	
Adherence	0.15	0.09	1.72 (15)	.11	–0.04 to 0.33	
BPI-intf[e]	2.55	0.88	2.91 (15)	.01	0.68 to 4.42	
BAI	0.74	0.34	2.17 (15)	.05	0.01 to 1.46	

[a]$F_{4,15}$=2.68; P=.07.

[b]BPI-sev: Brief Pain Inventory: Severity.

[c]BAI: Beck Anxiety Inventory.

[d]$F_{4,15}$=3.61; P=.03.

[e]BPI-intf: Brief Pain Inventory: Interference.

Hypothesis 2: Pain Moderates the Relationship Between Steps and BDI-II Improvement

There was evidence of a significant interaction between daily steps and BPI-sev: $b_{(\text{Steps} \times \text{BPI-sev})}$=–1.45; SE=0.62; t_{16}=–2.36; P=.03. The interaction effect explained an additional 22.21% of the variance in BDI-II improvement (ΔR^2=.22) over the variance explained by the additive effects of daily steps and BPI-sev (for the additive model, see Table S1 in Multimedia Appendix 1).

By contrast, the interaction between daily steps and BPI-intf was not significant: $b_{(\text{Steps} \times \text{BPI-intf})}$=–0.61; SE=0.38; t_{16}=–1.59; P=.13. The interaction effect explained an additional 9.45% of the variance in BDI-II improvement (ΔR^2=.09) over the variance explained by the additive effects of daily steps and BPI-intf (for the additive model, see Table S1 in Multimedia Appendix 1).

Table 3 summarizes the regression results, Figure 1 depicts the interaction, and Table 4 describes the conditional effect of daily steps on BDI-II improvement at low (25th percentile) and high (75th percentile) levels of BPI-sev and BPI-intf.

Overall, the effect of daily steps on MDD symptom improvement was positive at low levels of pain, but this effect tended to decrease as pain increased. Pain therefore appears to modify the antidepressant effect of steps.

Table 3. Regression of Beck Depression Inventory-II (BDI-II) improvement on steps, pain, and their interaction (N=20).

Variable	*b*	SE	*t* (*df*)	*P* value	95% CI	R^2
Model A						.36[a]
(Intercept)	–13.52	10.73	–1.26 (16)	.23	–36.26 to 9.22	
Daily steps (per 1000)	2.80	1.08	2.60 (16)	.02	0.52 to 5.09	
BPI-sev[b]	13.70	4.84	2.83 (16)	.01	3.44 to 23.96	
Daily steps × BPI-sev	–1.45	0.62	–2.36 (16)	.03	–2.76 to –0.15	
Model B						.41[c]
(Intercept)	–7.88	8.59	–0.92 (16)	.37	–26.08 to 10.32	
Daily steps (per 1000)	2.20	0.88	2.50 (16)	.02	0.34 to 4.06	
BPI-intf[d]	6.51	2.64	2.47 (16)	.03	0.92 to 12.11	
Daily steps × BPI-intf	–0.61	0.38	–1.59 (16)	.13	–1.43 to 0.20	

[a]$F_{3,16}$=3.05; *P*=.06.

[b]BPI-sev: Brief Pain Inventory: Severity.

[c]$F_{3,16}$=3.63; *P*=.04.

[d]BPI-intf: Brief Pain Inventory: Interference.

Figure 1. Visualization of the effect of steps on Beck Depression Inventory-II (BDI-II) improvement at high and low levels of pain severity and pain interference (N=20). At low levels of pain, more physical activity in the form of daily steps was associated with greater symptom improvement, whereas this positive association tended to weaken as pain levels increased. BPI-intf: Brief Pain Inventory: Interference; BPI-sev: Brief Pain Inventory: Severity.

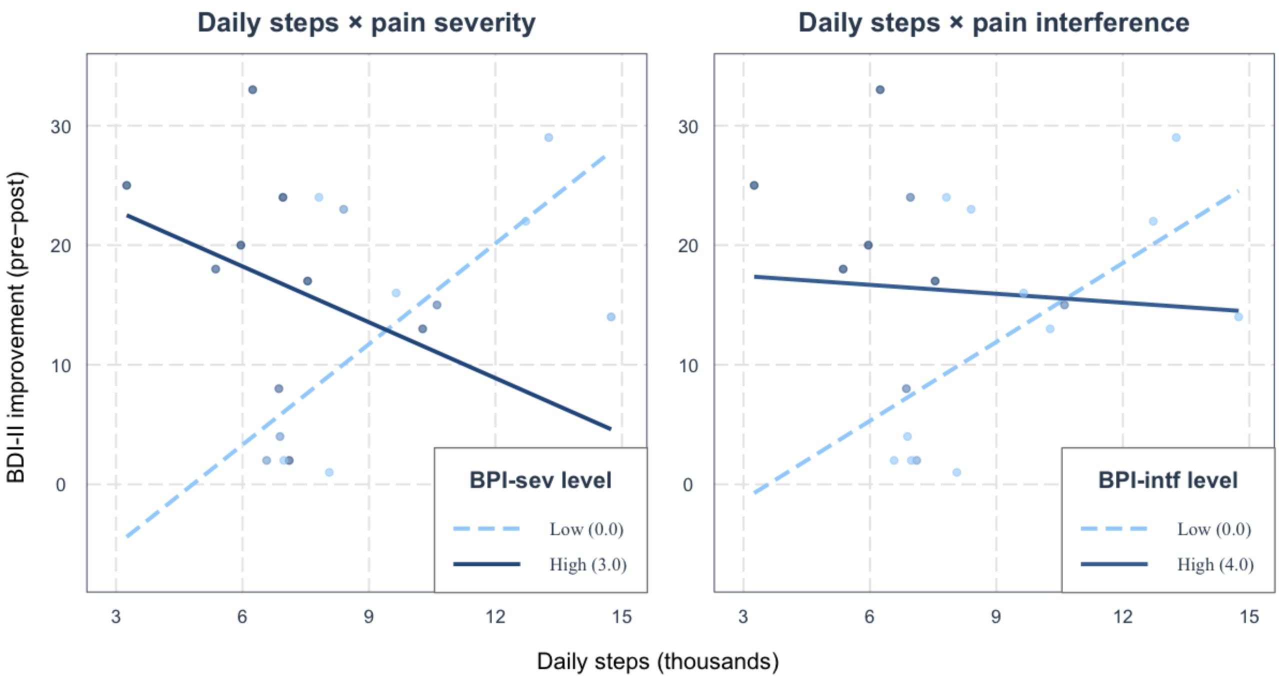


Table 4. Conditional effect of daily steps on Beck Depression Inventory-II (BDI-II) improvement at different levels of pain (N=20).

Variable	b_{Steps}	SE	*t* (*df*)	*P* value	95% CI
Pain severity level (BPI-sev)[a]					
0 (Low)	2.81	1.08	2.60 (16)	.02	0.52 to 5.09
3 (High)	–1.56	1.40	–1.11 (16)	.28	–4.53 to 1.41
Pain interference level (BPI-intf)[b]					
0 (Low)	2.20	0.88	2.50 (16)	.02	0.34 to 4.06
4 (High)	–0.25	1.44	–0.17 (16)	.87	–3.29 to 2.80

[a]BPI-sev: Brief Pain Inventory: Severity.

[b]BPI-intf: Brief Pain Inventory: Interference.

Hypothesis 3: Baseline Depression and Anxiety Levels Predict Week-To-Week Step Trajectories Over the Course of iMCBT

Conditional growth models were used to test whether changes in steps over time (trajectories) varied as a function of baseline depression or anxiety. We use b_{Time} to denote the effect of an additional week on the weekly average of steps, conditioned on the levels of baseline depression or baseline anxiety.

Depression ($b_{\text{Time × BDI-II}}$=–7.29) and anxiety ($b_{\text{Time × BAI}}$=–8.32) were significant moderators of step trajectories (Table 5), indicating that greater baseline BDI-II or BAI severity was associated with larger step reductions over the course of the intervention. Figure 2 depicts the interaction effects and Table 6 provides the corresponding trajectory estimates (b_{Time}) at low, medium, and high levels of BDI-II and BAI (levels calculated as mean ± SD).

Table 5. Results from longitudinal modeling of estimated step trajectories conditioned on baseline depression or baseline anxiety (N=20).

Variable	*b*	SE	*t* (*df*)	*P* value	95% CI
Model BDI-II[a]					
(Intercept)	9479.56	2344.89	4.04 (417)	<.001	4870.27 to 14,088.85
Time (Week)	170.77	67.11	2.54 (417)	.01	38.86 to 302.69
BDI-II	–32.22	77.50	–0.42 (18)	.68	–195.03 to 130.59
Time × BDI-II	–7.29	2.27	–3.21 (417)	.001	–11.76 to –2.83
Model BAI[b]					
(Intercept)	13124.06	1788.53	7.34 (417)	<.001	9608.40 to 16,639.72
Time (Week)	183.48	60.48	3.03 (417)	.003	64.59 to 302.37
BAI	–161.87	61.43	–2.64 (18)	.02	–290.92 to –32.82
Time × BAI	–8.32	2.21	–3.77 (417)	<.001	–12.66 to –3.98

[a]BDI-II: Beck Depression Inventory-II.

[b]BAI: Beck Anxiety Inventory.

Figure 2. Visualization of step trajectories over the course of internet-based, mindfulness-based cognitive behavioral therapy (iMCBT) at different levels of baseline depression and anxiety (N=20). Higher baseline levels of depression and anxiety were associated with larger decreases in physical activity over the course of treatment. BAI: Beck Anxiety Inventory; BDI-II: Beck Depression Inventory-II.

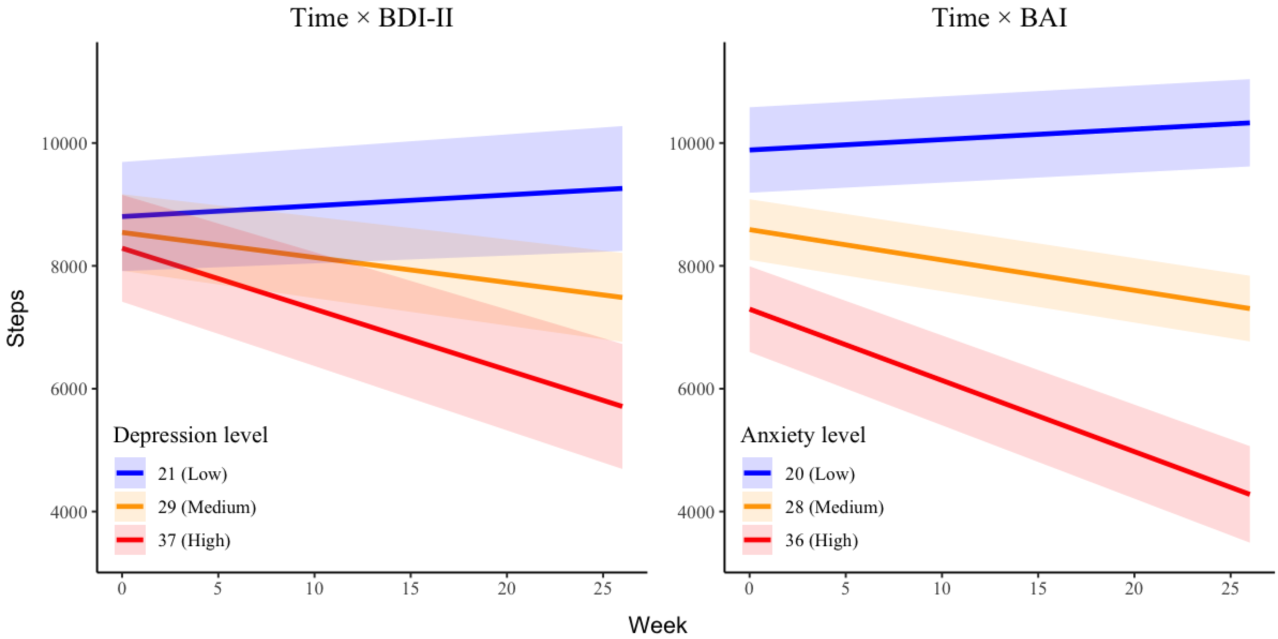


Table 6. Step trajectories (slopes) at different levels of baseline depression and anxiety (N=20).

Variable	b_{Time}	SE	t (df)	P value	95% CI
Depression level (BDI-II)[a]					
21 (Low)	17.60	25.11	0.70 (417)	.48	–31.77 to 66.97
29 (Medium)	–40.75	18.80	–2.17 (417)	.03	–77.70 to –3.80
37 (High)	–99.10	27.15	–3.65 (417)	<.001	–152.47 to –45.74
Anxiety level (BAI)[b]					
20 (Low)	17.06	22.40	0.76 (417)	.45	–26.98 to 61.10
28 (Medium)	–49.51	18.23	–2.72 (417)	.007	–85.34 to –13.67
36 (High)	–116.08	28.06	–4.14 (417)	<.001	–171.23 to –60.92

[a]BDI-II: Beck Depression Inventory-II.

[b]BAI: Beck Anxiety Inventory.

Hypothesis 4: Early Step Trajectories (Slopes) and Trajectories Across the Trial Are Positively Associated With Greater BDI-II Improvement

Early Step Trajectories and BDI-II Improvement

Regression models were used to test the relationship between early step increases (slopes) and BDI-II improvement, controlling for intercepts (estimated step levels at week 0), pain, and anxiety.

"Early" was defined as the first 6 weeks of iMCBT. Slopes represent the linear change in weekly step averages per week (eg, a slope estimate of 20 would indicate that each week, the weekly step average increased by 20). Thus, a 1-unit increase in an individual's slope would mean 1 additional average step per week. The average slope during the first 6 weeks of iMCBT was –35.47 (SD 403.73), and the average initial step level (intercept) was 8467.74 (SD 2847.34).

Per Table 7, the effect of slopes on BDI-II improvement was significant, controlling for estimated initial step levels, BPI-sev, and BAI (b=0.0137; SE=0.0060; t_{15}=2.26; P=.04). The effect of slopes was also significant when controlling for BPI-intf and the other 2 covariates (b=0.0165; SE=0.0051; t_{15}=3.23; P=.006).

In other words, each 100-unit increase (per week) in the weekly average of steps during early iMCBT predicted a 1.65 greater BDI-II improvement at the end of iMCBT, controlling for pain interference and covariates. The effect of early step increases was slightly weaker when controlling for pain severity and covariates.

Table 7. Regression of Beck Depression Inventory-II (BDI-II) change on 6-week step trajectories, adjusted for initial step levels, anxiety, and pain (N=20).

Variable	*b*	SE	*t* (df)	*P* value	95% CI	R^2
Model A						.30[a]
(Intercept)	3.49	14.17	0.25 (15)	.81	–26.71 to 33.70	
Slope	0.0137	0.0060	2.26 (15)	.04	0.0008 to 0.0265	
Step level (per 1000; week 0)	1.05	0.98	1.08 (15)	.30	–1.04 to 3.14	
BPI-sev[b]	3.20	1.81	1.77 (15)	.10	–0.65 to 7.04	
BAI[c]	–0.06	0.31	–0.20 (15)	.85	–0.73 to 0.61	
Model B						.52[d]
(Intercept)	–14.93	13.47	–1.11 (15)	.29	–43.65 to 13.79	
Slope	0.0165	0.0051	3.23 (15)	.006	0.0056 to 0.0274	
Step level (per 1000; week 0)	2.41	0.96	2.51 (15)	.02	0.36 to 4.46	
BPI-intf[e]	3.02	0.90	3.36 (15)	.004	1.10 to 4.94	
BAI	0.17	0.24	0.68 (15)	.51	–0.35 to 0.68	

[a] $F_{4,15}$=1.64; *P*=.22.

[b] BPI-sev: Brief Pain Inventory: Severity.

[c] BAI: Beck Anxiety Inventory.

[d] $F_{4,15}$=4.06; *P*=.02.

[e] BPI-intf: Brief Pain Inventory: Interference.

Step Trajectories Across iMCBT and BDI-II Improvement

BDI-II change was regressed on slopes across the full duration of iMCBT, controlling for covariates. A total of 3 participants were excluded due to adherence levels below 50%. The average slope across the entire trial was –20.68 (SD 85.74), and the average initial step level (intercept) was 8492.71 (SD 2887.52). Slopes across the trial were not significantly associated with MDD symptom improvement, controlling for

- initial step levels and BAI, *b*=0.0430; *P*=.40; 95% CI (–0.0635 to 0.1495)
- initial step levels and BPI-sev, *b*=–0.0223; *P*=.40; 95% CI (–0.0779 to 0.0332)
- initial step levels and BPI-intf, *b*=–0.0208; *P*=.40; 95% CI (–0.0728 to 0.0313)

Full regression results are presented in Table S2 in Multimedia Appendix 1.

Discussion

This study examined daily steps in relation to MDD symptom improvement and baseline characteristics in a youth sample who had completed a 24-week iMCBT treatment for mild to severe depression.

Principal Findings

The average patient in our sample was moderately to severely depressed at baseline (mean 29.20, SD 8.22). Over the course of treatment, participants took an average of 8269 steps per day, within the suggested optimal dose of 8000-10,000 [71,76]. We found that each additional 1000-step increment was significantly associated with a 2.66-point reduction in MDD symptoms (BDI-II) after controlling for adherence, anxiety, and pain interference. This finding is consistent with research showing that PA dose positively predicts greater antidepressant response [77]. The effect has clinical relevance given that in our sample, a 5-point reduction is considered a clinically meaningful improvement (based on a 17.5% reduction in BDI-II from baseline [78]). Thus, an additional 2000 steps daily would be associated with a meaningful reduction in depression symptoms. When we controlled for pain severity instead of pain interference, the 1000-step effect was of similar magnitude (*b*=1.95) but nonsignificant.

Given the observed bimodality in our sample's distribution of pain scores and previous evidence of pain's differential effects on PA-related benefits [28], we explored the interaction effect between pain (severity and interference) and daily steps on MDD symptom reduction. Analyses showed that as pain levels increased, the effect of steps on depression symptom reduction decreased. Our finding is consistent with previous work demonstrating pain's negative impact on depression treatment response [79-81], and suggests that while PA is consistently supported in the management of pain conditions [82], PA prescriptions for pain should be tailored to individual circumstances for optimal response [83-86]. In patients with fibromyalgia, for example, low to moderate PA is often associated with successful outcomes [87,88] whereas strenuous exercise induces hyperalgesia [28,89]. Additionally, dysfunction in stress response systems is often observed in patients with chronic pain [90], resulting in increased vulnerability to PA-related stress [84]. Thus, the smaller symptom reductions associated with excess PA among patients with higher pain levels may reflect impaired stress adaptation and inadequate restoration of homeostatic balance [84,91]. Future research is needed to verify these findings in larger, more diverse samples.

Additional study findings suggest that higher baseline levels of depression and anxiety were associated with larger decreases in PA over the course of iMCBT treatment. These results are consistent with other prospective research showing that initial anxiety and depression severity levels predict reduced PA over time [35,37]. Patients with greater depressive symptoms likely had less energy, confidence, or motivation [92] needed to persist in PAs. Additionally, anxiety disorders are often associated with social withdrawal and avoidance behaviors [93] that favor reduced PA and increased sedentariness [94,95]. Our findings are distinct from previous research in that the present associations occur in the context of a multimodal iMCBT intervention.

There was evidence that early PA increases predicted greater MDD symptom improvement at the end of the trial. Specifically, during the early weeks of iMCBT, each 100-unit increase in the weekly average of steps was associated with a 1.65-point reduction in BDI-II severity, controlling for pain interference, anxiety, and estimated step levels at baseline. Thus, a clinically meaningful reduction in BDI-II of 5 points would be expected if a patient's weekly step average increased by 300 steps each week during early iMCBT. This finding is consistent with previous evidence demonstrating the predictive value of early changes for later outcomes [96], with most studies focusing on early psychological gains [45,97,98] and fewer on early behavioral gains [47,48]. Our finding extends the literature by adding objectively measured PA to the accumulated evidence of early predictors of final treatment outcomes.

We did not find support for the hypothesis that, across the full duration of iMCBT, positive step trajectories would predict greater depression improvement. This was possibly due to the length of the trial combined with the heterogeneous nature of depression. It is well established that symptoms vary substantially between patients [99,100], with at least 1030 unique symptom profiles identified [101]. Therefore, as iMCBT treatment progressed, therapeutic tasks also progressed with increasing differentiation. For example, increasing approach behaviors may have been emphasized for patients prone to depressive avoidance [102], whereas calming mindfulness techniques may have been more important for patients with anxious distress [103]. Given the multimodal nature of iMCBT, varying therapeutic tasks were emphasized with differing implications for long-term PA trajectories.

Limitations

This study has several limitations. First, our sample size was modest, but this limitation was balanced against strict sampling criteria and intervention duration. Whether the study findings generalize will be confirmed in a future, larger trial targeting MDD in a broader age group [104].

Second, there was no measure of PA intensity (eg, moderate vs vigorous PA). Step count combined with intensity metrics might have enabled more detailed PA profiles, revealing more diverse impacts on depression.

Additionally, a basic measure like step count does not account for certain PAs (eg, strength training and water-based activities) that may contribute to depression improvement, though trial observations indicate that participants rarely engaged in such activities.

Lastly, while our data do not account for adherence to Fitbit monitoring at the within-day level, patients were instructed to wear their Fitbit monitors 24 hours daily. As an approximation of intraday adherence, we estimated day-to-day adherence (ie, days tracked divided by intervention period), which was included as a covariate in most of the multivariable regression models. To further account for possible missing intraday data, we removed daily step counts less than 100. Although a 100-step threshold may seem low given that previous researchers have used thresholds as high as 1000 [105], we considered potential vegetative symptoms commonly associated with major depression. Further, we evaluated the sensitivity of our results to a 1000-step threshold and found that it yielded only slightly different effect sizes and no substantive changes to inferential tests.

Conclusions

This is the first known prospective study to investigate objectively measured PA and its associations with depression symptom improvement during a multimodal cognitive-behavioral intervention. Our findings suggest that higher activity levels were associated with larger improvements and that this relationship may be influenced by chronic pain. We also extend the current literature by showing that initial symptom severity predicts PA change throughout treatment, and that early increases in PA predict better final outcomes. Given the role of PA in relapse prevention, it may be valuable to target activity levels in those with heightened symptoms early in the course of treatment. Future research is needed to examine how and to what extent PA changes can maximize treatment response in the context of multimodal interventions.

Acknowledgments

PR has received research support from the Canadian Institutes of Health Research and the Federal Development Program of Southern Ontario. JK is supported by a Canadian Institutes of Health Research Canada Research Chair in Health Psychology at York University.

Data Availability

The data sets generated during and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

PR currently receives research funding from the Canadian Institute for Health Research, the Social Sciences and Humanities Research Council of Canada, and the Canadian Pain Society. He receives software platform support from NexJ Health, Inc. and pilot research funding through the Digital Health Research Fund, administered through York University.
ZD received research and equipment in-kind support for an investigator-initiated study through Brainsway Inc and Magventure Inc. He is also on the scientific advisory board for Brainsway Inc. His work has been supported by the National Institutes of Mental Health, the Canadian Institutes of Health Research, Brain Canada, the Temerty Family Foundation, and the Grant Family Foundation.

Multimedia Appendix 1
Supplementary figures and and tables.
[DOCX File , 43 KB - ijmr_v12i1e46419_app1.docx]

Multimedia Appendix 2
CONSORT (Consolidated Standards of Reporting Trials) diagram. iMCBT: internet-based, mindfulness-based cognitive behavioral therapy.
[PNG File , 85 KB - ijmr_v12i1e46419_app2.png]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 944 KB - ijmr_v12i1e46419_app3.pdf]

References

1. Depression and Other Common Mental Disorders: Global Health Estimates. World Health Organization. 2017. URL: https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf [accessed 2023-09-07]
2. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. Am J Psychiatry 2006;163(11):1905-1917 [FREE Full text] [doi: 10.1176/ajp.2006.163.11.1905] [Medline: 17074942]
3. Levkovitz Y, Tedeschini E, Papakostas GI. Efficacy of antidepressants for dysthymia: a meta-analysis of placebo-controlled randomized trials. J Clin Psychiatry 2011;72(4):509-514. [doi: 10.4088/JCP.09m05949blu] [Medline: 21527126]
4. Cuijpers P, Karyotaki E, Ciharova M, Miguel C, Noma H, Furukawa TA. The effects of psychotherapies for depression on response, remission, reliable change, and deterioration: a meta-analysis. Acta Psychiatr Scand 2021;144(3):288-299 [FREE Full text] [doi: 10.1111/acps.13335] [Medline: 34107050]
5. Steinert C, Hofmann M, Kruse J, Leichsenring F. Relapse rates after psychotherapy for depression—stable long-term effects? A meta-analysis. J Affect Disord 2014;168:107-118 [FREE Full text] [doi: 10.1016/j.jad.2014.06.043] [Medline: 25043322]
6. WHO Guidelines on Physical Activity and Sedentary Behaviour. World Health Organization. 2020. URL: https://www.who.int/publications/i/item/9789240015128 [accessed 2023-09-07]
7. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Rep 1985;100(2):126-131 [FREE Full text] [Medline: 3920711]
8. Kvam S, Kleppe CL, Nordhus IH, Hovland A. Exercise as a treatment for depression: a meta-analysis. J Affect Disord 2016;202:67-86 [FREE Full text] [doi: 10.1016/j.jad.2016.03.063] [Medline: 27253219]
9. Mura G, Moro MF, Patten SB, Carta MG. Exercise as an add-on strategy for the treatment of major depressive disorder: a systematic review. CNS Spectr 2014;19(6):496-508. [doi: 10.1017/S1092852913000953] [Medline: 24589012]
10. Schuch FB, Vancampfort D, Richards J, Rosenbaum S, Ward PB, Stubbs B. Exercise as a treatment for depression: a meta-analysis adjusting for publication bias. J Psychiatr Res 2016;77:42-51 [FREE Full text] [doi: 10.1016/j.jpsychires.2016.02.023] [Medline: 26978184]
11. Robertson R, Robertson A, Jepson R, Maxwell M. Walking for depression or depressive symptoms: a systematic review and meta-analysis. Ment Health Phys Act 2012;5(1):66-75. [doi: 10.1016/j.mhpa.2012.03.002]
12. Nyström MBT, Neely G, Hassmén P, Carlbring P. Treating major depression with physical activity: a systematic overview with recommendations. Cogn Behav Ther 2015;44(4):341-352. [doi: 10.1080/16506073.2015.1015440] [Medline: 25794191]
13. Cooney GM, Dwan K, Greig CA, Lawlor DA, Rimer J, Waugh FR, et al. Exercise for depression. Cochrane Database Syst Rev 2013;2013(9):CD004366 Cochrane Common Mental Disorders Group, editor [FREE Full text] [doi: 10.1002/14651858.CD004366.pub6] [Medline: 24026850]
14. Singh B, Olds T, Curtis R, Dumuid D, Virgara R, Watson A, et al. Effectiveness of physical activity interventions for improving depression, anxiety and distress: an overview of systematic reviews. Br J Sports Med 2023 [FREE Full text] [doi: 10.1136/bjsports-2022-106195] [Medline: 36796860]

15. Schuch FB, Vancampfort D, Firth J, Rosenbaum S, Ward PB, Silva ES, et al. Physical activity and incident depression: a meta-analysis of prospective cohort studies. Am J Psychiatry 2018;175(7):631-648 [FREE Full text] [doi: 10.1176/appi.ajp.2018.17111194] [Medline: 29690792]
16. Schuch F, Vancampfort D, Firth J, Rosenbaum S, Ward P, Reichert T, et al. Physical activity and sedentary behavior in people with major depressive disorder: a systematic review and meta-analysis. J Affect Disord 2017;210:139-150 [FREE Full text] [doi: 10.1016/j.jad.2016.10.050] [Medline: 28033521]
17. Ainsworth B, Cahalin L, Buman M, Ross R. The current state of physical activity assessment tools. Prog Cardiovasc Dis 2015;57(4):387-395. [doi: 10.1016/j.pcad.2014.10.005] [Medline: 25446555]
18. Prince SA, Adamo KB, Hamel ME, Hardt J, Gorber SC, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act 2008;5:56 [FREE Full text] [doi: 10.1186/1479-5868-5-56] [Medline: 18990237]
19. Choi KW, Chen CY, Stein MB, Klimentidis YC, Wang MJ, Koenen KC, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium. Assessment of bidirectional relationships between physical activity and depression among adults: a 2-sample mendelian randomization study. JAMA Psychiatry 2019;76(4):399-408 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.4175] [Medline: 30673066]
20. Fukuoka Y, Haskell W, Vittinghoff E. New insights into discrepancies between self-reported and accelerometer-measured moderate to vigorous physical activity among women—the mPED trial. BMC Public Health 2016;16(1):761 [FREE Full text] [doi: 10.1186/s12889-016-3348-7] [Medline: 27514368]
21. Zaninotto L, Solmi M, Veronese N, Guglielmo R, Ioime L, Camardese G, et al. A meta-analysis of cognitive performance in melancholic versus non-melancholic unipolar depression. J Affect Disord 2016;201:15-24 [FREE Full text] [doi: 10.1016/j.jad.2016.04.039] [Medline: 27156095]
22. Zuckerman H, Pan Z, Park C, Brietzke E, Musial N, Shariq AS, et al. Recognition and treatment of cognitive dysfunction in major depressive disorder. Front Psychiatry 2018;9:655 [FREE Full text] [doi: 10.3389/fpsyt.2018.00655] [Medline: 30564155]
23. Vancampfort D, Firth J, Schuch FB, Rosenbaum S, Mugisha J, Hallgren M, et al. Sedentary behavior and physical activity levels in people with schizophrenia, bipolar disorder and major depressive disorder: a global systematic review and meta-analysis. World Psychiatry 2017;16(3):308-315 [FREE Full text] [doi: 10.1002/wps.20458] [Medline: 28941119]
24. Gianfredi V, Blandi L, Cacitti S, Minelli M, Signorelli C, Amerio A, et al. Depression and objectively measured physical activity: a systematic review and meta-analysis. Int J Environ Res Public Health 2020;17(10):3738 [FREE Full text] [doi: 10.3390/ijerph17103738] [Medline: 32466242]
25. Dhanju S, Kennedy SH, Abbey S, Katz J, Weinrib A, Clarke H, et al. The impact of comorbid pain and depression in the United States: results from a nationally representative survey. Scand J Pain 2019;19(2):319-325. [doi: 10.1515/sjpain-2018-0323] [Medline: 30759071]
26. Geneen LJ, Moore RA, Clarke C, Martin D, Colvin LA, Smith BH. Physical activity and exercise for chronic pain in adults: an overview of cochrane reviews. Cochrane Database Syst Rev 2017;4(4):CD011279 [FREE Full text] [doi: 10.1002/14651858.CD011279.pub3] [Medline: 28436583]
27. Lima LV, Abner TSS, Sluka KA. Does exercise increase or decrease pain? Central mechanisms underlying these two phenomena. J Physiol 2017;595(13):4141-4150 [FREE Full text] [doi: 10.1113/JP273355] [Medline: 28369946]
28. Rice D, Nijs J, Kosek E, Wideman T, Hasenbring MI, Koltyn K, et al. Exercise-induced hypoalgesia in pain-free and chronic pain populations: state of the art and future directions. J Pain 2019;20(11):1249-1266 [FREE Full text] [doi: 10.1016/j.jpain.2019.03.005] [Medline: 30904519]
29. Heissel A, Heinen D, Brokmeier LL, Skarabis N, Kangas M, Vancampfort D, et al. Exercise as medicine for depressive symptoms? A systematic review and meta-analysis with meta-regression. Br J Sports Med 2023;57(16):1049-1057 [FREE Full text] [doi: 10.1136/bjsports-2022-106282] [Medline: 36731907]
30. Schuch FB, Morres ID, Ekkekakis P, Rosenbaum S, Stubbs B. A critical review of exercise as a treatment for clinically depressed adults: time to get pragmatic. Acta Neuropsychiatr 2017;29(2):65-71 [FREE Full text] [doi: 10.1017/neu.2016.21] [Medline: 27145824]
31. Mammen G, Faulkner G. Physical activity and the prevention of depression: a systematic review of prospective studies. Am J Prev Med 2013;45(5):649-657. [doi: 10.1016/j.amepre.2013.08.001] [Medline: 24139780]
32. Stubbs B, Koyanagi A, Hallgren M, Firth J, Richards J, Schuch F, et al. Physical activity and anxiety: a perspective from the World Health Survey. J Affect Disord 2017;208:545-552 [FREE Full text] [doi: 10.1016/j.jad.2016.10.028] [Medline: 27802893]
33. Da Silva MA, Singh-Manoux A, Brunner EJ, Kaffashian S, Shipley MJ, Kivimäki M, et al. Bidirectional association between physical activity and symptoms of anxiety and depression: the whitehall II study. Eur J Epidemiol 2012;27(7):537-546 [FREE Full text] [doi: 10.1007/s10654-012-9692-8] [Medline: 22623145]
34. Gudmundsson P, Lindwall M, Gustafson DR, Östling S, Hällström T, Waern M, et al. Longitudinal associations between physical activity and depression scores in Swedish women followed 32 years. Acta Psychiatr Scand 2015;132(6):451-458 [FREE Full text] [doi: 10.1111/acps.12419] [Medline: 25865488]

35. Hiles SA, Lamers F, Milaneschi Y, Penninx BWJH. Sit, step, sweat: longitudinal associations between physical activity patterns, anxiety and depression. Psychol Med 2017;47(8):1466-1477. [doi: 10.1017/S0033291716003548] [Medline: 28137333]
36. Pinto Pereira SM, Geoffroy MC, Power C. Depressive symptoms and physical activity during 3 decades in adult life: bidirectional associations in a prospective cohort study. JAMA Psychiatry 2014;71(12):1373-1380 [FREE Full text] [doi: 10.1001/jamapsychiatry.2014.1240] [Medline: 25321867]
37. Roshanaei-Moghaddam B, Katon WJ, Russo J. The longitudinal effects of depression on physical activity. Gen Hosp Psychiatry 2009;31(4):306-315 [FREE Full text] [doi: 10.1016/j.genhosppsych.2009.04.002] [Medline: 19555789]
38. Teychenne M, Abbott G, Lamb KE, Rosenbaum S, Ball K. Is the link between movement and mental health a two-way street? Prospective associations between physical activity, sedentary behaviour and depressive symptoms among women living in socioeconomically disadvantaged neighbourhoods. Prev Med 2017;102:72-78 [FREE Full text] [doi: 10.1016/j.ypmed.2017.07.005] [Medline: 28694061]
39. Harris AHS, Cronkite R, Moos R. Physical activity, exercise coping, and depression in a 10-year cohort study of depressed patients. J Affect Disord 2006;93(1-3):79-85 [FREE Full text] [doi: 10.1016/j.jad.2006.02.013] [Medline: 16545873]
40. Ku PW, Fox KR, Chen LJ, Chou P. Physical activity and depressive symptoms in older adults: 11-year follow-up. Am J Prev Med 2012;42(4):355-362. [doi: 10.1016/j.amepre.2011.11.010] [Medline: 22424248]
41. Barth J, Munder T, Gerger H, Nüesch E, Trelle S, Znoj H, et al. Comparative efficacy of seven Psychotherapeutic interventions for patients with depression: a network meta-analysis. Focus (Am Psychiatr Publ) 2016;14(2):229-243 [FREE Full text] [doi: 10.1176/appi.focus.140201] [Medline: 31997951]
42. Cuijpers P, de Wit L, Taylor A. The effects of psychological treatments for adult depression on physical activity: a systematic review. Ment Health Phys Act 2014;7(1):6-8. [doi: 10.1016/j.mhpa.2014.01.002]
43. Fagan MJ, Faulkner G, Duncan M, Yun L, Blumberger D, Daskalakis ZJ, et al. Moving away from depression: physical activity changes in patients undergoing r-TMS for major depressive disorder. Ment Health Phys Act 2019;16:50-53. [doi: 10.1016/j.mhpa.2019.03.003]
44. Johansson P, Svensson E, Andersson G, Lundgren J. Trajectories and associations between depression and physical activity in patients with cardiovascular disease during participation in an internet-based cognitive behavioural therapy programme. Eur J Cardiovasc Nurs 2021;20(2):124-131 [FREE Full text] [doi: 10.1177/1474515120947250] [Medline: 33611380]
45. Lewis CC, Simons AD, Kim HK. The role of early symptom trajectories and pretreatment variables in predicting treatment response to cognitive behavioral therapy. J Consult Clin Psychol 2012;80(4):525-534. [doi: 10.1037/a0029131] [Medline: 22730951]
46. Lutz W, Stulz N, Köck K. Patterns of early change and their relationship to outcome and follow-up among patients with major depressive disorders. J Affect Disord 2009;118(1-3):60-68 [FREE Full text] [doi: 10.1016/j.jad.2009.01.019] [Medline: 19217669]
47. Katz MM, Tekell JL, Bowden CL, Brannan S, Houston JP, Berman N, et al. Onset and early behavioral effects of pharmacologically different antidepressants and placebo in depression. Neuropsychopharmacology 2004;29(3):566-579 [FREE Full text] [doi: 10.1038/sj.npp.1300341] [Medline: 14627997]
48. Dimidjian S, Goodman SH, Sherwood NE, Simon GE, Ludman E, Gallop R, et al. A pragmatic randomized clinical trial of behavioral activation for depressed pregnant women. J Consult Clin Psychol 2017;85(1):26-36 [FREE Full text] [doi: 10.1037/ccp0000151] [Medline: 28045285]
49. Byrne N, Regan C, Livingston G. Adherence to treatment in mood disorders. Curr Opin Psychiatry 2006;19(1):44-49. [doi: 10.1097/01.yco.0000191501.54034.7c] [Medline: 16612178]
50. Wilbur J, Zenk S, Wang E, Oh A, McDevitt J, Block D, et al. Neighborhood characteristics, adherence to walking, and depressive symptoms in midlife African American women. J Womens Health (Larchmt) 2009;18(8):1201-1210 [FREE Full text] [doi: 10.1089/jwh.2008.1168] [Medline: 19630546]
51. Herman S, Blumenthal JA, Babyak M, Khatri P, Craighead WE, Krishnan KR, et al. Exercise therapy for depression in middle-aged and older adults: predictors of early dropout and treatment failure. Health Psychol 2002;21(6):553-563. [doi: 10.1037/0278-6133.21.6.553]
52. Teychenne M, Costigan SA, Parker K. The association between sedentary behaviour and risk of anxiety: a systematic review. BMC Public Health 2015;15:513 [FREE Full text] [doi: 10.1186/s12889-015-1843-x] [Medline: 26088005]
53. Karlsson L, Gerdle B, Takala EP, Andersson G, Larsson B. Experiences and attitudes about physical activity and exercise in patients with chronic pain: a qualitative interview study. J Pain Res 2018;11:133-144 [FREE Full text] [doi: 10.2147/JPR.S149826] [Medline: 29379314]
54. Ritvo P, Daskalakis ZJ, Tomlinson G, Ravindran A, Linklater R, Chang MK, et al. An online mindfulness-based cognitive behavioral therapy intervention for youth diagnosed with major depressive disorders: protocol for a randomized controlled trial. JMIR Res Protoc 2019;8(7):e11591 [FREE Full text] [doi: 10.2196/11591] [Medline: 31359869]
55. Ritvo P, Knyahnytska Y, Pirbaglou M, Wang W, Tomlinson G, Zhao H, et al. Online mindfulness-based cognitive behavioral therapy intervention for youth with major depressive disorders: randomized controlled trial. J Med Internet Res 2021;23(3):e24380 [FREE Full text] [doi: 10.2196/24380] [Medline: 33688840]

56. Smarr KL, Keefer AL. Measures of depression and depressive symptoms: Beck Depression Inventory-II (BDI-II), Center for Epidemiologic Studies Depression Scale (CES-D), Geriatric Depression Scale (GDS), Hospital Anxiety and Depression Scale (HADS), and Patient Health Questionnaire-9 (PHQ-9). Arthritis Care Res (Hoboken) 2011;63(Suppl 11):S454-S466 [FREE Full text] [doi: 10.1002/acr.20556] [Medline: 22588766]
57. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59(Suppl 20):22-33 quiz 34-57. [Medline: 9881538]
58. Ahmad F, El Morr C, Ritvo P, Othman N, Moineddin R, Team MVC. An eight-week, web-based mindfulness virtual community intervention for students' mental health: randomized controlled trial. JMIR Ment Health 2020;7(2):e15520 [FREE Full text] [doi: 10.2196/15520] [Medline: 32074061]
59. El Morr C, Ritvo P, Ahmad F, Moineddin R, Team MVC. Effectiveness of an 8-week web-based mindfulness virtual community intervention for university students on symptoms of stress, anxiety, and depression: randomized controlled trial. JMIR Ment Health 2020;7(7):e18595 [FREE Full text] [doi: 10.2196/18595] [Medline: 32554380]
60. Beck AT, Steer RA, Brown G. Manual for the Beck Depression Inventory-II. San Antonio, TX, USA: Psychological Corporation; 1996.
61. Wang YP, Gorenstein C. Psychometric properties of the beck depression inventory-II: a comprehensive review. Braz J Psychiatry 2013;35(4):416-431 [FREE Full text] [doi: 10.1590/1516-4446-2012-1048] [Medline: 24402217]
62. Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. J Consult Clin Psychol 1988;56(6):893-897. [doi: 10.1037//0022-006x.56.6.893] [Medline: 3204199]
63. Bardhoshi G, Duncan K, Erford BT. Psychometric meta-analysis of the English version of the beck anxiety inventory. J Couns Dev 2016;94(3):356-373. [doi: 10.1002/jcad.12090]
64. Hewitt PL, Norton GR. The beck anxiety inventory: a psychometric analysis. Psychol Assess 1993;5(4):408-412. [doi: 10.1037/1040-3590.5.4.408]
65. Cleeland CS. The brief pain inventory user guide. 2009. URL: https://www.mdanderson.org/content/dam/mdanderson/documents/Departments-and-Divisions/Symptom-Research/BPI_UserGuide.pdf [accessed 2023-09-07]
66. Poquet N, Lin C. The Brief Pain Inventory (BPI). J Physiother 2016;62(1):52 [FREE Full text] [doi: 10.1016/j.jphys.2015.07.001] [Medline: 26303366]
67. Bai Y, Tompkins C, Gell N, Dione D, Zhang T, Byun W. Comprehensive comparison of Apple Watch and Fitbit monitors in a free-living setting. PLoS One 2021;16(5):e0251975 [FREE Full text] [doi: 10.1371/journal.pone.0251975] [Medline: 34038458]
68. Straiton N, Alharbi M, Bauman A, Neubeck L, Gullick J, Bhindi R, et al. The validity and reliability of consumer-grade activity trackers in older, community-dwelling adults: a systematic review. Maturitas 2018;112:85-93 [FREE Full text] [doi: 10.1016/j.maturitas.2018.03.016] [Medline: 29704922]
69. Tedesco S, Sica M, Ancillao A, Timmons S, Barton J, O'Flynn B. Validity evaluation of the Fitbit Charge2 and the Garmin vivosmart HR+ in free-living environments in an older adult cohort. JMIR Mhealth Uhealth 2019;7(6):e13084 [FREE Full text] [doi: 10.2196/13084] [Medline: 31219048]
70. Washburn BD, Ihm JM. Using step counts to prescribe physical activity: what is the optimal dose? Curr Sports Med Rep 2021;20(8):402-409 [FREE Full text] [doi: 10.1249/JSR.0000000000000868] [Medline: 34357886]
71. Paluch AE, Bajpai S, Bassett DR, Carnethon MR, Ekelund U, Evenson KR, Steps for Health Collaborative. Daily steps and all-cause mortality: a meta-analysis of 15 international cohorts. Lancet Public Health 2022;7(3):e219-e228 [FREE Full text] [doi: 10.1016/S2468-2667(21)00302-9] [Medline: 35247352]
72. Delahanty LM, Conroy MB, Nathan DM, Diabetes Prevention Program Research Group. Psychological predictors of physical activity in the diabetes prevention program. J Am Diet Assoc 2006;106(5):698-705 [FREE Full text] [doi: 10.1016/j.jada.2006.02.011] [Medline: 16647327]
73. van Stralen MM, De Vries H, Mudde AN, Bolman C, Lechner L. Determinants of initiation and maintenance of physical activity among older adults: a literature review. Health Psychol Rev 2009;3(2):147-207. [doi: 10.1080/17437190903229462]
74. Pinheiro J, Bates D. nlme: Linear and nonlinear mixed effects models. R Core Team. 2022. URL: https://cran.r-project.org/web/packages/nlme/index.html [accessed 2023-09-07]
75. International Council on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH E9: Statistical principles for clinical trials. European Medicines Agency. 1998. URL: https://www.ema.europa.eu/en/documents/scientific-guideline/ich-e-9-statistical-principles-clinical-trials-step-5_en.pdf [accessed 2023-09-07]
76. Master H, Annis J, Huang S, Beckman JA, Ratsimbazafy F, Marginean K, et al. Association of step counts over time with the risk of chronic disease in the All of Us Research Program. Nat Med 2022;28(11):2301-2308 [FREE Full text] [doi: 10.1038/s41591-022-02012-w] [Medline: 36216933]
77. Dunn AL, Trivedi MH, Kampert JB, Clark CG, Chambliss HO. Exercise treatment for depression: efficacy and dose response. Am J Prev Med 2005;28(1):1-8 [FREE Full text] [doi: 10.1016/j.amepre.2004.09.003] [Medline: 15626549]
78. Button KS, Kounali D, Thomas L, Wiles NJ, Peters TJ, Welton NJ, et al. Minimal clinically important difference on the Beck Depression Inventory—II according to the patient's perspective. Psychol Med 2015;45(15):3269-3279 [FREE Full text] [doi: 10.1017/S0033291715001270] [Medline: 26165748]

79. Bair MJ, Robinson RL, Katon W, Kroenke K. Depression and pain comorbidity: a literature review. Arch Intern Med 2003;163(20):2433-2445 [FREE Full text] [doi: 10.1001/archinte.163.20.2433] [Medline: 14609780]
80. Kroenke K, Shen J, Oxman TE, Williams JW, Dietrich AJ. Impact of pain on the outcomes of depression treatment: results from the RESPECT trial. Pain 2008;134(1-2):209-215. [doi: 10.1016/j.pain.2007.09.021] [Medline: 18022319]
81. Thielke SM, Fan MY, Sullivan M, Unützer J. Pain limits the effectiveness of collaborative care for depression. Am J Geriatr Psychiatry 2007;15(8):699-707. [doi: 10.1097/JGP.0b013e3180325a2d] [Medline: 17670998]
82. Belavy DL, Van Oosterwijck J, Clarkson M, Dhondt E, Mundell NL, Miller CT, et al. Pain sensitivity is reduced by exercise training: evidence from a systematic review and meta-analysis. Neurosci Biobehav Rev 2021;120:100-108 [FREE Full text] [doi: 10.1016/j.neubiorev.2020.11.012] [Medline: 33253748]
83. Ambrose KR, Golightly YM. Physical exercise as non-pharmacological treatment of chronic pain: why and when. Best Pract Res Clin Rheumatol 2015;29(1):120-130 [FREE Full text] [doi: 10.1016/j.berh.2015.04.022] [Medline: 26267006]
84. Daenen L, Varkey E, Kellmann M, Nijs J. Exercise, not to exercise, or how to exercise in patients with chronic pain? Applying science to practice. Clin J Pain 2015;31(2):108-114. [doi: 10.1097/AJP.0000000000000099] [Medline: 24662498]
85. Hu H, Xu A, Gao C, Wang Z, Wu X. The effect of physical exercise on rheumatoid arthritis: an overview of systematic reviews and meta-analysis. J Adv Nurs 2021;77(2):506-522. [doi: 10.1111/jan.14574] [Medline: 33176012]
86. Joelsson M, Bernhardsson S, Larsson MEH. Patients with chronic pain may need extra support when prescribed physical activity in primary care: a qualitative study. Scand J Prim Health Care 2017;35(1):64-74 [FREE Full text] [doi: 10.1080/02813432.2017.1288815] [Medline: 28277047]
87. Bidonde J, Busch AJ, Bath B, Milosavljevic S. Exercise for adults with fibromyalgia: an umbrella systematic review with synthesis of best evidence. Curr Rheumatol Rev 2014;10(1):45-79. [doi: 10.2174/1573403x10666140914155304] [Medline: 25229499]
88. Häuser W, Klose P, Langhorst J, Moradi B, Steinbach M, Schiltenwolf M, et al. Efficacy of different types of aerobic exercise in fibromyalgia syndrome: a systematic review and meta-analysis of randomised controlled trials. Arthritis Res Ther 2010;12(3):R79 [FREE Full text] [doi: 10.1186/ar3002] [Medline: 20459730]
89. Vierck CJ, Staud R, Price DD, Cannon RL, Mauderli AP, Martin AD. The effect of maximal exercise on temporal summation of second pain (windup) in patients with fibromyalgia syndrome. J Pain 2001;2(6):334-344 [FREE Full text] [doi: 10.1054/jpai.2001.25533] [Medline: 14622813]
90. Woda A, Picard P, Dutheil F. Dysfunctional stress responses in chronic pain. Psychoneuroendocrinol 2016;71:127-135. [doi: 10.1016/j.psyneuen.2016.05.017] [Medline: 27262345]
91. Gatchel RJ, Peng YB, Peters ML, Fuchs PN, Turk DC. The biopsychosocial approach to chronic pain: scientific advances and future directions. Psychol Bull 2007;133(4):581-624. [doi: 10.1037/0033-2909.133.4.581] [Medline: 17592957]
92. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 5th Edition, Text Revision. Washington, D.C: American Psychiatric Association Publishing; 2022.
93. Hofmann SG, Hay AC. Rethinking avoidance: toward a balanced approach to avoidance in treating anxiety disorders. J Anxiety Disord 2018;55:14-21 [FREE Full text] [doi: 10.1016/j.janxdis.2018.03.004] [Medline: 29550689]
94. De Herdt A, Knapen J, Vancampfort D, De Hert M, Brunner E, Probst M. Social anxiety in physical activity participation in patients with mental illness: a cross-sectional multicenter study. Depress Anxiety 2013;30(8):757-762. [doi: 10.1002/da.22059] [Medline: 23532913]
95. Martinsen EW. Physical activity in the prevention and treatment of anxiety and depression. Nord J Psychiatry 2008;62(Suppl 47):25-29 [FREE Full text] [doi: 10.1080/08039480802315640] [Medline: 18752115]
96. Wagner S, Engel A, Engelmann J, Herzog D, Dreimüller N, Müller MB, et al. Early improvement as a resilience signal predicting later remission to antidepressant treatment in patients with major depressive disorder: systematic review and meta-analysis. J Psychiatr Res 2017;94:96-106. [doi: 10.1016/j.jpsychires.2017.07.003] [Medline: 28697423]
97. van Calker D, Zobel I, Dykierek P, Deimel CM, Kech S, Lieb K, et al. Time course of response to antidepressants: predictive value of early improvement and effect of additional psychotherapy. J Affect Disord 2009;114(1-3):243-253 [FREE Full text] [doi: 10.1016/j.jad.2008.07.023] [Medline: 18849079]
98. Tadić A, Helmreich I, Mergl R, Hautzinger M, Kohnen R, Henkel V, et al. Early improvement is a predictor of treatment outcome in patients with mild major, minor or subsyndromal depression. J Affect Disord 2010;120(1-3):86-93 [FREE Full text] [doi: 10.1016/j.jad.2009.04.014] [Medline: 19428118]
99. Fried EI. Moving forward: how depression heterogeneity hinders progress in treatment and research. Expert Rev Neurother 2017;17(5):423-425 [FREE Full text] [doi: 10.1080/14737175.2017.1307737] [Medline: 28293960]
100. Monroe SM, Anderson SF. Depression: the shroud of heterogeneity. Curr Dir Psychol Sci 2015;24(3):227-231. [doi: 10.1177/0963721414568342]
101. Fried EI, Nesse RM. Depression is not a consistent syndrome: an investigation of unique symptom patterns in the STAR*D study. J Affect Disord 2015;172:96-102 [FREE Full text] [doi: 10.1016/j.jad.2014.10.010] [Medline: 25451401]
102. Trew JL. Exploring the roles of approach and avoidance in depression: an integrative model. Clin Psychol Rev 2011;31(7):1156-1168. [doi: 10.1016/j.cpr.2011.07.007] [Medline: 21855826]
103. Segal ZV, Williams JMG, Teasdale JD. Mindfulness-Based Cognitive Therapy for Depression. 2nd Edition. New York: Guilford Press; 2013.

104. Ritvo P, Gratzer D, Knyahnytska Y, Ortiz A, Walters C, Katz J, et al. Comparing online and on-site cognitive behavior therapy in major depressive disorder: protocol for a noninferiority randomized controlled trial. JMIR Res Protoc 2022;11(4):e29726 [FREE Full text] [doi: 10.2196/29726] [Medline: 35393942]
105. Ding D, Cheng M, Del Pozo Cruz B, Lin T, Sun S, Zhang L, et al. How COVID-19 lockdown and reopening affected daily steps: evidence based on 164,630 person-days of prospectively collected data from Shanghai, China. Int J Behav Nutr Phys Act 2021;18(1):40 [FREE Full text] [doi: 10.1186/s12966-021-01106-x] [Medline: 33731132]

Abbreviations

BAI: Beck Anxiety Inventory
BDI-II: Beck Depression Inventory-II
BPI: Brief Pain Inventory
BPI-intf: Brief Pain Inventory: Interference
BPI-sev: Brief Pain Inventory: Severity
CAMH: Centre for Addiction and Mental Health
CBT: cognitive behavioral therapy
DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
iMCBT: internet-based, mindfulness-based cognitive behavioral therapy
MCBT: mindfulness-based cognitive behavioral therapy
MDD: major depressive disorder
MINI: Mini-International Neuropsychiatric Interview
PA: physical activity
RCT: randomized controlled trial

Edited by T de Azevedo Cardoso; submitted 30.04.23; peer-reviewed by P Owen, N Mundell; comments to author 06.07.23; revised version received 15.08.23; accepted 28.08.23; published 08.12.23.

Please cite as:
Dang K, Ritvo P, Katz J, Gratzer D, Knyahnytska Y, Ortiz A, Walters C, Attia M, Gonzalez-Torres C, Lustig A, Daskalakis Z
The Role of Daily Steps in the Treatment of Major Depressive Disorder: Secondary Analysis of a Randomized Controlled Trial of a 6-Month Internet-Based, Mindfulness-Based Cognitive Behavioral Therapy Intervention for Youth
Interact J Med Res 2023;12:e46419
URL: https://www.i-jmr.org/2023/1/e46419
doi:10.2196/46419
PMID:38064262

Original Paper

An Artificial Intelligence Exercise Coaching Mobile App: Development and Randomized Controlled Trial to Verify Its Effectiveness in Posture Correction

Han Joo Chae[1*], PhD; Ji-Been Kim[2*], PhD; Gwanmo Park[1], MSc; David Michael O'Sullivan[2], PhD; Jinwook Seo[1], PhD; Jung-Jun Park[2], PhD

[1]Department of Computer Science and Engineering, Seoul National University, Seoul, Republic of Korea

[2]Division of Sports Science, Pusan National University, Busan, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Jung-Jun Park, PhD
Division of Sports Science
Pusan National University
Busandaehak-ro 63beon-gil, Geumjeong-gu
Busan, 46241
Republic of Korea
Phone: 82 10 3285 1626
Email: jjparkpnu@pusan.ac.kr

Abstract

Background: Insufficient physical activity due to social distancing and suppressed outdoor activities increases vulnerability to diseases like cardiovascular diseases, sarcopenia, and severe COVID-19. While bodyweight exercises, such as squats, effectively boost physical activity, incorrect postures risk abnormal muscle activation joint strain, leading to ineffective sessions or even injuries. Avoiding incorrect postures is challenging for novices without expert guidance. Existing solutions for remote coaching and computer-assisted posture correction often prove costly or inefficient.

Objective: This study aimed to use deep neural networks to develop a personal workout assistant that offers feedback on squat postures using only mobile devices—smartphones and tablets. Deep learning mimicked experts' visual assessments of proper exercise postures. The effectiveness of the mobile app was evaluated by comparing it with exercise videos, a popular at-home workout choice.

Methods: Twenty participants were recruited without squat exercise experience and divided into an experimental group (EXP) with 10 individuals aged 21.90 (SD 2.18) years and a mean BMI of 20.75 (SD 2.11) and a control group (CTL) with 10 individuals aged 22.60 (SD 1.95) years and a mean BMI of 18.72 (SD 1.23) using randomized controlled trials. A data set with over 20,000 squat videos annotated by experts was created and a deep learning model was trained using pose estimation and video classification to analyze the workout postures. Subsequently, a mobile workout assistant app, Home Alone Exercise, was developed, and a 2-week interventional study, in which the EXP used the app while the CTL only followed workout videos, showed how the app helps people improve squat exercise.

Results: The EXP significantly improved their squat postures evaluated by the app after 2 weeks (Pre: 0.20 vs Mid: 4.20 vs Post: 8.00, *P*=.001), whereas the CTL (without the app) showed no significant change in squat posture (Pre: 0.70 vs Mid: 1.30 vs Post: 3.80, *P*=.13). Significant differences were observed in the left (Pre: 75.06 vs Mid: 76.24 vs Post: 63.13, *P*=.02) and right (Pre: 71.99 vs Mid: 76.68 vs Post: 62.82, *P*=.03) knee joint angles in the EXP before and after exercise, with no significant effect found for the CTL in the left (Pre: 73.27 vs Mid: 74.05 vs Post: 70.70, *P*=.68) and right (Pre: 70.82 vs Mid: 74.02 vs Post: 70.23, *P*=.61) knee joint angles.

Conclusions: EXP participants trained with the app experienced faster improvement and learned more nuanced details of the squat exercise. The proposed mobile app, offering cost-effective self-discovery feedback, effectively taught users about squat exercises without expensive in-person trainer sessions.

Trial Registration: Clinical Research Information Service KCT0008178 (retrospectively registered); https://cris.nih.go.kr/cris/search/detailSearch.do/24006

(Interact J Med Res 2023;12:e37604) doi:10.2196/37604

KEYWORDS

home workout; mobile assistant; deep-learning; posture correction; physical activity; exercise; social distance; COVID-19; mobile device; workout

Introduction

A recent study on the relationship between physical activity and the risk of COVID-19 has shown that engaging in the recommended levels of physical activity decreases the likelihood of SARS-CoV-2 infection, severe COVID-19 illness, and COVID-19–related death [1]. However, social distancing and lockdown after the outbreak of COVID-19 have resulted in a vigorous decrease in physical activity [2]. Since such reduction in physical activity could increase risks of not only COVID-19 but also cardiovascular diseases (eg, obesity, hypertension, diabetes, and metabolic syndrome) and even sarcopenia [3-7], physical activity is highly recommended during the confinement through the performance of aerobic, strength, flexibility, and balance exercises [8]. Body weight exercises, such as squats, sit-ups, and push-ups, are some of the best options because they are easy to perform at home without additional equipment and involve various joints and muscle movements [9].

Maintaining proper posture is vital to reap meaningful benefits from workouts. An incorrect posture can cause abnormal muscle activation or apply unwanted pressure on the body's joints, leading to less effective session results or even injuries [10,11]. Unfortunately, avoiding incorrect postures is not an easy task for nonexperts. It is difficult for people to view their own bodies from an objective perspective, especially while exercising. A widely accepted solution is to have someone else watch and provide feedback. However, nonexperts often do not have proper knowledge of correct postures, and hence, cannot provide helpful feedback. But experts can provide valuable feedback and increase the quality of workout sessions; however, they are in short supply and are often expensive. Moreover, under the current COVID-19 pandemic, in-person meetings and exercise sessions should be avoided.

Remote coaching is one of the most effective solutions for this problem. The recent surge in sales of home-fitness gear [12] supports such a change in exercise trends. Some people purchase equipment and subscribe to web-based services. For example, Nike+ Kinect Training (Nike) [13] is a fitness game for Xbox 360 (Microsoft Corp) that uses Microsoft Kinect, a depth-sensing camera device, and Weelo (Alyce Healthcare) [14] is a web subscription service that analyzes training forms using a laptop. Several studies have been conducted on computer-assisted postural correction. For example, Chen et al [15] built a system that analyzed workout posture using pose estimation; however, they used geometric methods and dynamic time warping using small amounts of data. Han et al [16] proposed using deep neural networks to analyze skeleton data extracted using Microsoft Kinect but did not actually show the implementation of the results. All these systems either require specific equipment or use heuristic methods that are difficult to generalize.

Some of the recent studies used deep learning–based exercise programs. Liao et al [17] developed a system that recognized rehabilitation exercises and Soro et al [18] used the deep learning model to classify several different exercises. However, none of them provided feedback about correct postures. In contrast, this study tried to imitate experts' visual judgments on correct exercise postures learned from years of experience through deep learning, which is well-known for extracting patterns from image and video data. Furthermore, while various non–face-to-face exercise methods using apps or services [12-16] have surged during the prolonged COVID-19 outbreak, studies on the effects of these methods are insufficient. Quantitative evaluations of posture correction and muscle strength improvements are limited as most studies focus on the qualitative effects of behavioral changes or weight loss [19].

This study aimed to use deep neural networks to design and develop a personal workout assistant capable of providing feedback on squat postures using only mobile devices such as smartphones. In the first part of this study, a squat video data set was created and a deep learning model using a combination of pose estimation and video classification was trained to analyze workout postures. In the second part, a mobile workout assistant app was developed and an interventional study was conducted to show how the app helps people improve squat exercise, in contrast to simply following videos over the internet.

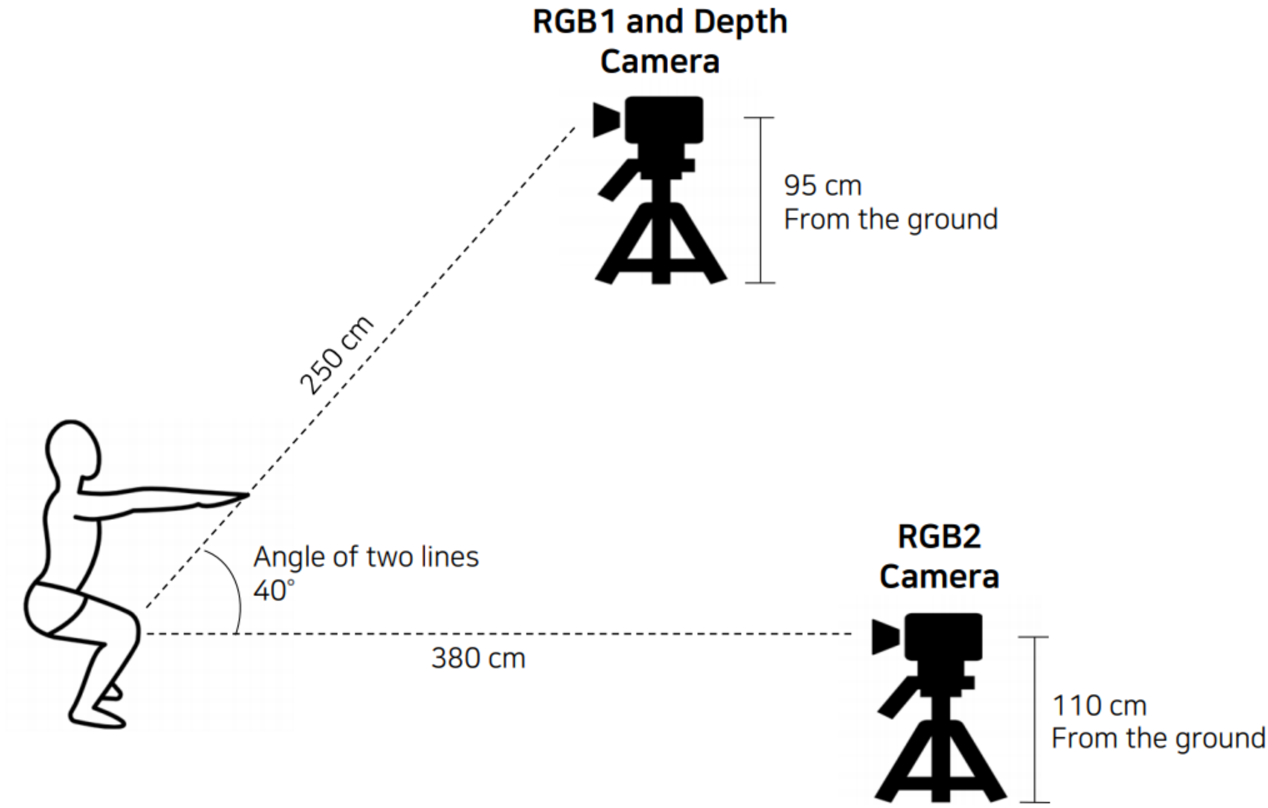
Methods

Part 1: Deep Learning Model

Data Collection

In the first part of the study, a squat posture data set was collected from participants to train the neural network model for a mobile app. Chosen participants were adults with no diagnosed musculoskeletal conditions and no pain in the ankles, knees, or lower back for the past 6 months. Those who could not perform the normal range of motion (ROM) for physiological reasons were excluded. A total of 52 participants (36 men and 16 women) were recruited, comprising a mix of 24 novices with little or no experience doing squats and 28 experts with extensive knowledge of correct workout postures. After receiving sufficient explanation about the experiment, participants signed a consent form. Each participant performed less than 100 squats in a single session and less than 200 squats in a single day, with sufficient breaks in between to minimize fatigue and possible injuries. Each participant performed an average of 400 squats, with each squat recorded individually. A Microsoft Kinect One was placed in front of the participant at a distance of 2.5 m and height of 0.95 m to capture the depth and red green blue (RGB) images; a laptop was placed at a distance of 3.8 m and height of 1.1 m to capture RGB images in a diagonal direction using its webcam (Figure 1). Both devices recorded at 30 fps. The EyesWeb program was used to synchronize the recording devices. Three workout experts were present during the recording to determine the correctness of each squat. Any disagreements were resolved by reviewing the video recordings. The problematic body parts were later labeled by experts through video analysis.

Figure 1. Camera placements during data collection. RGB: red green blue.



Data Preparation

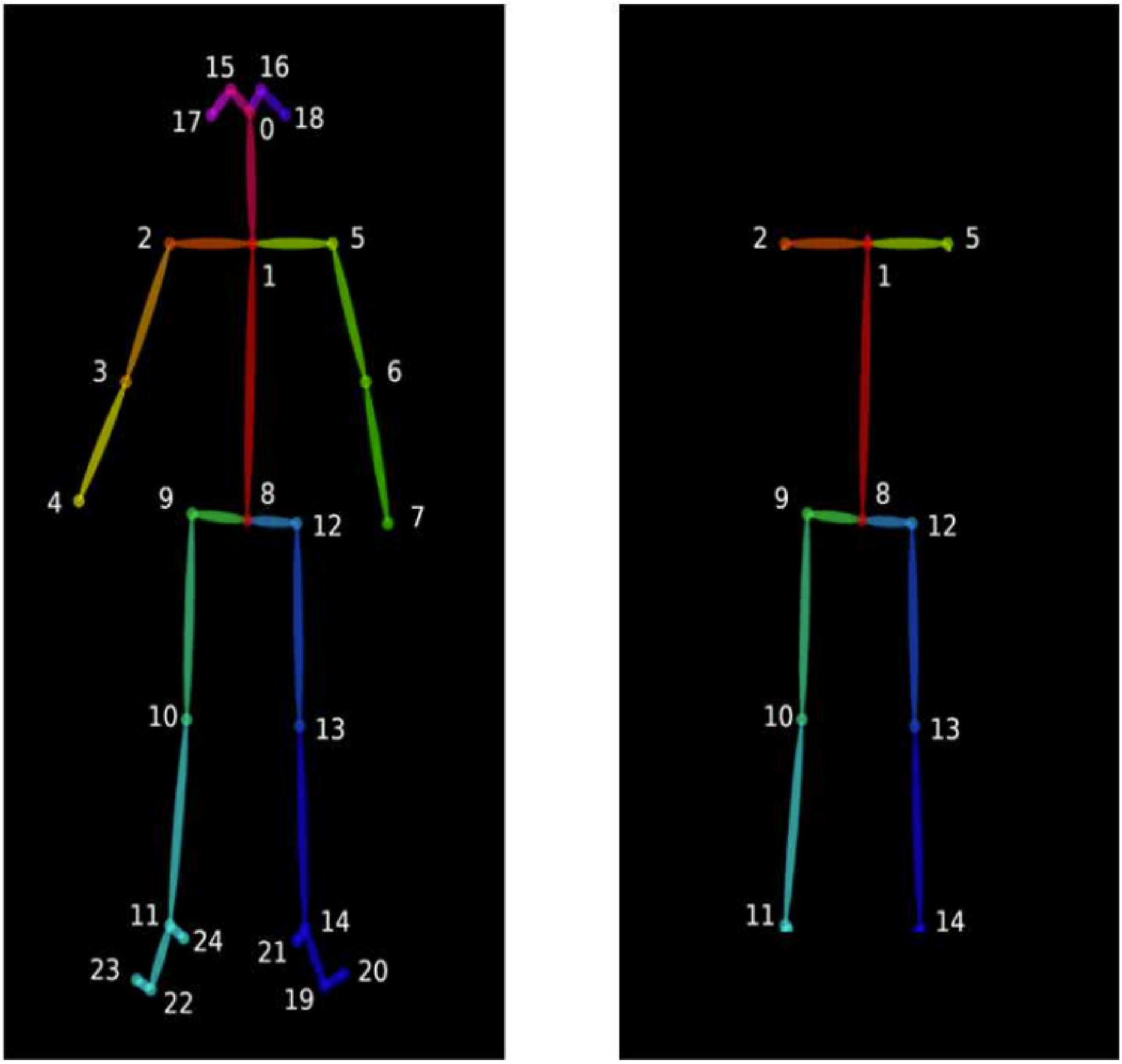
Of the collected data, only RGB videos recorded from the front were used to train the deep learning model in this study. This section describes the preprocessing steps performed to transform the videos into trainable data.

The videos recorded from the front camera were trimmed to accelerate the pose estimation process. OpenPose [20] was used to extract 2D skeleton key points from the videos, but only a certain portion of the output (see Figure 2) was used as some parts of the body, such as arms, head, and feet, do not have a significant impact on the correctness of the squat posture.

To normalize the key point data, a reference frame was selected from the early portion of the videos to ensure the highest chance that the target was upright and standing straight. Once the reference frame was chosen, a transform matrix that would map the coordinates of the hip joint key point to (0, 0) and the length of the torso to 1 in the reference frame was found and uniformly applied to all frames, ensuring that the coordinates of the key points would fall within a similar bound of values, thus normalizing data points from individuals with different heights. To mitigate the inevitable errors from pose estimation, a Gaussian filter was applied with an SD of 1 for each key point sequence, which smoothed out the movement of each key point, suppressing any unwanted sudden jumps. Finally, the per-frame changes in the normalized key point positions were calculated and used as input to the classification model. For example, if a person was detected in k frames of the video, the size of the input tensor to the model k was –1×20; if the person did not move throughout the video, the resulting input to the classification model was a tensor with all zeros.

The preprocessed key point data were then used as the input for the deep neural network, which classifies them as either correct or incorrect postures.

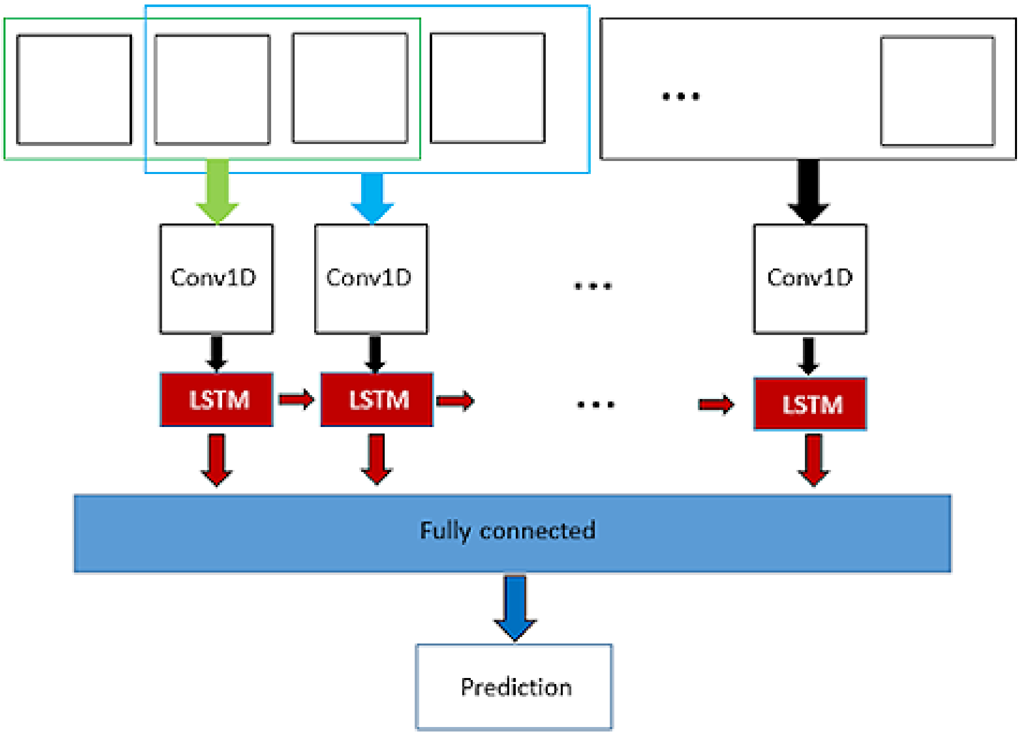
Figure 2. On the left: raw key point output. On the right: key points used for the training.



Architecture

The task of classifying squat postures can be viewed as a branch of the action recognition task in which videos are classified according to the actions performed by humans. It is better to understand the whole action instead of per-frame information; therefore, the temporal aspect of the data is crucial for the task. The concepts from long-term recurrent convolutional network [21] and convolutional 3D network [22] were combined to design a model with both temporal convolution layers and long short-term memory (LSTM) blocks. The architecture consisted of three 1D convolution layers, 1 bidirectional LSTM layer, and 1 fully connected layer, followed by a softmax layer. The convolution layers had temporal kernel depths of 5, 3, and 3. In other words, in the first convolution layer, 5 consecutive frames were convoluted to create new features; in the second and third convolution layers, 3 consecutive outputs were convoluted together. The number of filters on the convolution layers was 32, 64, and 64. The bidirectional LSTM block had 64 outputs, whereas the fully connected layer had 32 outputs. The model was trained using the Adam optimizer with a learning rate of 0.003 and a dropout rate of 0.3. These parameters were determined using a grid search, which yielded the highest validation accuracy of 0.9866. Figure 3 shows a simplified illustration of the proposed model architecture.

Figure 3. Proposed model architecture. LSTM: long short-term memory; Conv1D: 1D convolution layer.



Evaluation

The proposed deep learning model was evaluated using the test set described in section 3; the results are reported in Table 1. The model achieved an accuracy of 0.8498, a precision of 0.8713, and a recall of 0.8394. In addition to the model described in Section 4, models without convolution layers and without the LSTM blocks were trained and evaluated. As no suitable baseline performance for squat postures was found in prior work, the results of the Auto-ML Video Intelligence service of Google Cloud [23] were used as the baseline; although not designed for this task, it is still capable of classifying videos.

The results are presented in Table 1. The combination of LSTM blocks and temporal convolution layers exhibited the highest performance with a test accuracy of 85%. Convolution-only methods performed better (82% accuracy) than LSTM-only methods (78% accuracy), implying that the benefits of the temporal convolution layer are greater than those of the LSTM blocks. Google AutoML video intelligence service had the lowest accuracy (69%). The test accuracies were significantly lower than the training and validation accuracies for all 3 versions of the model, possibly due to overfitting on all 3 occasions, or it could mean that a large amount of the test set data is quite different from the data in the training set.

Table 1. Accuracy of the proposed model.

Model	Training accuracy	Validation accuracy	Test accuracy
LSTM[a] + Convolution	0.9929	0.9866	0.8498
LSTM only	0.8712	0.8532	0.7613
Convolution only	0.9546	0.9441	0.8203
Google Auto-ML	—[b]	—	0.6880

[a]LSTM: long short-term memory.

[b]Not available.

Part 2: Mobile Workout Assistant

Overview

After testing the proposed model, a mobile app, Home Alone Exercise (HALE), was designed and developed to assist users in learning squat exercises. In addition, an interventional study was conducted to demonstrate the effectiveness of the app by analyzing how the participants improved their exercises over time in terms of performance and joint ROM. This study was retrospectively registered on the Clinical Research Information Service (KCT0008178). The study design was unmodified after it began.

App Design

HALE was implemented for mobile devices, comprising 2 components: a server and a mobile app client. The server receives a single squat workout video each time, analyzes the posture through the pipeline, and responds with the classification score (the output of the softmax layer) along with the received video, but with the extracted key points rendered on top. The client allowed users to record their squat posture and send the recording to the server for analysis. The recording starts 3 seconds after the user touches the “start” button to give them time to position themselves correctly. The recording ended automatically after a certain amount of time configured by the user, and the recorded video was sent to the server. Visual and auditory cues are provided at the start and end of the recording session. When the client received a response from the server, the recorded video was shown to the user along with the extracted key points, and posture correctness was checked—such feedback was designed to guide the user to monitor one’s previous squats (eg, balance, joint angles, etc) so that they could learn the correct posture through self-discovery [24]. The user could configure the classification score threshold to adjust the difficulty of the workout session. The app’s best results were obtained when the device was placed at a height of approximately 130 cm (or just below the chest height) and far enough from the user’s workout spot to capture the user’s entire body. The user must face the camera, just like in the videos used in training. The server was implemented using Python and TensorFlow (Google Brain), and the client was implemented for Android devices using Android Studio (Google and JetBrains).

Recruitment

The volunteers for the test were recruited through a web-based forum on the Pusan National University web page. Preliminary screening was conducted over the phone using the Physical Activity Readiness Questionnaire for Everyone [25] to select individuals without any health issues. They were then asked whether they have learned squat exercise before and selected ones without any squat exercise experience. Volunteers who passed this preliminary selection underwent secondary in-person screening and were asked to perform 10 squats. Only those who received a “Good” mark from HALE less than 3 times and whose squat posture was considered incorrect by 3 experts (each with over 5 years of experience as a trainer with a master’s degree in sports science) were allowed to participate in the study. After receiving a sufficient explanation about the experiment, the selected participants signed a consent form. The participants were divided into experimental (EXP) and control groups (CTL) using block randomization. Out of the original 34 volunteers, 29 passed the preliminary screening, and 26 passed the second screening. Thus, the study began with 13 participants in each group (experimental and control); however, 3 participants from each group dropped out in each group. Consequently, the study had a total of 20 participants (20 women), 10 participants in each group: experimental (n=10) and control (n=10). None of the participants in the data set collection were allowed to participate in this part of the study. The physical characteristics of the participants in each group are described in Table 2.

Table 2. Physical characteristics of the participants per group.

Characteristics	EXP[a]	CTL[b]
Age (years), mean (SD)	21.90 (2.18)	22.60 (1.95)
Height (cm), mean (SD)	163.50 (6.00)	164.70 (5.07)
Weight (kg), mean (SD)	55.32 (4.40)	50.83 (4.74)
BMI (kg/m^2), mean (SD)	20.75 (2.11)	18.72 (1.23)
Muscle mass (kg), mean (SD)	36.87 (3.61)	35.52 (4.34)
Fat mass (kg), mean (SD)	16.03 (3.78)	12.99 (1.74)

[a]EXP: experimental group.

[b]CTL: control group.

Procedure

Isometric muscle function, muscle strength, and muscular endurance were measured using isokinetic equipment (Cybex 770, HUMAC NORM). Measurements were performed by measuring the extension and flexion values of the hip and knee joints. For muscle strength, angular velocity was measured at 60° per second 5 times, and muscular endurance was measured at 180° per second 15 times. The maximum torque and total workload of the knee and hip joints were also measured. All data were normalized to the participants’ body weight.

In addition, XSENS MTi-1 inertial measurement unit sensors (sampling at 100 Hz) were attached to identify 3D movements of the knee joints. The ROM of the left and right knee joints was calculated by combining the data from the inertial measurement unit sensors (Figure 4). Finally, participants’ body composition analysis results were recorded.

After the presession, each participant was given a smartphone with its app installed. The CTL was given a version of the app without feedback capabilities, whereas the EXP received the full version. Each group was asked to regularly practice squats using their respective apps for a predetermined amount of time, with their progress tracked through the logs left on the app server. The squat practice sessions were designed according to the American College of Sport Medicine guidelines as follows:

All participants were required to spend 30 minutes 5 days a week each session (5-minute warm-up, 20-minute main exercise, 5-minute cooldown). During the main exercise, participants performed squats for 1 minute and rested for 20-30 seconds. For the EXP, the last squat in each set was assessed as either “good” or “bad” during the resting period. The training sessions were recommended to be continuous for 30 min, but depending on the participant’s stamina and schedule, they could be split into multiple sessions; when that was the case, they were advised to have sessions no shorter than 10 min. Table 3 shows a breakdown of the training sessions.

Figure 4. Joint angle measurement using inertial measurement unit (IMU) sensors. Angle A on the rightmost image: knee joint angle.

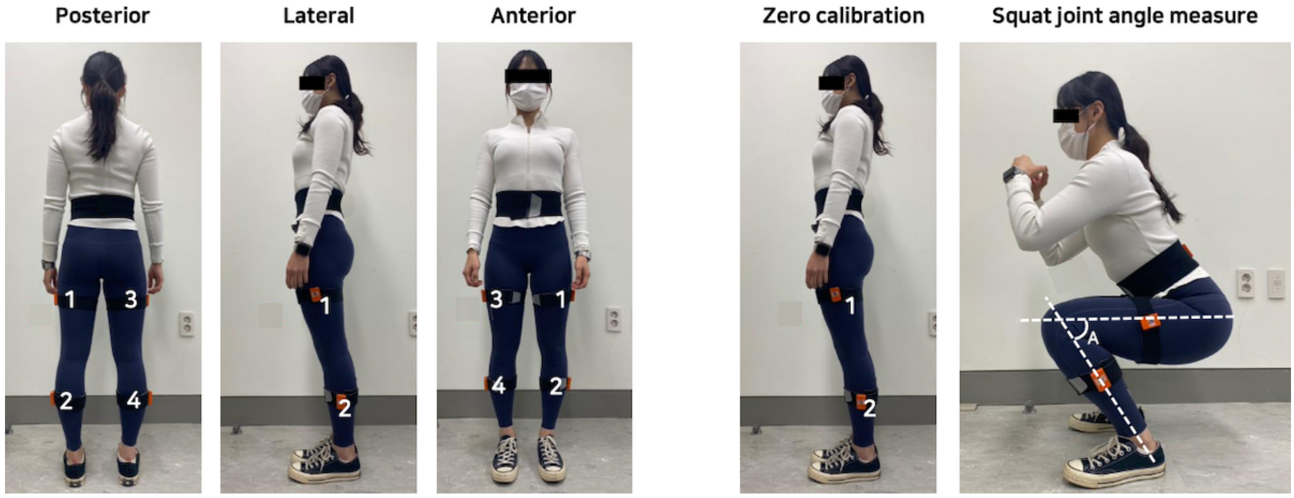


Table 3. App intervention program.

Order	Type	Time	Frequency
Warm-up	Stretching (neck, shoulder, waist, legs, and ankles)	5 minutes	5 days/week
Main exercise	(60-second squat work out + ≈20-30–second break) × 15 sets	≈20-22.5 minutes	5 days/week
Cooldown	Stretching (neck, shoulder, waist, legs, and ankles)	5 minutes	5 days/week

HALE was designed to send a notification if the participant did not open the app for longer than 24 hours, with the progress of each participant checked through the logs on the server. A check-in call was given on the following day whenever a participant failed to complete the training session on any given day.

After 2 weeks of training and the final recording session, all the smartphones were collected from the participants to collect and analyze data.

Statistical Analysis

Statistical analyses were performed on the 3 measures—the overall squat score assessed by HALE, the knee joint ROM, and the isokinetic muscle strength. Two-way mixed ANOVA [26] was used to analyze the interaction effects. Subsequently, 1-way repeated-measures ANOVA was applied for each group to analyze the effects of sessions per group. If any significant differences were detected, a post hoc comparison analysis using the least significant difference was performed.

Ethics Approval

After receiving sufficient explanation about the first part of the experiment, participants signed a consent form approved by the Pusan University Bioethics Committee (PNU IRB/2019_38_HR), and after receiving sufficient explanation about the second part of the experiment, the selected participants signed a consent form approved by the Pusan University Bioethics Committee (PNU IRB/2020_137_HR).

Results

As a result of Mauchly normality test [27], all items showed a normal distribution, but the CTL App test did not show normality; therefore, Greenhouse-Geisser function was used. The squat score measured by HALE was evaluated, and the results are presented in Table 4. The results of the repeated measures ANOVA for each group showed that there was a significant difference ($F_{2,18}$=11.174, P=.001) in squat score for the EXP, while no significant effect was observed for the CTL ($F_{2,18}$=2.675, P=.13). The post hoc analysis revealed that the participants performed significantly better during both the mid- and Post sessions than during the presession (P=.006 and P=.001, respectively). No interaction was found on the 2-way mixed ANOVA.

The results of the repeated-measures ANOVA for each group are shown in Table 5. For the EXP, there were significant differences in the left ($F_{2,18}$=5.811, P=.02) and right ($F_{2,18}$=4.736, P=.03) knee joint angles, while no significant effect was observed for the CTL in the left ($F_{2,18}$=0.398, P=.68) and right ($F_{2,18}$=0.517, P=.61) knee joint angles. Post hoc analysis revealed that the left knee joint angle in the postsession was smaller than that in the pre and midsessions (P=.048 and P=.02, respectively). In addition, the joint angle of the right knee in the postsession was smaller than that in the midsession (P=.02). No significant interaction was observed when 2-way mixed ANOVA was performed.

The changes in isokinetic muscle strength measured during each session were also analyzed. Regardless of the group, the participants' muscle strength and endurance of the trunk and knees either increased significantly or showed a tendency to increase.

Table 4. Results of the overall squat score measured by Home Alone Exercise (HALE).

Session	Squat posture, mean (SD)	Post hoc	1–β	ES[a]	95% CI	*F*(*df*; *P* value)
Group EXP[b] (N=10)		Pre<Mid (*P*=.006);Pre<Post (*P*=.001)	.945	0.554		11.174 (2,18; .001)
Pre	0.20 (0.42)				–0.102 to 0.502	
Mid	4.20 (3.43)				1.750 to 6.650	
Post	8.00 (5.20)				4.275 to 11.725	
Group CTL[c] (N=10)		NS[d]	.284	0.229		2.675 (2,18; .13)
Pre	0.70 (1.63)				–0.471 to 1.871	
Mid	1.30 (2.50)				–0.486 to 3.086	
Post	3.80 (5.43)				–0.086 to 7.686	

[a]ES: effect size.
[b]EXP: experimental group.
[c]CTL: control group.
[d]NS: nonsignificant; significance level at .05.

Table 5. Results of the joint angles.

Body	Session	Joint angle, mean (SD)	Post hoc	1–β	ES[a]	95% CI	*F* (*df*; *P* value)
Group EXP[b] (N=10)							
Right knee			Mid>Post (*P*=.018)	.722	0.904		4.736 (2,18; .03)
	Pre	71.99 (20.79)				54.612-89.383	
	Mid	76.68 (15.16)				64.011-89.364	
	Post	62.82 (14.88)				50.379-75.261	
Left knee			Pre>Post (*P*=.048); Mid<Post (*P*=.02)	.822	0.454		5.811 (2,18; .02)
	Pre	75.06 (22.16)				56.534-93.599	
	Mid	76.24 (19.37)				60.042-92.441	
	Post	63.13 (13.06)				52.213-74.064	
Group CTL[c] (N=10)							
Right knee			NS[d]	.061	0.054		0.517 (2,18; .61)
	Pre	70.82 (22.71)				54.579-87.071	
	Mid	74.02 (17.55)				61.470-86.582	
	Post	70.23 (15.42)				59.200-81.274	
Left knee			NS	.056	0.042		0.398 (2,18; .68)
	Pre	73.27 (23.15)				56.711-89.937	
	Mid	74.05 (17.81)				5.635-61.307	
	Post	70.70 (16.34)				5.169-59.015	

[a]ES: effect size.
[b]EXP: experimental group.
[c]CTL: control group.
[d]NS: nonsignificant; significance level at .05.

Discussion

Principal Results

The results of the isokinetic muscle strength test indicate that all the participants, regardless of their group, performed the squats as instructed. However, while they all gained muscle, only the participants in the EXP increased the ROM of the knee joint by 12.8% in the right leg and 15.9% in the left leg. In contrast, the ROM of the soft joint in the CTL increased by 0.9% in the right leg and by 3.6% in the left leg. The EXP showed significant improvement in the wider advanced squat technique compared to the CTL [28,29]. In addition, the results of the overall squat score and the joints' ROM indicated that the participants who used HALE effectively improved their squat skills, whereas those who did not use HALE showed no significant improvements. The ratio of squats in the correct posture increased by 52.3% in the EXP and 21.3% in the CTL, indicating that the developed app worked properly and was effective.

Moreover, while the exercise app only provided pass-or-fail feedback along with a skeleton overlaid video playback, it effectively guided participants to learn small details about squat exercises, such as how far they had to sit down to properly perform a full squat exercise. In other words, when an automatic evaluation and self-discovery [30] system is combined with physical training programs, people can effectively develop exercise skills without direct instruction. Such findings and approaches utilizing a deep learning–based evaluation and feedback process, are expected to benefit various communities seeking to develop effective exercise programs that can be remotely taught and learned.

Furthermore, to provide advanced feedback and guidance, the size and complexity of both data sets and deep learning models must increase drastically. As the deep learning model only was needed to determine whether a trainee correctly performed an exercise, the otherwise complex problem involving 3D motion guidance became as easy as a binary classification problem. In other words, the deep learning model was able to be trained with high accuracy within a short period of time using approximately 2000 squat video data taken from the front, which simplified data labeling, processing time, and effort. In addition, the study results indicate that the participants effectively improved their squats, even though the proposed solution did not provide detailed feedback or specific guidance for improvement. Such a simplified training method possesses great potential to boost web-based or remote exercise platforms, cover various exercises, and can be expanded to remote communities.

Limitations

Different from conventional training systems, HALE does not provide direct instructions. However, while such an approach may initially be mistaken as more difficult to learn, the study results indicate that the users experienced more effective workout sessions using the app even without direct instructions. In fact, the indirect feedback was able to help the users to engage and learn by themselves about the correct postures as they tried to improve their scores, suggesting that machine learning models can be trained effectively as training tools, even with minimal data and simple architectures. Choosing more complex model architectures, collecting more data to improve the model, or providing more detailed instructions to users might be viable options, but those approaches could be too costly compared to the benefits and do not even guarantee better teaching effects.

In addition, HALE might lack robustness; the user must place the smartphone at a predetermined distance, height, and angle, or the app cannot correctly assess their squat. However, this issue can be resolved by using the recently developed 3D pose estimation algorithms [31]. The current version of HALE only supports squats; however, it can be extended to other exercises by collecting new data sets and training new models.

Unfortunately, it was difficult to find more appropriate participants who had little or no experience with squats because most men had some exercise experience. However, while further investigation on additional men might be beneficial, the difference is likely to be marginal, as evidenced by prior work, which indicated no significant difference between men's and women's exercise efficacy [32,33].

Conclusions

This study demonstrated the effectiveness of a deep learning-based personal workout assistant that can provide feedback on squat postures using only mobile devices. In the first part of the study, the squat video data set was created and the deep learning model, which showed a test accuracy of 85%, was trained. In the second part, the mobile workout assistant app, HALE, was developed and the interventional study showed how it helped people improve squat exercise. As demonstrated by the improvements in the squat posture and joint ROM, the EXP trained with HALE experienced faster improvement and learned more nuanced details of the squat exercise. The proposed mobile app that is low cost and provides self-discovery feedback effectively taught users about squat exercises without expensive in-person sessions with a trainer.

Conflicts of Interest

None declared.

Editorial Notice

This randomized study was only retrospectively registered, explained by authors as an oversight in the methodology. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials, because the risk of bias appears low and the study was considered formative as a trial of feasibility. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 405 KB - ijmr_v12i1e37604_app1.pdf]

References

1. Lee SW, Lee J, Moon SY, Jin HY, Yang JM, Ogino S, et al. Physical activity and the risk of SARS-CoV-2 infection, severe COVID-19 illness and COVID-19 related mortality in South Korea: a nationwide cohort study. Br J Sports Med 2022;56(16):901-912 [FREE Full text] [doi: 10.1136/bjsports-2021-104203] [Medline: 34301715]
2. Castañeda-Babarro A, Arbillaga-Etxarri A, Gutiérrez-Santamaría B, Coca A. Physical activity change during COVID-19 confinement. Int J Environ Res Public Health 2020;17(18):6878 [FREE Full text] [doi: 10.3390/ijerph17186878] [Medline: 32967091]
3. Cruz-Jentoft AJ, Sayer AA. Sarcopenia. Lancet 2019;393(10191):2636-2646 Erratum in: Lancet. 2019 Jun 29;393(10191):2590. [doi: 10.1016/S0140-6736(19)31138-9] [Medline: 31171417]
4. Kirwan R, McCullough D, Butler T, de Heredia FP, Davies IG, Stewart C. Sarcopenia during COVID-19 lockdown restrictions: long-term health effects of short-term muscle loss. Geroscience 2020;42(6):1547-1578 [FREE Full text] [doi: 10.1007/s11357-020-00272-3] [Medline: 33001410]
5. Narici M, De Vito G, Franchi M, Paoli A, Moro T, Marcolin G, et al. Impact of sedentarism due to the COVID-19 home confinement on neuromuscular, cardiovascular and metabolic health: physiological and pathophysiological implications and recommendations for physical and nutritional countermeasures. Eur J Sport Sci 2021;21(4):614-635. [doi: 10.1080/17461391.2020.1761076] [Medline: 32394816]
6. Bae EJ, Kim YH. Factors affecting sarcopenia in Korean adults by age groups. Osong Public Health Res Perspect 2017;8(3):169-178 [FREE Full text] [doi: 10.24171/j.phrp.2017.8.3.03] [Medline: 28781939]
7. Pongchaiyakul C, Limpawattana P, Kotruchin P, Rajatanavin R. Prevalence of sarcopenia and associated factors among Thai population. J Bone Miner Metab 2013;31(3):346-350. [doi: 10.1007/s00774-013-0422-4] [Medline: 23377622]
8. Polero P, Rebollo-Seco C, Adsuar JC, Pérez-Gómez J, Rojo-Ramos J, Manzano-Redondo F, et al. Physical activity recommendations during COVID-19: narrative review. Int J Environ Res Public Health 2020;18(1):65 [FREE Full text] [doi: 10.3390/ijerph18010065] [Medline: 33374109]
9. Selseth A, Dayton M, Cordova ML, Ingersoll CD, Merrick MA. Quadriceps concentric EMG activity is greater than eccentric EMG activity during the lateral step-up exercise. J Sport Rehabil 2000;9(2):124-134. [doi: 10.1123/jsr.9.2.124]
10. Wallace BJ, Kernozek TW, Mikat RP, Wright GA, Simons SZ, Wallace KL. A comparison between back squat exercise and vertical jump kinematics: implications for determining anterior cruciate ligament injury risk. J Strength Cond Res 2008;22(4):1249-1258 [FREE Full text] [doi: 10.1519/JSC.0b013e31816d66a4] [Medline: 18545181]
11. Diggin D, O'Regan C, Whelan N, Daly S, McLoughlin V, McNamara L, et al. A biomechanical analysis of front versus back squat: injury implications. Portuguese J Sports Sci 2011;11(Suppl 2):643-646 [FREE Full text]
12. Shaban H. The pandemic's home-workout revolution may be here to stay. The Washington Post. 2021. URL: https://www.washingtonpost.com/road-to-recovery/2021/01/07/home-fitness-boom/ [accessed 2023-08-04]
13. Nike+ Kinect training. AKQA. URL: https://www.akqa.com/work/nike/kinect/ [accessed 2023-08-04]
14. Weelo. URL: https://weelo.fit/land [accessed 2023-08-04]
15. Chen S, Yang R. Pose trainer: correcting exercise posture using pose estimation. ArXiv. Preprint posted online on June 21 2020 2020 [FREE Full text] [doi: 10.48550/arXiv.2006.11718]
16. Han SH, Kim HG, Choi JH. Rehabilitation posture correction using deep neural network. 2017 Presented at: 2017 IEEE International Conference on Big Data and Smart Computing (BigComp); 13-16 February 2017; Jeju, Korea (South) p. 400-402. [doi: 10.1109/bigcomp.2017.7881743]
17. Liao Y, Vakanski A, Xian M. A deep learning framework for assessing physical rehabilitation exercises. IEEE Trans Neural Syst Rehabil Eng 2020;28(2):468-477. [doi: 10.1109/tnsre.2020.2966249]
18. Soro A, Brunner G, Tanner S, Wattenhofer R. Recognition and repetition counting for complex physical exercises with deep learning. Sensors (Basel) 2019;19(3):714 [FREE Full text] [doi: 10.3390/s19030714] [Medline: 30744158]
19. Joo SY, Lee CB, Joo NY, Kim CR. Feasibility and effectiveness of a motion tracking-based online fitness program for office workers. Healthcare (Basel) 2021;9(5):584 [FREE Full text] [doi: 10.3390/healthcare9050584] [Medline: 34068929]
20. Cao Z, Simon T, Wei S, Sheikh Y. Realtime multi-person 2D pose estimation using part affinity fields. 2017 Presented at: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 21-26 July 2017; Honolulu, HI, USA p. 1302-1310. [doi: 10.1109/cvpr.2017.143]
21. Dnahue J, Hendricks LA, Rohrbach M, Venugopalan S, Guadarrama S, Saenko K, et al. Long-term recurrent convolutional networks for visual recognition and description? ArXiv. Preprint posted online on May 31 2016 2014 [FREE Full text] [doi: 10.48550/arXiv.1411.4389]

22. Tran D, Bourdev L, Fergus R, Torresani L, Paluri M. Learning spatiotemporal features with 3D convolutional networks. 2015 Presented at: 2015 IEEE International Conference on Computer Vision (ICCV); 07-13 December 2015; Santiago, Chile. [doi: 10.1109/iccv.2015.510]
23. Video AI. Google Cloud. URL: https://cloud.google.com/video-intelligence [accessed 2023-08-04]
24. Olugbade TA, Bianchi-Berthouze N, Marquardt N, de C Williams AC. Human observer and automatic assessment of movement related self-efficacy in chronic pain: from exercise to functional activity. IEEE Trans Affective Comput 2020;11(2):214-229. [doi: 10.1109/taffc.2018.2798576]
25. American College of Sports Medicine. In: Riebe D, Liguori G, Ehrman JK, Magal M, editors. ACSM's Guidelines for Exercise Testing and Prescription, 10th Edition. Philadelphia, Pennsylvania: Wolters Kluwer; 2018.
26. Field A. GLM 5: Mixed designs. In: Discovering Statistics Using IBM SPSS Statistics, 5th Edition. Los Angeles: Sage Publications; 2018:704-733.
27. Field A. Correlation. In: Discovering Statistics Using IBM SPSS Statistics, 5th Edition. Los Angeles: Sage Publications; 2018:340-356.
28. Kitamura T, Kido A, Ishida Y, Kobayashi Y, Tsukamoto S, Tanaka Y. Muscle activity pattern with a shifted center of pressure during the squat exercise. J Sports Sci Med 2019;18(2):248-252 [FREE Full text] [Medline: 31191094]
29. Slater LV, Hart JM. Muscle activation patterns during different squat techniques. J Strength Cond Res 2017;31(3):667-676 [FREE Full text] [doi: 10.1519/JSC.0000000000001323] [Medline: 26808843]
30. Wagensveld B, Segers E, Kleemans T, Verhoeven L. Child predictors of learning to control variables via instruction or self-discovery. Instr Sci 2015;43(3):365-379. [doi: 10.1007/s11251-014-9334-5]
31. Bartol K, Bojanic D, Petkovic T, D'Apuzzo N, Pribanic T. A review of 3D human pose estimation from 2D images. 2020 Presented at: Proceedings of 3DBODY.TECH 2020 11th Int. Conference and Exhibition on 3D Body Scanning and Processing Technologies; 17-18 Nov. 2020; Online/Virtual p. 29 URL: https://www.3dbody.tech/cap/abstracts/2020/2029bartol.html [doi: 10.15221/20.29]
32. DiStefano LJ, Padua DA, DiStefano MJ, Marshall SW. Influence of age, sex, technique, and exercise program on movement patterns after an anterior cruciate ligament injury prevention program in youth soccer players. Am J Sports Med 2009;37(3):495-505. [doi: 10.1177/0363546508327542] [Medline: 19251685]
33. Barha CK, Davis JC, Falck RS, Nagamatsu LS, Liu-Ambrose T. Sex differences in exercise efficacy to improve cognition: a systematic review and meta-analysis of randomized controlled trials in older humans. Front Neuroendocrinol 2017;46:71-85. [doi: 10.1016/j.yfrne.2017.04.002] [Medline: 28442274]

Abbreviations

CTL: control group
EXP: experimental group
HALE: home alone exercise
LSTM: long short-term memory
RGB: red green blue
ROM: range of motion

Edited by G Eysenbach, T Leung; submitted 28.02.22; peer-reviewed by B Lai, A Videira-Silva; comments to author 13.07.22; revised version received 08.11.22; accepted 02.08.23; published 12.09.23.

Please cite as:
Chae HJ, Kim JB, Park G, O'Sullivan DM, Seo J, Park JJ
An Artificial Intelligence Exercise Coaching Mobile App: Development and Randomized Controlled Trial to Verify Its Effectiveness in Posture Correction
Interact J Med Res 2023;12:e37604
URL: https://www.i-jmr.org/2023/1/e37604
doi:10.2196/37604
PMID:37698913

Original Paper

Prediction of Male Coronary Artery Bypass Grafting Outcomes Using Body Surface Area Weighted Left Ventricular End-diastolic Diameter: Multicenter Retrospective Cohort Study

Zhihui Zhu[1,2*], MD; Yuehuan Li[1*], MD; Fan Zhang[1], MD; Stefanie Steiger[2], PhD; Cheng Guo[3], PhD; Nan Liu[1], MD; Jiakai Lu[1], MD; Guangpu Fan[4], MD; Wenbo Wu[5], MD; Mingying Wu[5], MD; Huaibin Wang[6], MD; Dong Xu[7], MD; Yu Chen[4], MD; Junming Zhu[1], MD; Xu Meng[1], MD; Xiaotong Hou[1], MD; Hans-Joachim Anders[2], MD; Jian Ye[8], MD; Zhe Zheng[9], MD; Chenyu Li[2*], MD; Haibo Zhang[1*], MD

[1]Department of Cardiac Surgery, Beijing Anzhen Hospital, Capital Medical University, Beijing, China
[2]Medizinische Klinik und Poliklinik IV, Hospital of the Ludwig-Maximilians-University, Munich, Germany
[3]Allianz Technology, Allianz, Munich, Germany
[4]Department of Cardiovascular Surgery, Peking University People's Hospital, Beijing, China
[5]Department of Cardiovascular Surgery, Beijing Tongren Hospital, Capital Medical University, Beijing, China
[6]Department of Cardiovascular Surgery, Beijing Hospital, Beijing, China
[7]Department of Cardiovascular Surgery, Beijing Xuanwu Hospital, Capital Medical University, Beijing, China
[8]Department of Cardiovascular Surgery, St. Paul's Hospital, University of British Columbia, Vancouver, BC, Canada
[9]Department of Cardiovascular Surgery, Fuwai Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing, China
[*]these authors contributed equally

Corresponding Author:
Haibo Zhang, MD
Department of Cardiac Surgery, Beijing Anzhen Hospital
Capital Medical University
2 Anzhen Road, Chaoyang District
Beijing, 100029
China
Phone: 86 13370103561
Fax: 86 64456884
Email: zhanghb2318@163.com

Abstract

Background: The presence of a high left ventricular end-diastolic diameter (LVEDD) has been linked to a less favorable outcome in patients undergoing coronary artery bypass grafting (CABG) procedures. However, by taking into consideration the reference of left ventricular size and volume measurements relative to the patient's body surface area (BSA), it has been suggested that the accuracy of the predicting outcomes may be improved.

Objective: We propose that BSA weighted LVEDD (bLVEDD) is a more accurate predictor of outcomes in patients undergoing CABG compared to simply using LVEDD alone.

Methods: This study was a comprehensive retrospective cohort study that was conducted across multiple medical centers. The inclusion criteria for this study were patients who were admitted for treatment between October 2016 and May 2021. Only elective surgery patients were included in the study, while those undergoing emergency surgery were not considered. All participants in the study received standard care, and their clinical data were collected through the institutional registry in accordance with the guidelines set forth by the Society of Thoracic Surgeons National Adult Cardiac Database. bLVEDD was defined as LVEDD divided by BSA. The primary outcome was in-hospital all-cause mortality (30 days), and the secondary outcomes were postoperative severe adverse events, including use of extracorporeal membrane oxygenation, multiorgan failure, use of intra-aortic balloon pump, postoperative stroke, and postoperative myocardial infarction.

Results: In total, 9474 patients from 5 centers under the Chinese Cardiac Surgery Registry were eligible for analysis. We found that a high LVEDD was a negative factor for male patients' mortality (odds ratio 1.44, $P<.001$) and secondary outcomes. For female patients, LVEDD was associated with secondary outcomes but did not reach statistical differences for morality. bLVEDD showed a strong association with postsurgery mortality (odds ratio 2.70, $P<.001$), and secondary outcomes changed in parallel with bLVEDD in male patients. However, bLVEDD did not reach statistical differences when fitting either mortality or severer outcomes in female patients. In male patients, the categorical bLVEDD showed high power to predict mortality (area under the curve [AUC] 0.71, $P<.001$) while BSA (AUC 0.62) and LVEDD (AUC 0.64) both contributed to the risk of mortality but were not as significant as bLVEDD ($P<.001$).

Conclusions: bLVEDD is an important predictor for male mortality in CABG, removing the bias of BSA and showing a strong capability to accurately predict mortality outcomes.

Trial Registration: ClinicalTrials.gov NCT02400125; https://clinicaltrials.gov/ct2/show/NCT02400125

(Interact J Med Res 2023;12:e45898) doi:10.2196/45898

KEYWORDS

body surface area; BSA; left ventricular end-diastolic diameter; LVEDD; coronary artery bypass grafting; CABG; outcomes

Introduction

Coronary artery disease causes angina pectoris, myocardial infarction, and ischemic heart failure and thereby contributes significantly to the cardiovascular disease being the leading cause of death worldwide [1-3]. Coronary artery bypass graft (CABG) surgery is the gold-standard treatment in many patients with complex multivessel coronary artery disease or left main disease [4,5]. Left ventricular enlargement is a powerful predictor of adverse outcomes such as all-cause death, cardiovascular death, heart failure hospitalization, and outcomes of cardiac surgery [6-9]. For those undergoing CABG, enlarged left ventricular end-diastolic diameter (LVEDD) is most commonly associated with ischemic cardiomyopathy and is known to increase the risk of postoperative adverse events [10,11]. A more detailed study on the effects of LVEDD is needed to add more evidence about the relationship between the LVEDD and perioperative prognosis in CABG.

Body surface area (BSA) is a simple calculation based on the patients' height and weight [12]. In contemporary cardiovascular care, BSA is used to normalize cardiac output to cardiac index, calculate glomerular filtration rate, is positively associated with blood pressure, and has been shown to be a relatively accurate representation of total body water [13-16]. The American Society of Echocardiography and the European Association of Cardiovascular Imaging have recommended using BSA to normalize echocardiographic parameters such as right ventricle size and left ventricle size [17]. Here, we propose to use the BSA to normalize the end-diastolic volume and to remove the bias of the BSA on LVEDD; thus, the BSA weighted LVEDD (bLVEDD) is defined as LVEDD divided by BSA. To this end, we intended to investigate the role of LVEDD, BSA, and bLVEDD in a specific clinical setting, and to evaluate whether the relationship among them after CABG will facilitate patient-specific care in cardiac surgery.

Hence, this study investigated the respective effect of LVEDD, BSA, and bLVEDD on early clinical outcomes in patients undergoing CABG by using clinical data from 5 top cardiac centers in China (Beijing Anzhen Hospital, Beijing Tongren Hospital, Beijing Hospital, Peking University People's Hospital, and Beijing Xuanwu Hospital) under the Chinese Cardiac Surgery Registry database, to reveal (1) the effects of LVEDD on the perioperative prognosis, (2) the relationship between BSA and LVEDD, and (3) whether bLVEDD was associated with perioperative complications and mortality in patients undergoing CABG.

Methods

Study Setting and Population

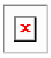
This study was a multicenter retrospective analysis of observational data. A total of 9474 inpatients across the nation (Multimedia Appendix 1) from 5 top cardiac centers in China under the Chinese Cardiac Surgery Registry database, admitted between October 2016 and May 2021, were included in this study. All patients were elective surgery patients, and emergency surgery patients were excluded. Clinical data were obtained via the institutional registry following the Society of Thoracic Surgeons National Adult Cardiac Database. The accuracy and completeness of these data were ensured through multiple procedures, which have been described previously [18].

Ethical Considerations

The study protocol had been approved by the Ethics Committee of Fuwai Hospital (Approval No. 2017-943). The study is registered at ClinicalTrials.gov (NCT02400125). To protect patient privacy, all patient data were deidentified (ie, patient names were replaced with the identification code, and all private patient information was deleted before analysis). The Peking University Clinical Research Institute has created a data committee to evaluate the data quality and supervise data collection. All patients were treated with standard care, and no additional intervention was performed as described previously [18].

Predictor and Outcomes

Patient demographics and clinical characteristics were collected and analyzed. This included the patient's past cardiovascular medical history (peripheral vascular disease, previous cerebrovascular event, previous myocardial infarction [MI], and previous percutaneous coronary intervention and New York Heart Association classification. The last preoperative test results of serum creatinine, total serum cholesterol, serum low-density lipoprotein, blood glucose, and estimated glomerular filtration rate (eGFR) were acquired. The patient's previous echocardiogram before surgery was also analyzed for LVEDD and left atrial dimension. Intraoperative factors, such as cardiopulmonary bypass time, and aortic cross-clamp time were also analyzed. Variables for concomitant cardiac drugs (ie, nitrate lipid drugs, catecholamines, β-blockers, angiotensin-converting enzyme inhibitor, angiotensin receptor blocker, statins, aspirin, clopidogrel, and ticagrelor) were documented as comprehensively as possible. The primary outcome was in-hospital all-cause mortality (30 days). The secondary outcomes were postoperative severe adverse events, including use of extracorporeal membrane oxygenation, multiorgan failure, use of intra-aortic balloon pump, postoperative stroke, and postoperative MI. The BSA is calculated as follows [12]:



Statistical Analysis

Variables with missing values or outliers warranted interpolation by multiple imputations using the MICE package [19]. Since the database was structurally designed and supervised by data committee, the missing values or outliers were less than 2% across all the indicators. We assumed that the data were missing or misrecording at random [20]; therefore, we performed predictive mean matching [21] to generate 5 complete imputed data sets that fit the logistic models. For multivariate logistic regression, we selected age, gender, smoking within 2 weeks before surgery, diabetes, hypertension, hyperlipidemia, last test of serum creatinine, total cholesterol, low-density lipoprotein, blood glucose, preoperative eGFR before surgery, and previous cerebrovascular events for adjustment based on clinical experience. The bLVEDD was optimally binned based on the weight of evidence binning by supervised tree-like segmentation; the process of generating the bin and threshold of the bLVEDD was follow by the reference pipeline of scorecard package [22]. The coefficient was calculated by Spearman correlation method. Categorical variables were compared using the chi-square test or Fisher exact test. Cochran-Armitage trend test was used for trend analysis. Continuous variables were compared using a 2-tailed *t* test or the Mann-Whitney *U* test. The area under the curve (AUC) of the receiver operating characteristics was compared using the DeLong method. The sample size calculation showed that an estimated 639 patients with a bLVEDD of ≥31.5 would be needed to provide 99% power for detecting a minimum clinically meaningful mortality rate of 5.23% with a 2-side α of .05 when compared with patients with a bLVEDD of <31.5. All analyses were performed using R version 3.4.2 (The R Project for Statistical Computing).

Results

Baseline

In total, 9474 patients were eligible for the final analysis, of which 7232 (76.34%) were male and 2242 (23.66%) were female. Among female patients, the mean age was 65.26 (SD 7.49) years, and 1032 (46.03%) had an LVEDD of <46 mm. In the male patients, mean age was 61.78 (SD 9.00) years, and 3615 (49.99%) had an LVEDD of <50 mm. Moreover, male patients with an LVEDD of <50 mm had a higher rate of smoking and comorbidity of hyperlipidemia, abnormal serum total cholesterol, and creatinine; a higher risk of previous MI; and a higher score of New York Heart Association (*P*<.05). Female patients with an LVEDD of <46 mm had more hypertension and previous MI and abnormal serum total cholesterol (*P*<.05). Both male and female patients with higher LVEDD had low left ventricular ejection fraction and left atrial dimension. Therefore, the preoperative baseline condition of patients with high LVEDD was worse than that of patients with low LVEDD (Table 1).

Table 1. Patient characteristics according to LVEDD[a] category.

Characteristics	Gender							
	Female				Male			
	Total	LVEDD<46	LVEDD≥46	*P* value	Total	LVEDD<50	LVEDD≥50	*P* value
Total, n	2242	1032	1210	N/A[b]	7232	3615	3617	N/A
Age (years), mean (SD)	65.26 (7.49)	65.29 (7.6)	65.24 (7.39)	.87	61.78 (9.00)	62.12 (8.74)	61.43 (9.23)	.01
BMI (kg/m^2), mean (SD)	25.34 (3.41)	24.76 (3.26)	25.82 (3.47)	<.001	25.77 (3.06	25.38 (3.04)	26..05 (3.05)	<.001
BSA[c], mean (SD)	1.66 (0.14)	1.63 (0.13)	1.68 (0.14)	<.001	1.87 (0.14)	1.85 (0.14)	1.89 (0.14)	<.001
Smoking[d], n (%)	193 (8.61)	85 (8.24)	108 (8.93)	.56	4075 (56.35)	1953 (54.02)	2122 (58.67)	<.001
Diabetes, n (%)	1027 (45.81)	473 (45.83)	554 (45.79)	.98	2708 (37.44)	1317 (36.43)	1391 (38.46)	.08
Hypertension, n (%)	1604 (71.54)	713 (69.09)	891 (73.64)	.02	4401 (60.85)	2189 (60.55)	2212 (61.16)	.60
Hyperlipidemia, n (%)	762 (33.99)	347 (33.62)	415 (34.3)	.74	2457 (33.97)	1289 (35.66)	1168 (32.29)	.003
Peripheral vascular disease, n (%)	73 (3.26)	32 (3.1)	41 (3.39)	.70	252 (3.48)	120 (3.32)	132 (3.65)	.44
Previous cerebrovascular event, n (%)	303 (13.51)	139 (13.47)	164 (13.55)	.95	986 (13.63)	500 (13.83)	486 (13.44)	.62
Previous MI[e], n (%)	311 (13.87)	134 (12.98)	177 (14.63)	<.001	1331 (18.40)	503 (12.81)	828 (22.89)	<.001
Previous PCI[f], n (%)	270 (12.04)	119 (11.53)	151 (12.48)	.49	1021 (14.12)	485 (13.42)	536 (14.82)	.86
NYHA[g], n (%)								
NYHA1	1801 (80.33)	831 (80.52)	970 (80.17)	.29	5595 (77.36)	2717 (75.16)	2878 (79.57)	<.001
NYHA2	1311 (58.47)	623 (60.37)	688 (56.86)	—[h]	4084 (56.47)	2015 (55.74)	2069 (57.2)	—
NYHA3	459 (20.47)	195 (18.9)	264 (21.82)	—	1423 (19.68)	677 (18.73)	746 (20.62)	—
NYHA4	31 (1.38)	13 (1.26)	18 (1.49)	—	88 (1.22)	25 (0.69)	63 (1.74)	—
Serum creatinine (umol/L)[i], mean (SD)	63.58 (20.44)	63.21 (21.18)	63.89 (19.79)	.43	78.02 (22.23)	76.77 (20.19)	79.26 (24.04)	<.001
Serum total cholesterol (mmol/L), mean (SD)	4.23 (1.04)	4.28 (1.02)	4.19 (1.05)	.04	3.92 (0.97)	3.95 (0.97)	3.9 (0.97)	.03
Serum low-density lipoprotein, mean (SD)	2.51 (0.86)	2.53 (0.85)	2.49 (0.86)	.19	2.34 (0.81)	2.35 (0.81)	2.34 (0.81)	.74
eGFR[j] (mL/min/1.73m^2), mean (SD)	99.91 (11.76)	100.14 (11.7)	99.71 (11.81)	.38	93.88 (10.95)	94.08 (10.5)	93.69 (11.39)	.12
Blood glucose (mmol/L), mean (SD)	6.65 (2.02)	6.57 (1.96)	6.72 (2.08)	.08	6.44 (2.09)	6.4 (1.92)	6.48 (2.25)	.11
LVEF[k], mean (SD)	61.65 (8.25)	63.34 (6.81)	60.22 (9.07)	<.001	59.31 (9.08)	62.38 (6.6)	56.24 (10.13)	<.001
LAD[l] (mm), mean (SD)	34.83 (7.94)	34 (7.18)	35.53 (8.48)	<.001	36.52 (7.78)	35.4 (6.92)	37.63 (8.4)	<.001
LVEDD (mm), mean (SD)	46.32 (4.96)	42.28 (2.55)	49.75 (3.78)	<.001	49.99 (5.9)	45.46 (3.09)	54.52 (4.38)	<.001
Normalized by weight, mean (SD)	0.75 (0.12)	0.7 (0.11)	0.78 (0.13)	<.001	0.68v(0.12)	0.63 (0.1)	0.73 (0.11)	<.001
Normalized by BMI, mean (SD)	1.86 (0.3)	1.74 (0.25)	1.96 (0.3)	<.001	1.97 (0.32)	1.79 (0.24)	2.09 (0.3)	<.001
Normalized by BSA, mean (SD)	28.05 (3.43)	26.09 (2.49)	29.73 (3.23)	<.001	26.85 (3.54)	24.72 (2.45)	28.99 (3.15)	<.001
Nitrate lipid drugs[m], n (%)	547 (24.4)	232 (22.48)	315 (26.03)	.51	1696 (23.45)	787 (21.77)	909 (25.13)	<.001
Catecholamines[n], n (%)	15 (0.67)	6 (0.58)	9 (0.74)	.64	33 (0.46)	17 (0.47)	16 (0.44)	.86

Characteristics	Gender							
	Female				Male			
	Total	LVEDD<46	LVEDD≥46	*P* value	Total	LVEDD<50	LVEDD≥50	*P* value
β-blockers[o], n (%)	1860 (82.96)	866 (83.91)	994 (82.15)	.27	6011 (83.12)	2974 (82.27)	3037 (83.96)	.05
ACEI[p] or ARB[q,r], n (%)	539 (24.04)	230 (22.29)	309 (25.54)	.07	1489 (20.59)	704 (19.47)	785 (21.7)	.02
Statins, n (%)	1519 (67.75)	708 (68.6)	811 (67.02)	.42	4942 (68.34)	2460 (68.05)	2482 (68.62)	.60
Aspirin[s], n (%)	672 (29.97)	295 (28.59)	377 (31.16)	.19	2234 (30.89)	1016 (28.11)	1218 (33.67)	<.001
Clopidogrel, n (%)	145 (6.47)	74 (7.17)	71 (5.87)	.21	547 (7.56)	276 (7.63)	271 (7.49)	.82
Ticagrelor, n (%)	86 (3.92)	46 (4.6)	40 (3.36)	.14	343 (4.84)	188 (5.37)	155 (4.33)	.04

[a]LVEDD: left ventricular end-diastolic diameter.
[b]N/A: not applicable.
[c]BSA: body surface area.
[d]Smoking within 2 weeks before surgery.
[e]MI: myocardial infarction.
[f]PCI: percutaneous coronary intervention.
[g]NYHA: New York Heart Association.
[h]Not available.
[i]Serum creatinine, serum total cholesterol, serum low-density lipoprotein, eGFR, blood glucose, LVEF, LVEDD, and LAD are the last tests before surgery.
[j]eGFR: estimated glomerular filtration rate.
[k]LVEF: left ventricular ejection fraction.
[l]LAD: left atrial dimension.
[m]Nitrate lipid drugs are administered intravenously 24 hours before surgery.
[n]Catecholamines are administered intravenously 48 hours before surgery.
[o]β-blockers and statins are administered orally 24 hours before surgery.
[p]ACEI: angiotensin-converting enzyme inhibitor.
[q]ARB: angiotensin receptor blocker.
[r]ACEI or ARB are administered orally 48 hours before surgery.
[s]Aspirin, clopidogrel, and ticagrelor are administered orally 5 days before surgery.

High LVEDD is a Negative Prognostic Factor for Male Patients

As shown in Table 2, both male and female patients with high LVEDD yielded lower eGFR (*P*<.001). Male patients with high LVEDD had longer mechanical ventilation duration, initial intensive care unit length of stay, higher serum creatinine, more use of intra-aortic balloon pump and extracorporeal membrane oxygenation, and higher mortality (*P*<.001), while in female patients, these do not reach statistical difference (Table 2). As a regression result, high LVEDD was a negative factor for male patients' mortality (adjusted odds ratio [OR] 1.44, 1.33-1.56, *P*<.001; Table 3, Figure 1B), indicating that each increase in the patient's LVEDD fifths classification increased the odd of mortality by 44%. Similarly, male patients with high LVEDD had more secondary adverse events (adjusted OR 1.19, 1.16-1.23, *P*<.001; Table 3, Figure 1B) by increasing the odd by 16% for each increase in the LVEDD classification. For female patients, the LVEDD was associated with secondary outcomes (OR 1.13, 1.07~1.19, *P*=.03), but did not reach statistical difference for morality (Multimedia Appendix 2). Therefore, high LVEDD is a negative prognostic factor for both postoperative survival and severe events in male patients.

Table 2. Patient outcomes according to left ventricular end-diastolic diameter (LVEDD) category.

Characteristics	Gender							
	Female				Male			
	Total	LVEDD<46	LVEDD≥46	*P* value	Total	LVEDD<50	LVEDD≥50	*P* value
Perioperative blood transfusion, n (%)	1672 (74.58)	763 (73.93)	909 (75.12)	.52	4740 (65.54)	2393 (66.2)	2347 (64.89)	.24
Mechanical ventilation duration (hour), mean (SD)	26.22 (29.58)	25.57 (28.2)	26.77 (30.7)	.34	24.43 (26.9)	23.05 (23.29)	25.8 (30.03)	<.001
Initial ICU[a] length of stay (hour), mean (SD)	37.79 (39.99)	36.5 (38.19)	38.89 (41.45)	.16	36.45 (39.83)	32.9 (33.39)	39.99 (45.08)	<.001
Perioperative blood loss (ml), mean (SD)	915.1 (809.39)	899.53 (809.24)	928.38 (809.62)	.40	1087.39 (900.03)	1081.72 (929.07)	1093.05 (870.12)	.59
Serum creatinine[b] (umol/L), mean (SD)	78.45 (43.49)	77.08 (45.72)	79.62 (41.48)	.17	91.7 (40.37)	89.9 (38.2)	93.51 (42.37)	<.001
eGFR[c] (mL/min/1.73m2), mean (SD)	105.15 (34.46)	107.53 (34.24)	103.12 (34.53)	.003	87.6 (28.3)	88.64 (27.82)	86.55 (28.74)	.002
AKI[d], n (%)	313 (13.96)	136 (13.18)	177 (14.63)	.32	715 (9.89)	348 (9.63)	367 (10.15)	.46
Use of IAPB[e], n (%)	155 (6.91)	65 (6.3)	90 (7.44)	.29	522 (7.22)	208 (5.75)	314 (8.68)	<.001
Use of ECMO[f], n (%)	18 (0.8)	9 (0.87)	9 (0.74)	.73	53 (0.73)	28 (0.77)	25 (0.69)	.67
Multiorgan failure, n (%)	31 (1.38)	12 (1.16)	19 (1.57)	.41	64 (0.88)	25 (0.69)	39 (1.08)	.08
Reoperation, n (%)	49 (2.19)	24 (2.33)	25 (2.07)	.90	151 (2.09)	73 (2.02)	78 (2.16)	.58
Postoperative MI[g], n (%)	10 (0.45)	4 (0.39)	6 (0.5)	.70	64 (0.88)	29 (0.8)	35 (0.97)	.45
Postoperative stroke, n (%)	33 (1.47)	14 (1.36)	19 (1.57)	.68	64 (0.88)	30 (0.83)	34 (0.94)	.62
Reintubation, n (%)	21 (0.94)	9 (0.87)	12 (0.99)	.77	77 (1.06)	22 (0.61)	55 (1.52)	<.001
Re-enter ICU, n (%)	45 (2.01)	16 (1.55)	29 (2.4)	.15	118 (1.63)	56 (1.55)	62 (1.71)	.58
Dead, n (%)	54 (2.41)	23 (2.23)	31 (2.56)	.61	107 (1.48)	33 (0.91)	74 (2.05)	<.001

[a]ICU: intensive care unit.

[b]Serum creatinine is the maximum serum creatinine after surgery.

[c]eGFR: estimated glomerular filtration rate. This is the minimum eGFR after surgery.

[d]AKI: acute kidney injury.

[e]IAPB: intra-aortic balloon pump.

[f]ECMO: extracorporeal membrane oxygenation.

[g]MI: myocardial infarction.

Table 3. Adjusted and unadjusted logistic regression model of the association between body surface area weighted left ventricular end-diastolic diameter (bLVEDD) and prognosis of male patients[a].

Variables	Secondary outcomes					Mortality				
	Univariate		Multivariate		AUC[b]	Univariate		Multivariate		AUC
	OR[c]	*P* value	OR	*P* value		OR	*P* value	OR	*P* value	
bLVEDD[d]										
Numerical bLVEDD	1.13 (1.12~1.15)	<.001	1.12 (1.1~1.13)	<.001	0.61	1.21 (1.18~1.24)	<.001	1.18 (1.15~1.21)	<.001	0.71
Categorized bLVEDD	1.76 (1.66~1.86)	<.001	1.64 (1.55~1.73)	<.001	0.59	3.08 (2.69~3.53)	<.001	2.7 (2.36~3.1)	<.001	0.71
<22.5	0.73 (0.57~0.92)	.17	0.65 (0.51~0.82)	0.07	N/A[e]	0 (0~0)	.98	0 (0~0)	.98	N/A
(22.5, 26)	1	N/A	1	N/A	N/A	1	N/A	1	N/A	N/A
(26, 31.5)	1.39 (1.26~1.53)	<.001	1.32 (1.19~1.45)	<.001	N/A	2.7 (2.07~3.52)	<.001	2.44 (1.87~3.19)	<.001	N/A
(31.5, infinity)	3.6 (3.2~4.06)	<.001	3 (2.65~3.39)	<.001	N/A	8.63 (6.52~11.43)	<.001	6.54 (4.9~8.73)	<.001	N/A
LVEDD[f]										
Numerical LVEDD	1.07 (1.06~1.07)	<.001	1.06 (1.05~1.07)	<.001	0.59	1.1 (1.09~1.12)	<.001	1.09 (1.08~1.11)	<.001	0.65
Categorized LVEDD	1.22 (1.18~1.26)	<.001	1.19 (1.16~1.23)	<.001	0.58	1.48 (1.36~1.6)	<.001	1.44 (1.33~1.56)	<.001	0.64
<45	1	N/A	1	N/A	N/A	1	N/A	1	N/A	N/A
(45, 48)	1.18 (1.01~1.39)	.30	1.32 (1.12~1.55)	.09	N/A	1.93 (1.19~3.13)	.18	2.33 (1.43~3.79)	.08	N/A
(48, 51)	1.17 (0.98~1.38)	.37	1.25 (1.05~1.49)	.2	N/A	2.03 (1.23~3.35)	.16	2.3 (1.39~3.8)	.10	N/A
(51, 54)	1.24 (1.07~1.44)	.15	1.37 (1.18~1.6)	.04	N/A	2.33 (1.47~3.68)	.07	2.85 (1.8~4.53)	.02	N/A
(54, infinity)	2.29 (1.99~2.64)	<.001	2.23 (1.93~2.58)	<.001	N/A	5.45 (3.53~8.41)	<.001	5.58 (3.61~8.64)	<.001	N/A
BSA[g]										
Numerical BSA	0.3 (0.23~0.41)	<.001	0.36 (0.27~0.49)	<.001	0.54	0.07 (0.04~0.14)	<.001	0.11 (0.06~0.23)	<.001	0.61
Categorized BSA	0.88 (0.85~0.91)	<.001	0.9 (0.87~0.93)	.002	0.54	0.72 (0.67~0.77)	<.001	0.76 (0.7~0.82)	<.001	0.62
<1.75	1	N/A	1	N/A	N/A	1	N/A	1	N/A	N/A
(1.75, 1.83)	0.75 (0.64~0.87)	.06	0.75 (0.64~0.88)	.06	N/A	0.68 (0.5~0.92)	.20	0.7 (0.52~0.96)	.25	N/A
(1.83, 1.91)	0.67 (0.58~0.77)	<.001	0.68 (0.58~0.79)	.01	N/A	0.54 (0.4~0.73)	.04	0.57 (0.42~0.78)	.07	N/A
(1.91, 1.99)	0.6 (0.52~0.7)	<.001	0.63 (0.54~0.73)	.003	N/A	0.35 (0.25~0.48)	<.001	0.39 (0.28~0.55)	<.001	N/A
(1.99, infinity)	0.56 (0.48~0.65)	<.001	0.6 (0.52~0.71)	<.001	N/A	0.26 (0.19~0.37)	<.001	0.33 (0.23~0.47)	<.001	N/A

[a]Age, gender, smoking within 2 weeks before surgery, diabetes, hypertension, hyperlipidemia, last test of serum creatinine before surgery, last test of serum total cholesterol before surgery, last test of serum low-density lipoprotein before surgery, last test of blood glucose before surgery, use of cardiopulmonary bypass, preoperative estimated glomerular filtration rate, and previous cerebrovascular events were used for the multivariate regression. bLVEDD was categorized into 4 groups based on a weight of tree-like segmentation binning.

[b]AUC: area under the curve.

[c]OR: odds ratio.

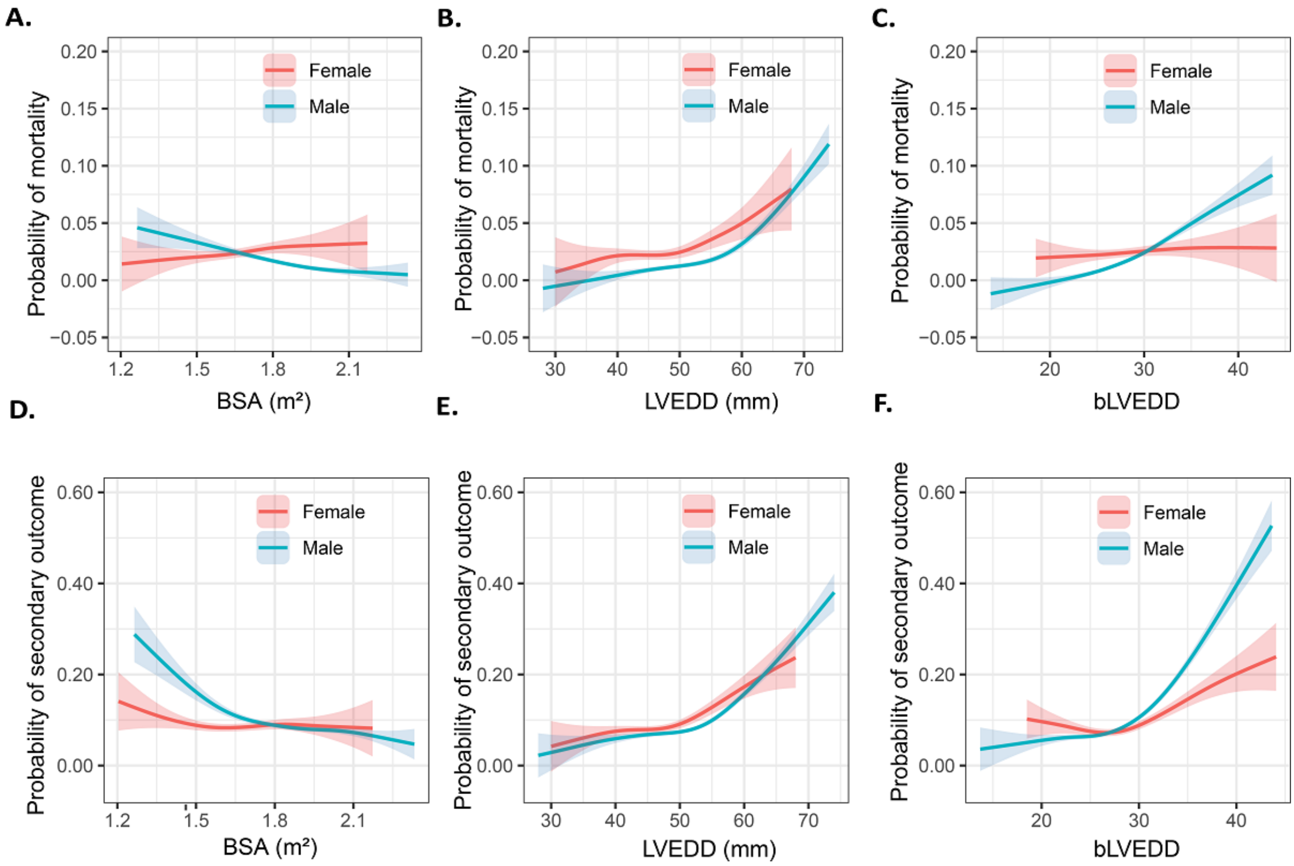
[d]bLVEDD is LVEDD divided by BSA.

[e]N/A: not applicable.

[f]LVEDD: left ventricular end-diastolic diameter.

[g]BSA: body surface area.

Figure 1. Generalized additive models of body surface area (BSA), left ventricular end-diastolic diameter (LVEDD), and body surface area weighted left ventricular end-diastolic diameter (bLVEDD) for primary and secondary outcomes. (A-C) Probability of mortality. (D-F) Secondary outcomes using restricted cubic splines.



LVEDD Normalized for BSA

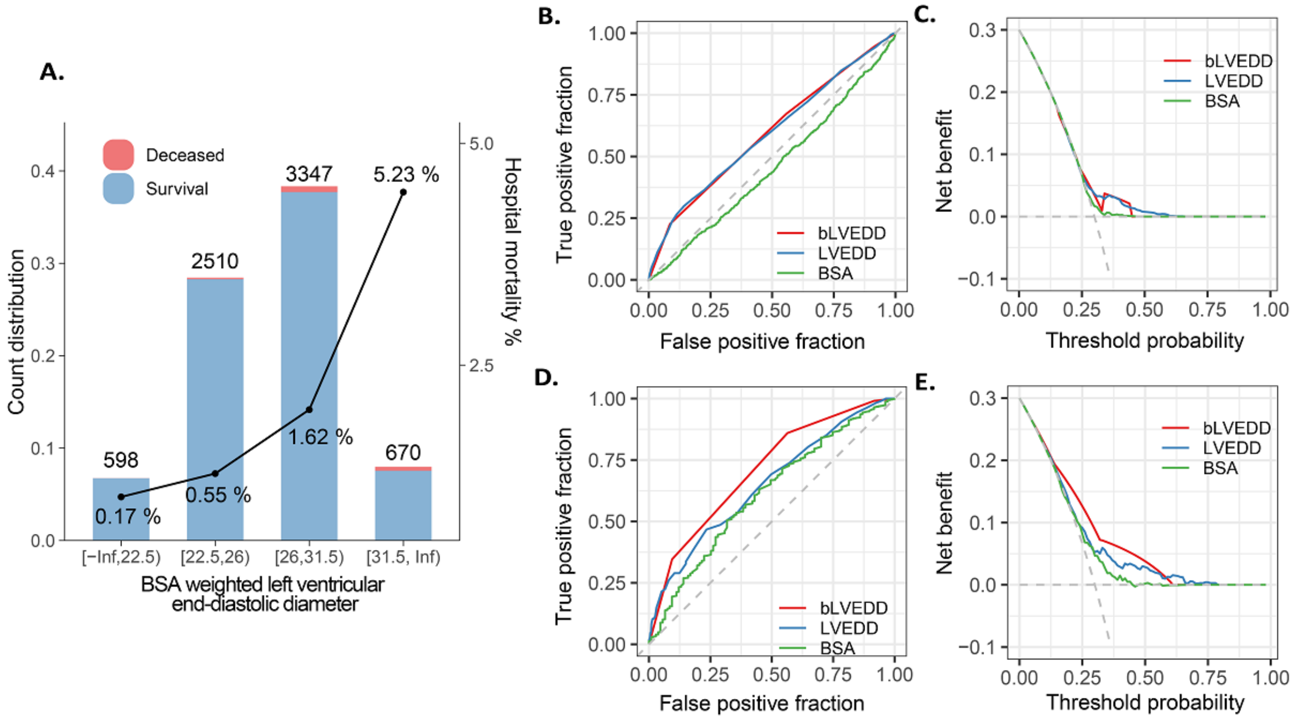
In our cohort, the mean BSA was 1.66 m^2 (SD 0.14 m^2) for female and 1.87 m^2 (SD 0.14 m^2) for male patients, which showed a slightly positive relationship (*R*=0.20 for female and *R*=0.15 for male patients, *P*<.001; Multimedia Appendix 3) with the LVEDD and reached statistical difference when compared in high and low LVEDD groups (*P*<.001). In male patients, high BSA was significantly associated with mortality (adjusted OR 0.76, 0.70-0.82, *P*<.001; Table 3, Figure 1A) and secondary outcomes (adjusted OR 0.90, 0.87-0.93, *P*<.001; Table 3, Figure 1D). However, in female patients, the BSA was not associated with either mortality or secondary outcomes.

Since the LVEDD has not been analyzed together with BSA in patients in CABG previously, we speculated that bLVEDD, defined as LVEDD divided by BSA, could better predict postsurgery prognosis. To identify whether bLVEDD increases the risk of postoperative mortality and secondary outcomes as well as whether it is the better predictor of outcomes, both univariate and multivariate logistic regression analyses were performed. As a result, in male patients, bLVEDD showed a strong association with postsurgery mortality; that is, the risk of mortality (adjusted OR 2.70, 2.39-3.10, *P*<.001; Table 3, Figure 1E). Secondary outcomes (adjusted OR 1.64, 1.55-1.73, *P*<.001; Table 3, Figure 1F) changed in parallel with a rise in bLVEDD, suggesting that bLVEDD represents a predictor for mortality and secondary outcomes. However, the bLVEDD did not reach statistical difference when fitting either mortality or severer outcomes in female patients.

The bLVEDD is a Robust Indicator for Mortality in Male Patients

To make bLVEDD more practical for male patients undergoing CABG, a weight of tree-like segmentation was used to binning bLVEDD to a categorical variable. As a result, our data set generated a categorization of (0, 22.5), (22.5, 26), (26, 31.5), (31.5, infinity; adjusted Kolmogorov-Smirnov 0.38, *P*=.01; Figure 2A, Multimedia Appendix 4). The categorical bLVEDD showed a quite similar power to a numerical form to predict mortality (AUC 0.71, *P*<.001). However, compared with BSA and LVEDD, bLVEDD was the most effective variable that fitted with mortality. The compositions of bLVEDD, such as BSA (AUC 0.62) and LVEDD (AUC 0.64), all slightly contributed to the risk of mortality with low AUC (DeLong test, *P*<.001) but were not as significant as that of bLVEDD (Figure 2B-E). Importantly, male patients with a bLVEDD of <31.5 faced higher mortality risk than those with a bLVEDD of ≥31.5 (OR 5.09, 4.14~6.26, *P*<.001; Figure 2D and E).

Figure 2. Segmentation of bLVEDD and its ability to predict clinical outcome in male patients. (A) Supervised tree-like segmentation of bLVEDD; (B) receiver operating characteristics; (C) precision-recall for secondary outcomes; and (D,E) motility. bLVEDD: body surface area weighted left ventricular end-diastolic diameter; BSA: body surface area; LVEDD: left ventricular end-diastolic diameter.



Discussion

Principal Findings

In this multicenter cohort study, we reported the following: (1) a high LVEDD is a negative prognostic factor for both postoperative survival and secondary outcomes in male patients; (2) in male patients, high BSA was significantly associated with mortality and secondary outcomes, while female patients' BSA is not associated with either mortality or secondary outcomes; (3) bLVEDD showed a strong association with postsurgery mortality; that is, the risk of mortality and secondary outcomes changed in parallel with bLVEDD increasing in male patients, but female patients' bLVEDD did not reach statistical difference; (4) A bLVEDD of 31.5 is the threshold to categorize male patients undergoing CABG with a high or low risk of mortality.

It is known that severe left ventricular dysfunction is associated with increased mortality in patients undergoing CABG [23-25]. Left ventricular function was usually described as the ejection fraction (EF) [26-28]; however, it is unclear whether EF is the most meaningful index of left ventricular function in a CABG situation. The left ventricular ejection fraction is the fraction of the end-diastolic volume that is ejected with each beat; that is, stroke volume divided by end-diastolic volume. Low EF may be caused by poor contractile function duo to extensive myocardial damage, or infarct expansion and stretching of the myocardial scar [29]. Thus, LVEDD might be a more meaningful predictor than EF, which is merely an arithmetical term based on 2 values. Zhu et al [10] reported that left ventricular geometry was an independent and incremental prognostic factor for death in patients undergoing CABG. Yan et al [11] found that left ventricular hypertrophy and left ventricular enlargement were associated with an increased risk of postoperative mortality after CABG in patients with heart failure with reduced ejection fraction. Categorizing left ventricular structural patterns with left ventricular hypertrophy and left ventricular enlargement contributes to risk stratification and provides incremental predictive ability. In our study, we also found that patients with high LVEDD had a poor baseline and suffered from more comorbidities, and LVEDD is an adverse prognostic factor for both postoperative survival and secondary outcomes in male patients undergoing CABG, which is consistent with previous studies.

Echocardiography is widely used in the diagnosis of cardiac diseases, especially for patients undergoing cardiac surgery. The measurement of the size of the left ventricle should be a part of every echocardiography report, because it provides diagnostic clues and prognostic information and enables the clinician to follow up with patients in respect of disease progression [30,31]. In clinical practice, surgeons often evaluate echocardiographic indicators using unstandardized absolute values; however, the structural characteristics of the heart should be related to human body measures such as height and weight under normal physiological conditions. Simply evaluating the absolute value of the left ventricle is not conducive to an accurate diagnosis of cardiac disease [32]. BSA is a critical index of physiologic functions, and it is used in several medical disciplines, including cardiology, oncology, burn management, and nephrology [33]. Some studies reported that the left ventricular diameter is a relatively crude and simplified assessment of a 3D structure, which cannot consider more complex variations in ventricular shape or size [34]. Using BSA to normalize echocardiographic parameters has been recommended by guidelines [17]. In the recommendations for chamber quantification, orthogonal long-axis views and the Simpson biplane method allow a more accurate calculation of the left ventricular volume, which may be corrected for size by normalizing to BSA [35]. LVEDD is data within echocardiography and is seen when the ventricle is the largest, shortly before the mitral valve closes and the mitral annulus descends. LVEDD is recognized as a negative risk factor for CABG. However, the definition of the normal range of the LVEDD is based on the entire population, including groups with different clinical characteristics. In our study, we found that the larger the LVEDD, the higher the perioperative mortality, even in the normal range. This may lead to misjudgment by cardiac surgeons, who believe that heart function is relatively safe when CABG is performed in patients with a "normal" LVEDD. Therefore, the risk factor of the left ventricle should be considered comprehensively when predicting the perioperative outcomes of CABG, and LVEDD should be normalized to remove the bias. We used BSA normalized LVEDD more accurately to predict mortality in patients undergoing CABG.

There is evidence to suggest gender inequality in CABG [36]. Studies have shown that female patients are at a higher risk of short-term mortality and other complications after surgery [37,38]. In our study, we also found that female patients have higher mortality compared to male patients undergoing CABG (2.41% vs 1.48%, *P*<.001). Some studies showed that the type of diffuse coronary disease is more commonly seen in female patients, which may be a contributing factor to poor outcomes [39,40]. Furthermore, some cardiovascular risk factors, such as diabetes and smoking, have a severer influence among female patients [39]. It is suggested that the baseline characteristics and unequal risk factors are the reasons for the difference in the outcomes after CABG [36]. For the LVEDD, guidelines also suggested that female and male patients have different normal ranges [17]. In our results, the bLVEDD also showed a disparity between different genders. In male patients undergoing CABG, bLVEDD showed a strong association with postsurgery mortality, while in female patients, the bLVEDD did not reach statistical difference when fitting either mortality or severer outcomes. Therefore, the gender disparity observed in bLVEDD should also be further studied in larger cohorts.

Limitations

There are some limitations in this study. First, our study is a retrospective cohort study. The data collection of patients in the past is limited, the preoperative activity tolerance of patients is difficult to obtain, the follow-up time was long, and the rate of loss to follow-up was high. Second, intractable heart failure and atrial fibrillation are also common complications after CABG, but they were not included in this study because they were difficult to record accurately during follow-up. Third, BSA is an empirical formula based on weight and height and cannot directly give the true numerical value of the human surface. Especially in this context, BSA is confounded with age, gender, race, etc; thus, a further study is needed to study the factors that have collinearity with BSA.

Conclusions

The bLVEDD is an important predictor for male mortality in CABG, removing the bias of BSA and showing a strong capability to accurately predict mortality outcomes. In predicting perioperative outcomes of CABG, it is important to comprehensively consider the risk factor of left ventricular enlargement and normalize LVEDD by BSA to eliminate bias in male patients. This research highlights significant benefits for enhancing the treatment standards of cardiac surgery and increasing the survival rate of patients following CABG.

Acknowledgments

This work is supported by the Beijing Science and Technology Commission of China (grant D171100002917001 and D171100002917003). The funder had no role in the study design, data collection, data analysis, data interpretation, or writing of the manuscript.

Data Availability

The data sets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Regional distribution of enrolled patients.

[PNG File , 247 KB - ijmr_v12i1e45898_app1.png]

Multimedia Appendix 2

Adjusted and unadjusted logistic regression model of the association between body surface area weighted left ventricular end-diastolic diameter (bLVEDD) and prognosis of female patients.
[DOCX File , 19 KB - ijmr_v12i1e45898_app2.docx]

Multimedia Appendix 3
Scatter plot and density distribution of left ventricular end-diastolic diameter (LVEDD) and body surface area (BSA).
[PNG File , 388 KB - ijmr_v12i1e45898_app3.png]

Multimedia Appendix 4
Supervised tree-like segmentation of bLVEDD and its evaluation and validation. AUC: Area Under the Curve; F-1: F1 Score; FPR: False Positive Rate; K-S: Kolmogorov-Smirnov; P-R: Precision-Recall; ROC: Receiver Operating Characteristic; TPR: True Positive Rate.
[PNG File , 229 KB - ijmr_v12i1e45898_app4.png]

References

1. Hu S, Zheng Z, Yuan X, Wang Y, Normand ST, Ross JS, et al. Coronary artery bypass graft: contemporary heart surgery center performance in China. Circ Cardiovasc Qual Outcomes 2012 Mar 01;5(2):214-221 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.111.962365] [Medline: 22396587]
2. Doenst T, Haverich A, Serruys P, Bonow RO, Kappetein P, Falk V, et al. PCI and CABG for Treating Stable Coronary Artery Disease: JACC Review Topic of the Week. J Am Coll Cardiol 2019 Mar 05;73(8):964-976 [FREE Full text] [doi: 10.1016/j.jacc.2018.11.053] [Medline: 30819365]
3. Hao Y, Liu J, Liu J, Yang N, Smith SC, Huo Y, et al. Sex Differences in In-Hospital Management and Outcomes of Patients With Acute Coronary Syndrome. Circulation 2019 Apr 09;139(15):1776-1785. [doi: 10.1161/CIRCULATIONAHA.118.037655] [Medline: 30667281]
4. Caliskan E, de Souza DR, Böning A, Liakopoulos OJ, Choi Y, Pepper J, et al. Saphenous vein grafts in contemporary coronary artery bypass graft surgery. Nat Rev Cardiol 2020 Mar;17(3):155-169. [doi: 10.1038/s41569-019-0249-3] [Medline: 31455868]
5. Hillis LD, Smith PK, Anderson JL, Bittl JA, Bridges CR, Byrne JG, et al. 2011 ACCF/AHA Guideline for Coronary Artery Bypass Graft Surgery: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. Circulation 2011 Dec 06;124(23):e652-e735. [doi: 10.1161/CIR.0b013e31823c074e] [Medline: 22064599]
6. Nestico PF, Hakki AH, Iskandrian AS. Left ventricular dilatation. Prognostic value in severe left ventricular dysfunction secondary to coronary artery disease. Chest 1985 Aug;88(2):215-220. [doi: 10.1378/chest.88.2.215] [Medline: 2990823]
7. McManus DD, Shah SJ, Fabi MR, Rosen A, Whooley MA, Schiller NB. Prognostic value of left ventricular end-systolic volume index as a predictor of heart failure hospitalization in stable coronary artery disease: data from the Heart and Soul Study. J Am Soc Echocardiogr 2009 Feb;22(2):190-197 [FREE Full text] [doi: 10.1016/j.echo.2008.11.005] [Medline: 19084372]
8. Vasan RS, Larson MG, Benjamin EJ, Evans JC, Levy D. Left ventricular dilatation and the risk of congestive heart failure in people without myocardial infarction. N Engl J Med 1997 May 08;336(19):1350-1355. [doi: 10.1056/NEJM199705083361903] [Medline: 9134875]
9. Oh JK, Velazquez EJ, Menicanti L, Pohost GM, Bonow RO, Lin G, STICH Investigators. Influence of baseline left ventricular function on the clinical outcome of surgical ventricular reconstruction in patients with ischaemic cardiomyopathy. Eur Heart J 2013 Jan;34(1):39-47 [FREE Full text] [doi: 10.1093/eurheartj/ehs021] [Medline: 22584648]
10. Zhu P, Dai Y, Qiu J, Xu H, Liu J, Zhao Q. Prognostic implications of left ventricular geometry in coronary artery bypass grafting patients. Quant Imaging Med Surg 2020 Dec;10(12):2274-2284 [FREE Full text] [doi: 10.21037/qims-19-926] [Medline: 33269226]
11. Yan P, Zhang K, Cao J, Dong R. Left Ventricular Structure is Associated with Postoperative Death After Coronary Artery Bypass Grafting in Patients with Heart Failure with Reduced Ejection Fraction. Int J Gen Med 2022;15:53-62 [FREE Full text] [doi: 10.2147/IJGM.S341145] [Medline: 35018113]
12. Mosteller RD. Simplified calculation of body-surface area. N Engl J Med 1987 Oct 22;317(17):1098. [doi: 10.1056/NEJM198710223171717] [Medline: 3657876]
13. Hume R, Weyers E. Relationship between total body water and surface area in normal and obese subjects. J Clin Pathol 1971 Apr;24(3):234-238 [FREE Full text] [doi: 10.1136/jcp.24.3.234] [Medline: 5573437]
14. Heerspink HJL, Stefánsson BV, Correa-Rotter R, Chertow GM, Greene T, Hou F, DAPA-CKD Trial CommitteesInvestigators. Dapagliflozin in Patients with Chronic Kidney Disease. N Engl J Med 2020 Oct 08;383(15):1436-1446 [FREE Full text] [doi: 10.1056/NEJMoa2024816] [Medline: 32970396]
15. Verbraecken J, Van de Heyning P, De Backer W, Van Gaal L. Body surface area in normal-weight, overweight, and obese adults. A comparison study. Metabolism 2006 Apr;55(4):515-524. [doi: 10.1016/j.metabol.2005.11.004] [Medline: 16546483]

16. Lang RM, Badano LP, Mor-Avi V, Afilalo J, Armstrong A, Ernande L, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. J Am Soc Echocardiogr 2015 Jan;28(1):1-39.e14 [FREE Full text] [doi: 10.1016/j.echo.2014.10.003] [Medline: 25559473]
17. Zafrir B, Salman N, Crespo-Leiro MG, Anker SD, Coats AJ, Ferrari R, Heart Failure Long-Term Registry Investigators. Body surface area as a prognostic marker in chronic heart failure patients: results from the Heart Failure Registry of the Heart Failure Association of the European Society of Cardiology. Eur J Heart Fail 2016 Jul;18(7):859-868 [FREE Full text] [doi: 10.1002/ejhf.551] [Medline: 27198159]
18. Rao C, Zhang H, Gao H, Zhao Y, Yuan X, Hua K, Chinese Cardiac Surgery Registry Collaborative Group. The Chinese Cardiac Surgery Registry: Design and Data Audit. Ann Thorac Surg 2016 Apr;101(4):1514-1520. [doi: 10.1016/j.athoracsur.2015.09.038] [Medline: 26652141]
19. Zhang Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. Ann Transl Med 2016 Jan;4(2):30 [FREE Full text] [doi: 10.3978/j.issn.2305-5839.2015.12.63] [Medline: 26889483]
20. Ibrahim JG, Chu H, Chen M. Missing data in clinical studies: issues and methods. J Clin Oncol 2012 Sep 10;30(26):3297-3303 [FREE Full text] [doi: 10.1200/JCO.2011.38.7589] [Medline: 22649133]
21. Schafer JL, Graham JW. Missing data: our view of the state of the art. Psychol Methods 2002 Jun;7(2):147-177. [Medline: 12090408]
22. Credit Risk Scorecard. The Comprehensive R Archive Network. URL: https://cran.r-project.org/web/packages/scorecard/index.html [accessed 2023-03-01]
23. Yamaguchi A, Ino T, Adachi H, Murata S, Kamio H, Okada M, et al. Left ventricular volume predicts postoperative course in patients with ischemic cardiomyopathy. Ann Thorac Surg 1998 Feb;65(2):434-438. [doi: 10.1016/s0003-4975(97)01155-7] [Medline: 9485241]
24. Velazquez EJ, Lee KL, Deja MA, Jain A, Sopko G, Marchenko A, STICH Investigators. Coronary-artery bypass surgery in patients with left ventricular dysfunction. N Engl J Med 2011 Apr 28;364(17):1607-1616 [FREE Full text] [doi: 10.1056/NEJMoa1100356] [Medline: 21463150]
25. Park S, Ahn J, Kim TO, Park H, Kang D, Lee PH, IRIS-MAIN Registry Investigators. Revascularization in Patients With Left Main Coronary Artery Disease and Left Ventricular Dysfunction. J Am Coll Cardiol 2020 Sep 22;76(12):1395-1406 [FREE Full text] [doi: 10.1016/j.jacc.2020.07.047] [Medline: 32943156]
26. Potter E, Marwick TH. Assessment of Left Ventricular Function by Echocardiography: The Case for Routinely Adding Global Longitudinal Strain to Ejection Fraction. JACC Cardiovasc Imaging 2018 Feb;11(2 Pt 1):260-274. [doi: 10.1016/j.jcmg.2017.11.017] [Medline: 29413646]
27. Velazquez EJ, Bonow RO. Revascularization in severe left ventricular dysfunction. J Am Coll Cardiol 2015 Feb 17;65(6):615-624 [FREE Full text] [doi: 10.1016/j.jacc.2014.10.070] [Medline: 25677320]
28. Gaasch WH, Zile MR. Left ventricular diastolic dysfunction and diastolic heart failure. Annu Rev Med 2004;55:373-394. [doi: 10.1146/annurev.med.55.091902.104417] [Medline: 14746527]
29. White HD, Norris RM, Brown MA, Brandt PW, Whitlock RM, Wild CJ. Left ventricular end-systolic volume as the major determinant of survival after recovery from myocardial infarction. Circulation 1987 Jul;76(1):44-51. [doi: 10.1161/01.cir.76.1.44] [Medline: 3594774]
30. Silbiger JJ. Pathophysiology and Echocardiographic Diagnosis of Left Ventricular Diastolic Dysfunction. J Am Soc Echocardiogr 2019 Feb;32(2):216-232.e2. [doi: 10.1016/j.echo.2018.11.011] [Medline: 30717860]
31. Sandhu AT, Parizo J, Moradi-Ragheb N, Heidenreich PA. Association Between Offering Limited Left Ventricular Ejection Fraction Echocardiograms and Overall Use of Echocardiography. JAMA Intern Med 2018 Sep 01;178(9):1270-1272 [FREE Full text] [doi: 10.1001/jamainternmed.2018.3317] [Medline: 30039163]
32. Vitale G, Galderisi M, Pivonello R, Spinelli L, Ciccarelli A, de Divitiis O, et al. Prevalence and determinants of left ventricular hypertrophy in acromegaly: impact of different methods of indexing left ventricular mass. Clin Endocrinol (Oxf) 2004 Mar;60(3):343-349. [doi: 10.1111/j.1365-2265.2004.01985.x] [Medline: 15009000]
33. Nafiu OO, Owusu-Bediako K, Chiravuri SD. Effect of Body Mass Index Category on Body Surface Area Calculation in Children Undergoing Cardiac Procedures. Anesth Analg 2020 Feb;130(2):452-461. [doi: 10.1213/ANE.0000000000004016] [Medline: 30676349]
34. Bellenger NG, Burgess MI, Ray SG, Lahiri A, Coats AJ, Cleland JG, et al. Comparison of left ventricular ejection fraction and volumes in heart failure by echocardiography, radionuclide ventriculography and cardiovascular magnetic resonance; are they interchangeable? Eur Heart J 2000 Aug;21(16):1387-1396. [doi: 10.1053/euhj.2000.2011] [Medline: 10952828]
35. Lang RM, Bierig M, Devereux RB, Flachskampf FA, Foster E, Pellikka PA, Chamber Quantification Writing Group, American Society of Echocardiography's Guidelines and Standards Committee, European Association of Echocardiography. Recommendations for chamber quantification: a report from the American Society of Echocardiography's Guidelines and Standards Committee and the Chamber Quantification Writing Group, developed in conjunction with the European Association of Echocardiography, a branch of the European Society of Cardiology. J Am Soc Echocardiogr 2005 Dec;18(12):1440-1463. [doi: 10.1016/j.echo.2005.10.005] [Medline: 16376782]

36. Blasberg JD, Schwartz GS, Balaram SK. The role of gender in coronary surgery. Eur J Cardiothorac Surg 2011 Sep;40(3):715-721. [doi: 10.1016/j.ejcts.2011.01.003] [Medline: 21349733]
37. Nicolini F, Vezzani A, Fortuna D, Contini GA, Pacini D, Gabbieri D, RERIC (Registro dell'Emilia Romagna degli Interventi Cardiochirurgici) Investigators. Gender differences in outcomes following isolated coronary artery bypass grafting: long-term results. J Cardiothorac Surg 2016 Sep 30;11(1):144 [FREE Full text] [doi: 10.1186/s13019-016-0538-4] [Medline: 27716382]
38. Vaccarino V, Abramson JL, Veledar E, Weintraub WS. Sex differences in hospital mortality after coronary artery bypass surgery: evidence for a higher mortality in younger women. Circulation 2002 Mar 12;105(10):1176-1181. [doi: 10.1161/hc1002.105133] [Medline: 11889010]
39. Haider A, Bengs S, Luu J, Osto E, Siller-Matula JM, Muka T, et al. Sex and gender in cardiovascular medicine: presentation and outcomes of acute coronary syndrome. Eur Heart J 2020 Apr 01;41(13):1328-1336 [FREE Full text] [doi: 10.1093/eurheartj/ehz898] [Medline: 31876924]
40. Dixon LK, Di Tommaso E, Dimagli A, Sinha S, Sandhu M, Benedetto U, et al. Impact of sex on outcomes after cardiac surgery: A systematic review and meta-analysis. Int J Cardiol 2021 Nov 15;343:27-34. [doi: 10.1016/j.ijcard.2021.09.011] [Medline: 34520795]

Abbreviations

AUC: area under the curve
bLVEDD: body surface area weighted left ventricular end-diastolic diameter
BSA: body surface area
CABG: coronary artery bypass grafting
EF: ejection fraction
eGFR: estimated glomerular filtration rate
LVEDD: left ventricular end-diastolic diameter
MI: myocardial infarction
OR: odds ratio

Edited by T Leung, G Eysenbach; submitted 20.01.23; this is a non–peer-reviewed article;accepted 21.02.23; published 23.03.23.

Please cite as:
Zhu Z, Li Y, Zhang F, Steiger S, Guo C, Liu N, Lu J, Fan G, Wu W, Wu M, Wang H, Xu D, Chen Y, Zhu J, Meng X, Hou X, Anders HJ, Ye J, Zheng Z, Li C, Zhang H
Prediction of Male Coronary Artery Bypass Grafting Outcomes Using Body Surface Area Weighted Left Ventricular End-diastolic Diameter: Multicenter Retrospective Cohort Study
Interact J Med Res 2023;12:e45898
URL: https://www.i-jmr.org/2023/1/e45898
doi:10.2196/45898
PMID:36951893

Review

Accelerometer-Measured Inpatient Physical Activity and Associated Outcomes After Major Abdominal Surgery: Systematic Review

Mikita Fuchita[1], MD; Kyle J Ridgeway[2,3], PT, DPT; Clinton Kimzey[4], BA; Edward L Melanson[5,6,7], PhD; Ana Fernandez-Bustamante[1], MD, PhD

[1]Department of Anesthesiology, University of Colorado Anschutz Medical Campus, Aurora, CO, United States
[2]Department of Inpatient Rehabilitation Therapy, University of Colorado Hospital, University of Colorado Health, Aurora, CO, United States
[3]Department of Physical Medicine and Rehabilitation, University of Colorado Anschutz Medical Campus, Aurora, CO, United States
[4]University of Colorado School of Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, United States
[5]Department of Medicine, Division of Endocrinology and Metabolism, University of Colorado Anschutz Medical Campus, Aurora, CO, United States
[6]Department of Medicine, Division of Geriatric Medicine, University of Colorado Anschutz Medical Campus, Aurora, CO, United States
[7]Eastern Colorado Veterans Affairs Geriatric Research, Education and Clinical Center, US Department of Veterans Affairs, Aurora, CO, United States

Corresponding Author:
Mikita Fuchita, MD
Department of Anesthesiology
University of Colorado Anschutz Medical Campus
12401 East 17th Avenue, 7th Floor
Aurora, CO, 80045
United States
Phone: 1 848 848 3691
Email: mikita.fuchita@cuanschutz.edu

Abstract

Background: It remains unclear how inpatient physical activity after major abdominal surgery affects outcomes. Accelerometer research may provide further evidence for postoperative mobilization.

Objective: We aimed to summarize the current literature evaluating the impact of accelerometer-measured postoperative physical activity on outcomes after major abdominal surgery.

Methods: We searched PubMed and Google Scholar in October 2021 to conduct a systematic review. Studies were included if they used accelerometers to measure inpatient physical behaviors immediately after major abdominal surgery, defined as any nonobstetric procedures performed under general anesthesia requiring hospital admission. Studies were eligible only if they evaluated the effects of physical activity on postoperative outcomes such as postoperative complications, return of gastrointestinal function, hospital length of stay, discharge destination, and readmissions. We excluded studies involving participants aged <18 years. Risk of bias was assessed using the risk-of-bias assessment tool for nonrandomized studies (RoBANS) for observational studies and the revised Cochrane risk-of-bias tool for randomized trials (RoB 2) for randomized controlled trials (RCTs). Findings were summarized by qualitative synthesis.

Results: We identified 15 studies. Risk of bias was high in 14 (93%) of the 15 studies. Most of the studies (11/15, 73%) had sample sizes of <100. Of the 15 studies, 13 (87%) included the general surgery population, 1 (7%) was a study of patients who had undergone gynecologic surgery, and 1 (7%) included a mixed (abdominal, thoracic, gynecologic, and orthopedic) surgical population. Of the 15 studies, 12 (80%) used consumer-grade accelerometers to measure physical behaviors. Step count was the most commonly reported physical activity outcome (12/15, 80%). In the observational studies (9/15, 60%), increased physical activity during the immediate postoperative period was associated with earlier return of gastrointestinal function, fewer surgical and pulmonary complications, shorter hospital length of stay, and fewer readmissions. In the RCTs (6/15, 40%), only 1 (17%) of the 6 studies demonstrated improved outcomes (shorter time to flatus and hospital length of stay) when a mobility-enhancing intervention was compared with usual care. Notably, mobility-enhancing interventions used in 4 (67%) of the 6 RCTs did not result in increased postoperative physical activity.

Conclusions: Although observational studies show strong associations between postoperative physical activity and outcomes after major abdominal surgery, RCTs have not proved the benefit of mobility-enhancing interventions compared with usual care.

The overall risk of bias was high, and we could not synthesize specific recommendations for postoperative mobilization. Future research would benefit from improving study design, increasing methodologic rigor, and measuring physical behaviors beyond step counts to understand the impact of postoperative mobilization on outcomes after major abdominal surgery.

(Interact J Med Res 2023;12:e46629) doi:10.2196/46629

KEYWORDS

abdominal surgery; accelerometry; early mobilization; physical activity; postoperative care; wearable

Introduction

Background

Clinicians used to prescribe strict bed rest for 2 to 3 weeks after abdominal surgery until pioneers such as Leithauser started challenging this dogma in the 1940s [1]. Through a series of case studies, these pioneers reported that immobility caused harm and that early mobilization was safe and feasible [2-4]. In 2005, the Enhanced Recovery After Surgery (ERAS) Society published its first perioperative guidelines for patients undergoing colorectal surgery [5] and promoted the uptake of early mobilization efforts by clinicians. The guidelines recommended that patients spend 2 hours out of bed on the day of surgery and 6 hours per day out of bed until discharge [5]. Today, ERAS guidelines have expanded to >20 adult specialties, all of them describing early mobilization as a vital component of postoperative care [6].

Despite the widespread acceptance of the ERAS guidelines, the recommendations on postoperative mobilization are built on expert consensus with little to no data supporting the specific mobility goals [7,8]. Early mobilization remains poorly defined in the literature [7-9], and protocols vary substantially between institutions and studies [8,9]. Hence, optimal methods to achieve early mobilization and the impact of specific physical activity components (such as timing, type, duration, frequency, and intensity) [10,11] on postoperative outcomes are still unknown [7-9].

Objectives

Accelerometers have gained popularity as consumer- and research-grade activity-tracking devices [12]. Their ability to quantitatively measure and summarize physical behaviors has attracted many researchers, and, as a result, the number of publications using accelerometers has grown exponentially in recent years [13]. In this systematic review, we aimed to summarize the current literature on accelerometer-measured postoperative physical activity in the acute inpatient setting and its impact on clinical outcomes after major abdominal surgery.

Methods

Overview

We first searched PROSPERO [14] to verify the absence of existing or ongoing research on this topic. We then outlined a written protocol (not registered) according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [15] before conducting the literature search. We adhered to our protocol and the PRISMA statement throughout the review process (Multimedia Appendix 1).

Search Strategy

We searched PubMed and Google Scholar using comprehensive search strategies developed with assistance from an institutional librarian. The search strategy included the Medical Subject Headings (MeSH) terms “postoperative period,” “postoperative care,” “accelerometry,” “wearable electronic devices,” “fitness trackers,” and their related terms (Multimedia Appendix 2). The database included all publications up to October 14, 2021. For Google Scholar, we screened the first 100 articles as described previously [16]. Reference lists of related studies were also used to identify relevant articles.

Inclusion and Exclusion Criteria

We included studies that used accelerometers to measure physical activity during the hospital stay immediately after major abdominal surgery. We defined major abdominal surgery as any nonobstetric procedures performed under general anesthesia requiring hospital admission. Both open and laparoscopic surgeries were eligible. Studies were eligible only if they evaluated the effects of physical activity on postoperative outcomes. Outcomes of interest included postoperative complications, return of gastrointestinal function, hospital length of stay, discharge destination, and readmissions.

We excluded studies if they (1) involved participants aged <18 years; (2) did not include physical activity measurements during the acute inpatient period; (3) only reported descriptive analysis of physical activity measures without evaluating their impact on clinical outcomes; or (4) were case reports, study protocols, or conference abstracts. Review articles were used to search for additional articles not captured by the database search.

Study Selection

After the literature search, we used Covidence systematic review software (Veritas Health Innovation Ltd) to facilitate study selection, data extraction, and quality assessment. First, 2 independent reviewers (MF and KJR) screened titles and abstracts of all articles identified from the database searches and the studies identified from the reference lists. Studies were excluded if they were not relevant (eg, nonabdominal surgery and use of accelerometers for purposes other than physical activity measurements). The included studies underwent a full-text review by 2 independent reviewers (MF and KJR) using the aforementioned inclusion and exclusion criteria. Any disagreements between the reviewers for each of these steps were resolved by a third reviewer (AFB).

Data Extraction

Data extraction was performed by a single investigator (MF). The data collection form contained the following variables: study design, type of surgery, number of participants, patient characteristics, descriptions of interventions (if applicable), device name, device setup (eg, sampling rate, filter, and epoch), device wear location, data collection period, reported measures of physical activity (including but not limited to step count, postural transition, activity duration, time-to-mobilization events, and activity trend over time), and clinical outcomes. We extrapolated the duration of device wear in the hospital based on the methods described in each study.

Quality Assessment

We used the risk-of-bias assessment tool for nonrandomized studies (RoBANS) [17] to assess the risk of bias in observational studies and the revised Cochrane risk-of-bias tool for randomized trials (RoB 2) [18] to assess the risk of bias in randomized controlled trials (RCTs). For observational studies, we predefined the following factors as confounding variables based on the literature: (1) preoperative level of physical activity [19], (2) American Society of Anesthesiologists physical status classification [19-21], (3) performed procedure [22], (4) open versus laparoscopic approach [20,23-25], (5) duration of surgery [20,21,23], and (6) postoperative intensive care unit admission [20,23]. Two independent assessors (MF and CK) evaluated each study, and conflicts were resolved through consensus. We used R statistical software (version 4.1.1; R Foundation for Statistical Computing) and the R package *robvis* [26] for data visualization.

Data Synthesis

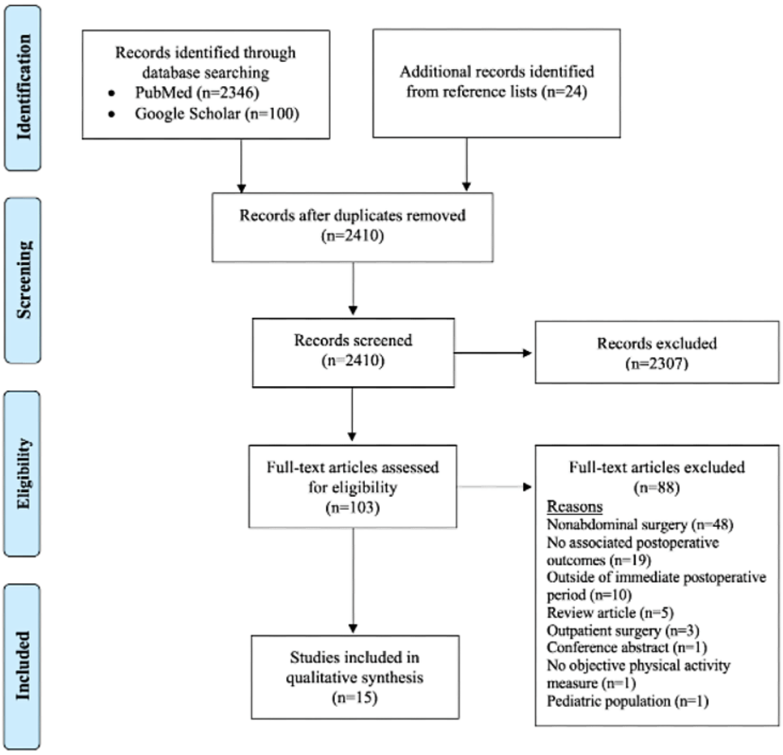
Because of the heterogeneity of study populations, methods, and device models, the data were not amenable for a meta-analysis. We performed a qualitative synthesis by summarizing the findings in three themes: (1) device use, (2) metrics used to describe physical activity, and (3) clinical outcomes analyzed in association with physical activity. Observational studies and RCTs were organized separately, given the differences in study designs. The key findings of individual studies were summarized in tables by tabulating the following variables: type of surgery, patient characteristics, main predictor (observational studies), intervention and control (RCTs), and main findings.

Results

Literature Search

We identified 2470 articles: 2446 (99.03%) through the database searches and 24 (0.97%) from the review of reference lists. After screening the titles and abstracts of all 2470 articles, 103 (4.17%) underwent a full-text review. Of these 103 articles, 15 (14.6%) met our selection criteria [27-41]. The reasons for exclusion are detailed in Figure 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of literature search and study selection.



Characteristics of the Included Studies

All articles were published between 2017 and 2021, except for the study by Browning et al [27], which was published in 2007. Of the 15 articles, 9 (60%) were observational studies [27-35], and 6 (40%) were RCTs [36-41]. The median sample sizes were 54 (IQR 50-94) for observational studies and 98 (IQR 64-107) for RCTs. Of the 15 articles, 13 (87%) studied the general surgery population, 1 (7%) was a study of patients who had undergone gynecologic surgery [40], and 1 (7%) included a mixed (abdominal, thoracic, gynecologic, and orthopedic) surgical population [30]. The results of individual studies are summarized in Tables 1-3.

Table 1. Summary of observational studies: physical activity predictors and associated outcomes.

Studies	Type of surgery (number of patients)	Patient characteristics	Main predictor	Main findings
Browning et al [27], 2007	Upper abdominal surgery (50)	Age: mean 61 (SD 12) years; BMI: mean 27.1 (SD 4.3) kg/m^2; ASA[a] classification I and II: 62% and ASA classification III: 38%; laparoscopic surgery: 0%; and LOS[b]: median 8 (IQR 3-121) days	Duration of uptime (standing or walking) during the first 4 postoperative days	A shorter uptime during the first 4 postoperative days was predictive of longer LOS (R^2=0.50; *P*<.001); patients who developed pulmonary complications spent shorter uptime during the first 4 postoperative days
Low et al [28], 2018	Hyperthermic intraperitoneal chemotherapy with cytoreductive surgery (54)	Age: mean 57 (SD 11) years; BMI: mean 27.3 (SD 5.8) kg/m^2; ASA classification II: 23% and ASA classification III and IV: 77%; laparoscopic surgery: 0%; and LOS: mean 12 (SD 7) days	Postoperative step count	Taking more steps during the inpatient recovery period predicted a lower risk of *30-day readmission* (OR[c] 0.83, 95% CI 0.72-0.96; *P*=.01) for each additional 100 steps taken per day
Jonsson et al [29], 2018	Acute high-risk abdominal surgery (50)	Age: mean 61 (SD 17) years; BMI: mean 25.0 (SD 5.8) kg/m^2; ASA classification I and II: 66% and ASA classification III and IV: 34%; laparoscopic surgery: 16%; and LOS: median 12 (IQR 7-22) days	Independent ambulation within the first operative week	Patients who achieved independent ambulation within the first postoperative week had fewer pulmonary complications than those who did not achieve independent ambulation (14% vs 53%, respectively; *P*=.01) and had a shorter LOS (8 days vs 22 days; *P*=.001)
Daskivich et al [30], 2019	Abdominal, thoracic, gynecologic, and orthopedic surgery (100)	Age: mean 53 (SD 18) years; BMI: mean 31 (SD 12) kg/m^2; abdominal surgery: 79%; and LOS: median 4 (IQR 3-6) days	Step count on postoperative day 1	Higher step count on postoperative day 1 was associated with a lower probability of a prolonged LOS (OR 0.63, 95% CI 0.45-0.84; *P*=.003) for every additional 100 steps taken; no further reduction in LOS was observed beyond 1000 steps
Martin et al [31], 2020	Colorectal surgery (50)	Age: mean 59 (SD 18) years; BMI: mean 25.4 (SD 4.3) kg/m^2; ASA classification I and II: 86% and ASA classification III and IV: 14%; minimally invasive surgery: 88%; and LOS: prolonged; >5 days	Postoperative step count during the first 3 postoperative days	Patients with postoperative complications took fewer steps during the first 3 postoperative days than those without complications (daily average: 1101, SD 2198 vs 1243, SD 1641 steps, respectively; *P*=.02); daily average step count was negatively correlated with LOS (*r*=−0.31; *P*=.03)
Nevo et al [32], 2021	Major abdominal surgery (91)	Age: mean 55 (SD 14) years; BMI: median 25.9 (IQR 20.9-30.9) kg/m^2; ASA classification I and II: 59% and ASA classification III: 41%; laparoscopic surgery: 41%; and LOS: median 6 (IQR 4-8) days	Step count on postoperative day 2	Patients who took >1050 steps on postoperative day 2 had fewer postoperative complications (32% vs 71%; *P*<.05) and had shorter time to flatus (2.4 days vs 3.3 days; *P*<.01), time to bowel movement (1.8 days vs 3.2 days; *P*<.01), and LOS (5.4 days vs 8.8 days; *P*<.01), as well as lower readmission rate (*P*<.05)
Iida et al [33], 2021	Hepatectomy (147)	Age: >60 years; BMI: >20.0 kg/m^2; laparoscopic surgery: 59%; and median LOS: 9, 14, and 12 days for upward slope, bell curve, and flat types, respectively	Physical activity trend patterns	Postoperative complications occurred in 4.5%, 76.9%, and 65.2% for upward slope, bell curve, and flat types, respectively (*P*<.001). Pneumonia was only observed among the flat type
Yi et al [34], 2021	Bowel surgery (37)	Age: mean 39 (SD 14) years; BMI: mean 27.7 (SD 8.3) kg/m^2; ASA classification II: 70% and ASA classification III: 30%; laparoscopic or robotic surgery: 84%; and LOS: 6 days	Postoperative step count	Postoperative step count was not associated with LOS
Kane et al [35], 2021	Colorectal surgery (94)	Age: median 55.5 (IQR 25.5-61.5) years[d] and median 58.0 (IQR 42.0-65.0) years[e]; BMI: median 29.1 (IQR 23.0-36.1) kg/m^{2d} and median 28.5 (IQR 23.5-30.4) kg/m^{2e}; ASA classification II: 53% and ASA classification III: 47%; laparoscopic surgery: 46%; and LOS: median 3 (IQR 2-4) days	Step count on the day of discharge	A higher step count on the day of discharge was associated with a lower 30-day readmission risk; each 10% increase in return to preoperative baseline steps was associated with a 40% decrease in risk of 30-day readmission (OR 0.60, 95% CI 0.39-0.91; *P*=.02)

[a]ASA: American Society of Anesthesiologists.

[b]LOS: length of stay.

[c]OR: odds ratio.

[d]Patients who were readmitted.

[e]Patients who were not readmitted.

Table 2. Summary of randomized controlled trials (patient population).

Studies	Type of surgery	Patient characteristics	
		Control	Intervention
Fiore et al [36], 2017	Colorectal surgery	n=49; age: median 63 (IQR 48-72) years; BMI: median 26.2 (IQR 23.1-30.7) kg/m^2; ASA[a] classification I and II: 84%; laparoscopic surgery: 80%; and LOS[b]: median 3 (IQR 3-4) days	n=50; age: median 65 (IQR 51-71) years; BMI: median 26.6 (24.0-29.2) kg/m^2; ASA classification: I and II: 86%; laparoscopic surgery: 82%; and LOS: median 4 (IQR 2-4) days
Ni et al [37], 2018	Hepatectomy	n=60; age: mean 49 (SD 15) years; and LOS: mean 7.7 (SD 2.1) days	n=59; age: mean 51 (SD 17) years; and LOS: mean 6.6 (SD 2.3) days
Wolk et al [38], 2019	Major visceral surgery	n=54; age: mean 57 years; BMI: mean 26.1 kg/m^2; ASA classification I and II: 65% and ASA classification III: 31%; laparoscopic surgery: 50%; and LOS: 12 days	n=56; age: mean 60 years; BMI: mean 25.7 kg/m^2; ASA classification I and II: 59% and ASA classification III: 39%; laparoscopic surgery: 52%; and LOS: 13 days
Waller et al [39], 2021	Elective colorectal surgery	n=23; age: 54 (SD 18) years; laparoscopic surgery: 52%	n=20; age: 54 (SD 13) years; laparoscopic surgery: 60%
No et al [40], 2021	Gynecologic midline laparotomy	n=28; age: 55 (SD 12) years; ASA classification I and II: 93% and ASA classification III: 7%; and LOS: median 6 (range 4-26) days	n=35; age: 53 (SD 10) years; ASA classification I and II: 100%; and LOS: median 7 (range 4-58) days
Steffens et al [41], 2021	Liver, gastric, and pancreatic cancer	n=49; age: median 64 (IQR 53-71) years; BMI: median 26.2 (IQR 21.9-29.2) kg/m^2; laparoscopic surgery: 27%; and LOS: median 9 (IQR 6-15) days	n=47; age: median 65 (IQR 54-73) years; BMI: median 25.0 (22.5-29.7) kg/m^2; laparoscopic surgery: 21%; and LOS: median 11 (IQR 7-17) days

[a]ASA: American Society of Anesthesiologists.

[b]LOS: length of stay.

Table 3. Summary of randomized controlled trials (interventions and main findings).

Studies	Control	Intervention	Main findings
Fiore et al [36], 2017	Usual care: preoperative instruction on early mobility, facilitation of postoperative mobilization by a nurse or nursing assistant, physiotherapy referral as needed, transfer to chair for 2 hours (day of surgery), and out of bed for at least 6 hours (from POD[a] 1 until discharge)	Facilitated mobilization: in addition to usual care, physiotherapy education and assistance on mobilization (once on the day of surgery and 3 times daily from POD 1 until discharge), as well as a minimal 200-m walking target per session	More patients in the intervention arm got out of bed on the day of surgery and spent at least 6 hours out of bed on POD 1 and POD 2; step counts were at least 2-fold greater in the intervention arm on POD 1 and POD 2; primary and secondary outcomes were similar between the control and intervention groups (primary outcome: proportion of patients who returned to baseline 6MWT[b] on postoperative week 4, 54% vs 51%; *P*=.58; secondary outcomes: median time to recovery of gastrointestinal function, 52.9 hours vs 46.6 hours; *P*=.60; median time to readiness to discharge, 3 days vs 3 days; *P*=.45; and complication rate, 48% vs 43%)
Ni et al [37], 2018	Usual care: bed activities (POD 1 and POD 2), bedside standing (POD 3), and ambulation (POD 4 and POD 5)	Early ambulation: preoperative education on early mobility, passive range of motion activities (the day of surgery), sit at the edge of the bed (POD 1), ambulate 2 to 3 times (POD 2), and ambulate more than 5 times (POD 3)	Patients in the intervention group took more steps from POD 2 to POD 5 and had shorter time to flatus (2.3 days vs 3.1 days; *P*=.04) and shorter LOS[c] (6.6 days vs 7.7 days; *P*=.01); rates of postoperative complications were similar between the groups
Wolk et al [38], 2019	Usual care: ERAS[d] protocol and blinded activity tracker wristband with no feedback or step count targets	Daily activity feedback: ERAS protocol and unblinded activity tracker wristband with daily activity feedback as well as predefined daily step count targets that progressed during the first 5 PODs	The intervention resulted in higher step counts during the first 5 PODs among patients who underwent laparoscopic surgery; postoperative complication rates and LOS were similar between the control and intervention groups for both patients who underwent open surgery and those who underwent laparoscopic surgery
Waller et al [39], 2021	Usual care: ERAS protocol, activity tracker without alarms, and daily ambulation target of 450 steps	Activity tracker with alarms: ERAS protocol, activity tracker with 5 daily alarms, and daily ambulation target of 600 steps	The intervention had no effect on postoperative step counts; postoperative outcomes (duration of ileus, incidences of pulmonary complications, venous thromboembolism, and LOS) were similar between the control and intervention groups
No et al [40], 2021	Usual care: encouragement for postoperative ambulation and blinded activity tracker with no feedback or step count targets	Activity tracker with feedback: in addition to usual care, patient self-monitoring of step counts and encouragement to meet daily targets until POD 5	Patients in the intervention arm took more steps on POD 4 and POD 5 and had higher percentage of recovery in steps from preoperative baseline than those in the control arm; however, the differences were not significant after adjusting for ASA[e] classification (*P*=.90); outcomes were similar between the 2 groups (first flatus: 4 days vs 3 days; first diet: 3 days vs 3 days; ileus: 7% vs 9%; venous thromboembolism: none vs none; and LOS: 7 days vs 6 days)
Steffens et al [41], 2021	Usual care: preoperative counseling, daily physiotherapy sessions, and no activity tracker	Individualized target steps: in addition to usual care, individualized, progressive mobilization protocol with a study physiotherapist until hospital discharge	The intervention did not result in increased activity levels based on patient self-report (no objective activity data were collected from the control group); the patients in the intervention arm were more fatigued upon discharge; outcomes were similar between the intervention and control arms (postoperative complications: 45% vs 29%; *P*=.14; LOS: 11 days vs 9 days; *P*=.15; discharge home: 94% vs 92%; *P*=.99; and 30-day readmission: 17% vs 12%; *P*=.57)

[a]POD: postoperative day.

[b]6MWT: 6-minute walk test.

[c]LOS: length of stay.

[d]ERAS: Enhanced Recovery After Surgery.

[e]ASA: American Society of Anesthesiologists.

Quality Assessment

Figure 2 summarizes the risk-of-bias assessment for the observational studies (n=9), all of which used an accelerometer to collect physical activity data and were at low risk of bias for domain 3 (measurement of exposure). By contrast, none adequately controlled for confounders and were determined to be at high risk of bias for domain 2 (confounding variables). Of the 9 studies, only 3 (33%) clearly stated their study objectives [28,30,35]; the remaining 6 (67%) were exploratory, putting them at high risk of bias for domain 6 (selective outcome reporting) [27,29,31-34].

Figure 3 shows the summary of the risk-of-bias assessment for the RCTs (n=6), all of which had at least some methodological concerns and were determined to have an overall high risk of bias, except for the study by Fiore et al [36], which was the most rigorous among these studies.

Figure 2. Quality assessment of observational studies using the risk-of-bias assessment tool for nonrandomized studies (RoBANS).

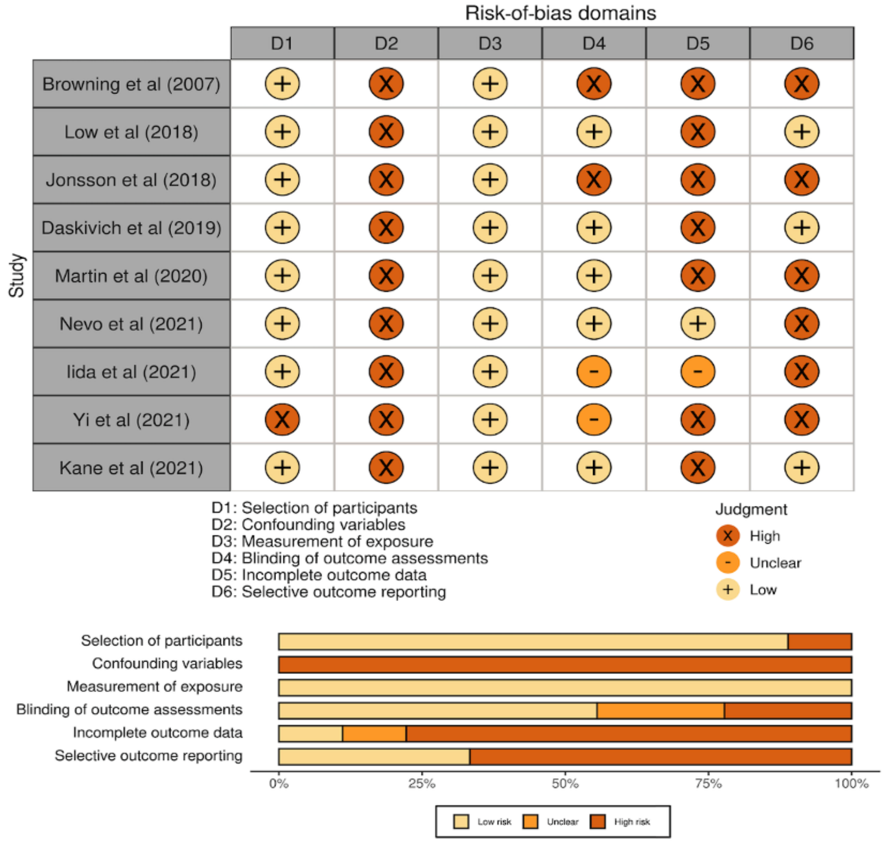
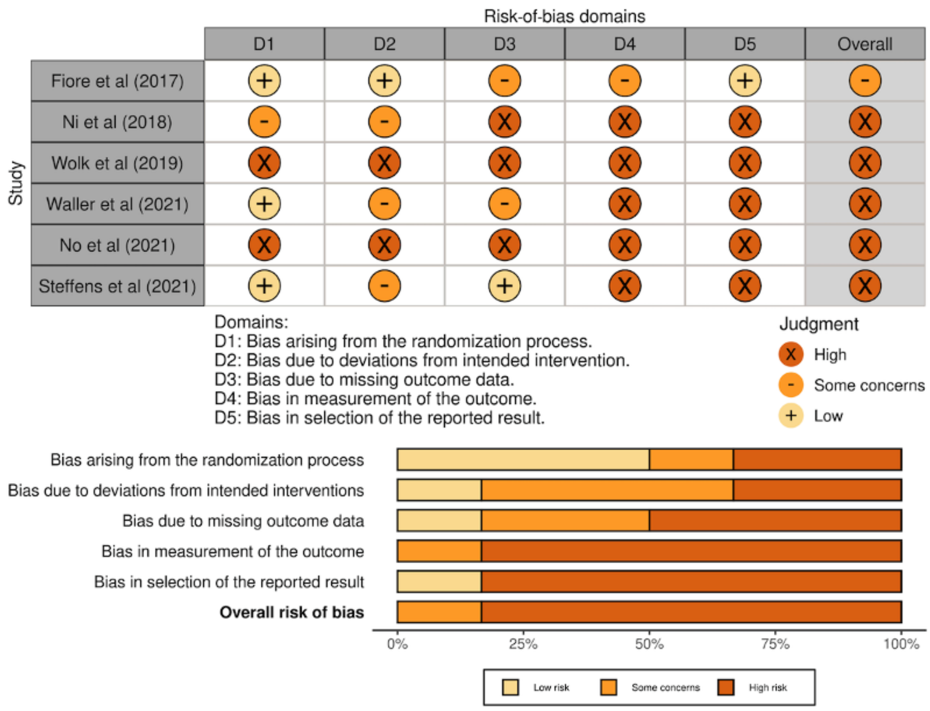


Figure 3. Quality assessment of randomized controlled trials using the revised Cochrane risk-of-bias tool for randomized trials (RoB 2).



Choice of Accelerometers

The device choice, use, and reported outcomes are summarized in Multimedia Appendix 3. Of the 15 studies, 3 (20%) used research-grade accelerometers [27,29,36], whereas the remaining 12 (80%) used consumer-grade devices [28,30-35,37-41]. Among the commercially available devices, the Fitbit series (Google LLC) was used most frequently (7/12, 58%) [28,30,34,35,37,39,41]. Of the 3 studies using research-grade accelerometers, only 1 (33%) study, which used the ActiGraph GT3X25 (ActiGraph LLC), described the accelerometry setup (such as sampling rate, filter, epoch, and analysis algorithm or software) [36].

Device Wear Period

All studies used accelerometers during the acute inpatient period (per our inclusion criteria). Of the 15 studies, 12 (80%) applied the device within a day after surgery (n=7, 58%, studies on the day of surgery [30-32,35,36,38,39] and n=5, 42%, studies on postoperative day 1 or within 24 hours after surgery [27,29,33,37,40]). Of the 15 studies, 2 (13%) started the device wear after patients were transferred to the floor from the intensive care unit, which occurred on postoperative day 2 or 3 on average [28,41], and 1 (7%) did not report the timing of initiation [34]. The mean in-hospital wear duration was 5.2 (SD 2.0; range 3-10) days. Most of the studies (11/15, 73%) described continuous 24-hour device wear with or without brief interruptions for battery charging or patient showering [27-30,32-34,36,38,39,41], whereas others implied continuous measurement but did not explicitly describe it [31,35,37,40]. Of the 15 studies, 3 (20%) obtained preoperative baseline data, ranging from 2 to 30 days before surgery [31,35,40], and 1 (7%) followed patients after discharge until postoperative day 30 [35].

Metrics Used to Describe Physical Activity

Step Count

Step count was the most commonly reported physical activity outcome, used in 12 (80%) of the 15 studies [28,30-32,34-41]. Of these 12 studies, 9 (75%) reported step count for each postoperative day [30,32,34-40]. Of these 9 studies, 3 (33%) averaged daily step counts over the entire postoperative study period, which ranged from 3 to 10 days [28,31,38], and 1 (33%) of these 3 studies also reported cumulative steps over the first 5 postoperative days [38].

Of the 12 studies that reported step count as the physical activity outcome, 3 (25%) measured preoperative step counts and also examined postoperative changes from baseline [31,35,40]. In the study by Martin et al [31], the average daily step counts decreased from 6444 (SD 4095) steps 5 days immediately before surgery to 1191 (SD 1864) steps 3 days after surgery (82% reduction). No et al [40] used preoperative step counts to calculate the percentage of recovery from baseline on postoperative days 4 and 5 to determine the efficacy of their activity-promoting intervention. Kane et al [35] calculated the percentage of return to preoperative baseline upon discharge and used it to predict 30-day readmission.

Activity Duration

Of the 15 studies, 2 (13%) observed that patients spent little time standing or walking (up to 0.6 hours per day) during the first week after surgery [27,29]. Fiore et al [36] measured the time spent out of bed (whether sitting or standing), which ranged from 0.6 to 0.9 hours on the day of surgery and from 6.7 to 10.3 hours between postoperative days 1 and 3. Wolk et al [38] reported an increase from 290 to 482 minutes of postoperative *activity time* per day, but *activity time* was not clearly defined.

Activity Trend

Many of the studies (11/15, 73%) presented daily activity trends using various outcome measures, including step counts [30-32,35-39], activity duration and sit-to-stand transitions [27,29], and energy expenditure [33]. These studies demonstrated that physical activity gradually increased after hitting the nadir immediately after surgery [27,29,30,32,33,35-39]. The studies by Daskivich et al [30] and Nevo et al [32] also showed that the recovery speed (measured in daily step counts) was different depending on the procedure type. Patients undergoing laparoscopic surgery had a more steady and faster recovery than those undergoing open surgery [33]. Nevo et al [32] and Iida et al [33] further analyzed how different recovery trajectories could inform the risks of developing postoperative complications (described in more detail in the *Surgical Complications* subsection).

Other Activity Metrics

Studies that used thigh-worn accelerometers reported daily numbers of sit-to-stand transitions [27,29]. Of the 15 studies, 2 (13%) tracked time-to-mobilization milestones such as sitting out of bed, standing, and walking [27,36]. Fiore et al [36] replicated the mobilization goals described in the 2005 ERAS guidelines (described in the *Introduction* section) [5].

Clinical Outcomes Analyzed in Association With Physical Activity

Hospital Length of Stay

Of the 6 observational studies, 5 (83%) analyzed the impact of physical activity on hospital length of stay and found that being more active during the immediate postoperative period was associated with a shorter length of stay [27,29-32]; for example, in 1 (20%) of these 5 studies, every additional 100 steps up to 1000 steps on postoperative day 1 was associated with a shorter length of stay (odds ratio 0.63, 95% CI 0.45-0.84; P=.003) [30]. In another study, patients who achieved >1050 steps on postoperative day 2 had a shorter stay in the hospital than those who did not achieve that milestone (5.4 days vs 8.8 days, respectively; P<.01) [32].

By contrast, only 1 (17%) of the 6 RCTs demonstrated a reduction in hospital length of stay from interventions to enhance postoperative mobilization [37]. However, in 4 (67%) of the 6 RCTs, physical activity performances were similar between the control and the intervention arms [38-41]. In the study by Fiore et al [36], patients in the facilitated mobilization group spent more time out of bed and took twice as many steps on postoperative days 1 and 2 but had hospital lengths of stay similar to those of the patients in the control group.

Surgical Complications

Of the 9 observational studies, 3 (33%) tracked surgical complications and identified postoperative activity as a predictor of surgical complications, as defined by the Clavien-Dindo classification of surgical complications [31-33]. In these studies, higher step counts during the first 3 days [31,32] and a steady recovery trajectory during the first postoperative 7 days (*upward slope type*) [33] were predictive of fewer surgical complications. Conversely, Nevo et al [32] found that an acute drop in daily step count (a drop of >50% from the previous day and <500 steps) was strongly associated with severe complications (Clavien-Dindo grade ≥III; odds ratio 7.87, 95% CI 1.63-27.9; P=.007). Iida et al [33] described this pattern as a *bell curve* *type* and also noted a high complication rate among patients in this category (76.9%). In the same study, approximately one-third of the patients showed minimal progression in activity levels during the first 7 days after surgery (*flat type*), and 65.2% of them experienced complications [33].

Regarding the RCTs, mobility-enhancing interventions did not reduce surgical complications in any of the studies that evaluated these outcomes (4/6, 67%) [36-38,41], including the studies (2/4, 50%) that successfully increased physical activity in the intervention group compared with the control group [36,37].

Readmission

Of the 9 observational studies, 3 (33%) used inpatient activity to predict hospital readmission and found that higher step counts across the inpatient period (postoperative day 2 [32], inpatient average [28], and on the day of discharge [35]) were predictive of a lower likelihood of readmission after hospital discharge.

By contrast, of the 6 RCTS, 1 (17%) looked at 30-day readmission and did not find any difference between the intervention group and the control group (17% vs 12%, respectively; *P*=.57) [41].

Return of Gastrointestinal Function

Of the 9 observational studies, 1 (11%) found that patients who achieved a step count of >1050 on postoperative day 2 had a shorter time to first flatus (2.4 days vs 3.3 days; *P*<.01) and time to first bowel movement (3.2 days vs 4.9 days; *P*<.01) than those with fewer step counts [32].

Of the 6 RCTs, 4 (67%) evaluated the return of gastrointestinal function; only the study by Ni et al [37] demonstrated a shorter time to gastrointestinal recovery from a mobility-enhancing intervention (time to flatus: 2.3 vs 3.1 days; *P*=.04). The remaining RCTs (3/4, 75%), including the well-conducted study by Fiore et al [36], found no such effect from mobility-enhancing interventions [39,40].

Postoperative Pulmonary Complications

Of the 15 studies, 4 (27%; n=2, 50% of observational studies [27,29] and n=2, 50% of RCTs [36,39]) analyzed the effect of physical activity on postoperative pulmonary complications. Each study defined postoperative pulmonary complications differently, and the reported incidence rate ranged from 0% to 34%. In the observational studies, patients who developed pulmonary complications spent shorter times in upright positions (standing or walking) during the first 4 to 7 postoperative days than those who did not develop complications [27,29]. By contrast, the RCTs found no differences in pulmonary complications between the intervention and control groups. The overall incidence rates of pulmonary complications were very low in these RCTs (4% and 0%) [36,39].

Venous Thromboembolism

Of the 6 RCTS, 2 (33%) examined the incidence of venous thromboembolism [36,39]. Three patients developed venous thromboembolic complications in the study by Fiore et al [36], with no significant difference between the intervention and control groups. None of the 43 patients in the study by Waller et al [39] developed venous thromboembolism.

Discussion

Principal Findings

In this systematic review, we found 15 articles that used accelerometers to evaluate the effects of postoperative physical activity on outcomes after major abdominal surgery, with 14 (93%) published within the last 5 years. Although the observational studies (9/15, 60%) consistently showed that increased physical activity during the immediate postoperative period was associated with improved patient outcomes, only 1 (17%) of the 6 RCTs demonstrated that a mobility-enhancing intervention was beneficial compared with usual care. These findings confirm that physical activity is an important predictor of outcomes, but leave important questions unanswered—what is the optimal postoperative mobilization strategy or the dose of mobilization associated with better outcomes? Because of the high risks of bias, we could not synthesize specific mobility recommendations. However, our study illustrates how accelerometers can be a powerful tool for quantifying objective, continuous measures of physical behaviors in the hospital and provides guidance for future research to improve methodological rigors and study design.

We found from this systematic review that physical behaviors follow certain patterns after abdominal surgery. First, surgery causes a steep drop in physical activity from the preoperative baseline [31,35]. This effect is more significant after open abdominal surgery than after laparoscopic surgery [33] and varies by procedure type [30]. Second, the recovery of physical activity is slow, often requiring >1 month to return to baseline [30,33,35,39]. The recovery speed is also different, depending on the procedure [30-32], which is consistent with previous literature [24,25]. In the observational studies (n=9), increased physical activity during the immediate postoperative period was associated with improved clinical outcomes regarding surgical complications, return of gastrointestinal function, postoperative pulmonary complications, hospital length of stay, and hospital readmission [27-35]. These findings suggest that physical behaviors are important predictors of outcomes. In more practical terms, clinicians could use certain physical behaviors to predict or identify patients at risk for adverse outcomes after surgery.

Notably, 4 (67%) of the 6 mobility-enhancing interventions used in the RCTs did not increase postoperative mobilization compared with usual care [38-41]. The mobility-enhancing interventions ranged from step count feedback with encouragement to designated study physiotherapists assisting patients to achieve set mobility milestones. We speculate several reasons why many of these interventions (4/6, 67%) did not enhance mobility performances beyond usual care: (1) the selected interventions were simply ineffective, (2) the selected activity measure (step count was the most commonly used) was not sensitive enough to detect changes in mobility performances, and (3) mobility performances were nonmodifiable. Furthermore, the RCTs (2/6, 33%) that successfully enhanced physical activity showed conflicting effects on clinical outcomes. In the study by Ni et al [37], patients in the intervention arm achieved higher step counts from postoperative days 2 to 5 and had a faster return of gastrointestinal function and shorter hospital length of stay. By contrast, in the study by Fiore et al [36], more patients in the intervention arm were out of bed (sitting or standing) from the day of surgery through postoperative day 2 and took more steps on postoperative days 1 and 2 but had similar outcomes on return of gastrointestinal function and hospital length of stay.

Several factors could explain why postoperative mobilization had little effect on clinical outcomes when studied prospectively in the RCTs. First, the sample sizes of these RCTs were relatively small (median 98, IQR 64-107). Therefore, they could have lacked the statistical power to detect differences in clinical outcomes. Second, postoperative physical activity may be a prognostic indicator of outcomes rather than a modifiable factor. This theory is plausible, given that the factors associated with reduced postoperative mobilization and worse clinical outcomes often overlap, such as preoperative physical activity level [19-21], open versus minimally invasive approach [20,23-25], and duration of surgery [20,21,23]. Third, it is possible that the achieved differences in mobilization dosage (such as timing, type, duration, frequency, and intensity) [10,11] were not significant enough to affect clinical outcomes. Fourth, routine care that involves basic mobility may be sufficient to prevent immobility harm. Fifth and last, the effects of specific physical activity measures on postoperative outcomes remain unknown [8]; for example, it is unclear whether sitting out of bed (static positioning) is as effective as standing and walking (active mobility) in improving clinical outcomes. Thus, the choice of reported mobility metrics could have affected the researchers' ability to detect clinically meaningful differences in activity exposures.

The particularly well-conducted study by Fiore et al [36] is worth special attention. The authors defined physical activity as "out of bed at all on the day of surgery and out of bed for at least 6 hours on postoperative day 1-3," which directly reflects the recommendation described in the original ERAS guidelines [5]. This RCT found no benefit from the authors' facilitated mobilization intervention, including the 6-minute walk test at 4 weeks (primary outcome), time to gastrointestinal recovery, time to readiness to discharge, length of stay, and 30-day complications. The negative result may be partly due to patient selection because 80% (80/99) of the study participants received laparoscopic surgery. Laparoscopic surgeries have been shown to expedite recovery [24,25,38], and treatment effects from mobility-enhancing interventions may be less pronounced in patients undergoing laparoscopic surgeries than in those undergoing open surgeries, especially in environments with optimal usual care. In the case of the study by Fiore et al [36], patients in the usual care group reached activity levels similar to those reached by patients in the intervention arm by postoperative day 3.

We found that most of the studies (12/15, 80%) used consumer-grade accelerometers to characterize physical behaviors. Commercially available devices have appealing features such as patient familiarity, user-friendly interfaces, and fashionable designs, all of which could improve wear compliance. In addition, measures such as step count are intuitive and easy to interpret among many users. However, consumer-grade accelerometers are different from research-grade accelerometers in that they use proprietary algorithms to compute and report physical behavior measures such as step count and energy expenditure. Furthermore, they do not give researchers access to accelerometer settings such as filter, sampling rate, epochs, and software algorithm. As patients who are hospitalized are distinct from the free-living population in that they spend most of their wakeful time sedentary or in bed [42-44], walk significantly slower, and may hold on to an intravenous pole or an assistive device when ambulating [43], researchers may benefit from using research-grade accelerometers because of their flexibility in terms of data collection and analysis [45].

Importantly, most validation studies of accelerometer devices are derived from laboratory and free-living conditions. These studies show that measurements can vary substantially by device manufacturer [46-50], wear location [47,49-52], and data-processing algorithm [46,48]. For step count, the most commonly reported physical activity outcome in our review (12/15, 80%), the discrepancy in measurement can be as much as 120%, depending on the device and wear location [50]. Moreover, the study population and setting can affect device accuracy; for example, older adults, who tend to walk slower than younger adults, walked at a speed of 0.74 meters per second as outpatients but recorded a speed of 0.46 meters per second as inpatients [42]. One study found 20% absolute percentage errors in step counts at a gait speed of 0.42 meters per second and 45% errors at an even slower pace [49]. Many accelerometers available in the market, including research-grade devices, still await validation in acute inpatient settings [42,53]. As is the case with laboratory biomarkers, digital biomarkers derived from biometric monitoring technologies require multistep validation before they can be applied reliably to a specific patient population and clinical setting [54]. It is critical to be mindful of these limitations when interpreting the results or conducting research using accelerometers because the reported outcomes, particularly step count, are not directly comparable [50-52].

Strengths and Limitations

There are 2 major limitations related to the conduct of this systematic review. First, database searches were limited to PubMed and Google Scholar owing to our time constraints; therefore, we could have missed articles available in other databases. To supplement this, we used the reference lists of the included studies and related review articles to identify relevant studies. Second, the heterogeneity of study designs, patient populations, and accelerometer use made it difficult to compare study findings. To minimize the risk of bias resulting from data synthesis, we developed and followed a written protocol using rigorous systematic review processes.

Future Directions

Overall, the quality of available evidence was poor, and we could not synthesize specific recommendations for postoperative mobilization. On the basis of the limitations we identified in the included studies, we recommend that researchers (1) select a patient population that is more likely to benefit from mobility-enhancing interventions (eg, patients undergoing open abdominal surgery rather than laparoscopic surgery and patients with frailty rather than those who are young, healthy, and fit); (2) clearly define and measure timing, type, duration, frequency, and intensity of a mobility-enhancing intervention to delineate the differences in mobility performances achieved by patients in different treatment groups; (3) measure all relevant data (such as patient, surgical, and postoperative factors) to control for confounders adequately; and (4) measure physical behaviors beyond step counts (such as static positioning and in-bed activities) because patients are highly sedentary after surgery [55,56], and step counts only capture snapshots of patients' mobility status.

Conclusions

In conclusion, although observational studies showed strong associations between postoperative physical activity and outcomes after major abdominal surgery, RCTs have not proven the benefit of mobility-enhancing interventions compared with usual care. To understand the optimal postoperative mobilization strategy or the impact of individual physical activity components such as timing, type, duration, frequency, and intensity, future accelerometer research would benefit from improved study designs, increased methodologic rigor, and more consistent reporting of accelerometer methods [57].

Acknowledgments

The authors thank Benjamin D Harnke, MLIS, education and research informationist at the Strauss Health Sciences Library of the University of Colorado Anschutz Medical Campus (Aurora, Colorado, United States), for helping to develop comprehensive search strategies for the database searches. AF-B received funding from the National Heart, Lung, and Blood Institute (UH3-HL140177). The funding agency had no role in the conduct of this research. MF received salary support from the Department of Anesthesiology at the University of Colorado Anschutz Medical Campus to conduct this project as a research fellow.

Data Availability

The data sets generated or analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.

[DOCX File , 30 KB - ijmr_v12i1e46629_app1.docx]

Multimedia Appendix 2

PubMed and Google Scholar search strategies.

[DOCX File , 16 KB - ijmr_v12i1e46629_app2.docx]

Multimedia Appendix 3

Device choice, use, and reported outcomes.

[DOCX File , 232 KB - ijmr_v12i1e46629_app3.docx]

References

1. Brieger GH. Early ambulation. A study in the history of surgery. Ann Surg 1983 Apr;197(4):443-449. [doi: 10.1097/00000658-198304000-00012] [Medline: 6338846]
2. Leithauser DJ. Confinement to bed for only twenty-four hours after operation. Arch Surg 1943 Aug 01;47(2):203. [doi: 10.1001/archsurg.1943.01220140085008]
3. Leithauser D. Early ambulation and related procedures in surgical management. JAMA 1947 Jul 19;134(12):1055. [doi: 10.1001/jama.1947.02880290069033]
4. Harper C, Lyles Y. Physiology and complications of bed rest. J Am Geriatr Soc 1988 Nov;36(11):1047-1054. [doi: 10.1111/j.1532-5415.1988.tb04375.x] [Medline: 3049751]
5. Fearon K, Ljungqvist O, Von Meyenfeldt M, Revhaug A, Dejong C, Lassen K, et al. Enhanced recovery after surgery: a consensus review of clinical care for patients undergoing colonic resection. Clin Nutr 2005 Jun;24(3):466-477. [doi: 10.1016/j.clnu.2005.02.002] [Medline: 15896435]
6. Guidelines. Enhanced Recovery After Surgery Society. URL: https://erassociety.org/guidelines/ [accessed 2022-06-23]
7. Gustafsson UO, Scott MJ, Hubner M, Nygren J, Demartines N, Francis N, et al. Guidelines for perioperative care in elective colorectal surgery: enhanced recovery after surgery (ERAS) society recommendations: 2018. World J Surg 2019 Mar;43(3):659-695. [doi: 10.1007/s00268-018-4844-y] [Medline: 30426190]

8. Castelino T, Fiore JF, Niculiseanu P, Landry T, Augustin B, Feldman LS. The effect of early mobilization protocols on postoperative outcomes following abdominal and thoracic surgery: a systematic review. Surgery 2016 Apr;159(4):991-1003. [doi: 10.1016/j.surg.2015.11.029] [Medline: 26804821]
9. Pashikanti L, Von Ah D. Impact of early mobilization protocol on the medical-surgical inpatient population: an integrated review of literature. Clin Nurse Spec 2012;26(2):87-94. [doi: 10.1097/NUR.0b013e31824590e6] [Medline: 22336934]
10. Ammann BC, Knols RH, Baschung P, de Bie RA, de Bruin ED. Application of principles of exercise training in sub-acute and chronic stroke survivors: a systematic review. BMC Neurol 2014 Aug 22;14(1):167 [FREE Full text] [doi: 10.1186/s12883-014-0167-2] [Medline: 25162455]
11. Bell DR, Pfeiffer KA, Cadmus-Bertram LA, Trigsted SM, Kelly A, Post EG, et al. Objectively measured physical activity in patients after anterior cruciate ligament reconstruction. Am J Sports Med 2017 Jul 18;45(8):1893-1900 [FREE Full text] [doi: 10.1177/0363546517698940] [Medline: 28419817]
12. Wearables unit shipments worldwide from 2014 to 2020. Statista. URL: https://www.statista.com/statistics/437871/wearables-worldwide-shipments/ [accessed 2022-06-23]
13. Shin G, Jarrahi MH, Fei Y, Karami A, Gafinowitz N, Byun A, et al. Wearable activity trackers, accuracy, adoption, acceptance and health impact: a systematic literature review. J Biomed Inform 2019 May;93:103153 [FREE Full text] [doi: 10.1016/j.jbi.2019.103153] [Medline: 30910623]
14. PROSPERO search. National Institute for Health Research. URL: https://www.crd.york.ac.uk/prospero/#searchadvanced [accessed 2021-10-10]
15. Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ 2021 Mar 29;372:n160 [FREE Full text] [doi: 10.1136/bmj.n160] [Medline: 33781993]
16. Haddaway NR, Collins AM, Coughlin D, Kirk S. The role of Google scholar in evidence reviews and its applicability to grey literature searching. PLoS One 2015 Sep 17;10(9):e0138237 [FREE Full text] [doi: 10.1371/journal.pone.0138237] [Medline: 26379270]
17. Kim SY, Park JE, Lee YJ, Seo H, Sheen S, Hahn S, et al. Testing a tool for assessing the risk of bias for nonrandomized studies showed moderate reliability and promising validity. J Clin Epidemiol 2013 Apr;66(4):408-414. [doi: 10.1016/j.jclinepi.2012.09.016] [Medline: 23337781]
18. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898 [FREE Full text] [doi: 10.1136/bmj.l4898] [Medline: 31462531]
19. Jonker LT, Hendriks S, Lahr MM, van Munster BC, de Bock GH, van Leeuwen BL. Postoperative recovery of accelerometer-based physical activity in older cancer patients. Eur J Surg Oncol 2020 Nov;46(11):2083-2090 [FREE Full text] [doi: 10.1016/j.ejso.2020.06.012] [Medline: 32682650]
20. Rivas E, Cohen B, Pu X, Xiang L, Saasouh W, Mao G, et al. Pain and opioid consumption and mobilization after surgery: post hoc analysis of two randomized trials. Anesthesiology 2022 Jan 01;136(1):115-126 [FREE Full text] [doi: 10.1097/ALN.0000000000004037] [Medline: 34780602]
21. Asada J, Ida M, Sato M, Okamoto N, Kawaguchi M. Associated factors with delayed ambulation after abdominal surgery. J Anesth 2019 Dec 16;33(6):680-684. [doi: 10.1007/s00540-019-02696-4] [Medline: 31620877]
22. Robinson TN, Carmichael H, Hosokawa P, Overbey DM, Goode CM, Barnett CC, et al. Decreases in daily ambulation forecast post-surgical re-admission. Am J Surg 2022 May;223(5):857-862. [doi: 10.1016/j.amjsurg.2021.08.011] [Medline: 34392912]
23. Kovar A, Carmichael H, Jones TS, Hosokawa P, Goode CM, Overbey DM, et al. Early identification of patients at risk for delayed recovery of ambulation after elective abdominal surgery. Surg Endosc 2022 Jul 09;36(7):4828-4833. [doi: 10.1007/s00464-021-08829-9] [Medline: 34755234]
24. Inouez Y, Kimura T, Fujita S, Noro H, Nishikawa K, Yumiba T, et al. A new parameter for assessing postoperative recovery of physical activity using an accelerometer. Surg Today 2003 Sep 1;33(9):645-650. [doi: 10.1007/s00595-003-2582-7] [Medline: 12928838]
25. Inoue Y, Kimura T, Noro H, Yoshikawa M, Nomura M, Yumiba T, et al. Is laparoscopic colorectal surgery less invasive than classical open surgery? Quantitation of physical activity using an accelerometer to assess postoperative convalescence. Surg Endosc 2003 Aug 1;17(8):1269-1273. [doi: 10.1007/s00464-002-8876-9] [Medline: 12799886]
26. McGuinness LA. Risk of bias plots. In: Doing Meta-Analysis with R: A Hands-On Guide. Boca Raton, Florida, United States: CRC Press; 2021.
27. Browning L, Denehy L, Scholes RL. The quantity of early upright mobilisation performed following upper abdominal surgery is low: an observational study. Aust J Physiother 2007;53(1):47-52. [doi: 10.1016/s0004-9514(07)70061-2] [Medline: 17326738]
28. Low C, Bovbjerg D, Ahrendt S, Choudry MH, Holtzman M, Jones HL, et al. Fitbit step counts during inpatient recovery from cancer surgery as a predictor of readmission. Ann Behav Med 2018 Jan 05;52(1):88-92 [FREE Full text] [doi: 10.1093/abm/kax022] [Medline: 29538623]

29. Jønsson LR, Ingelsrud LH, Tengberg LT, Bandholm T, Foss NB, Kristensen MT. Physical performance following acute high-risk abdominal surgery: a prospective cohort study. Can J Surg 2018 Feb 01;61(1):42-49 [FREE Full text] [doi: 10.1503/cjs.012616] [Medline: 29368676]
30. Daskivich TJ, Houman J, Lopez M, Luu M, Fleshner P, Zaghiyan K, et al. Association of wearable activity monitors with assessment of daily ambulation and length of stay among patients undergoing major surgery. JAMA Netw Open 2019 Feb 01;2(2):e187673 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.7673] [Medline: 30707226]
31. Martin D, Romain B, Pache B, Vuagniaux A, Guarnero V, Hahnloser D, et al. Physical activity and outcomes in colorectal surgery: a pilot prospective cohort study. Eur Surg Res 2020 Jun 3;61(1):23-33. [doi: 10.1159/000507578] [Medline: 32492676]
32. Nevo Y, Shaltiel T, Constantini N, Rosin D, Gutman M, Zmora O, et al. Activity tracking after surgery: does it correlate with postoperative complications? Am Surg 2022 Feb 30;88(2):226-232. [doi: 10.1177/0003134820988818] [Medline: 33522277]
33. Iida H, Maehira H, Mori H, Takebayashi K, Kojima M, Ueki T, et al. Usefulness of measuring temporal changes in physical activity levels using an accelerometer for prediction and early detection of postoperative complications after hepatectomy. HPB (Oxford) 2022 Jan;24(1):57-64 [FREE Full text] [doi: 10.1016/j.hpb.2021.05.011] [Medline: 34158231]
34. Yi Y, Sossenheimer PH, Erondu AI, Skowron KB, Rai V, Singer JM, et al. Using wearable biosensors to predict length of stay for patients with IBD after bowel surgery. Dig Dis Sci 2022 Mar 24;67(3):844-853 [FREE Full text] [doi: 10.1007/s10620-021-06910-w] [Medline: 33761092]
35. Kane WJ, Hassinger TE, Myers EL, Chu DL, Charles AN, Hoang SC, et al. Wearable technology and the association of perioperative activity level with 30-day readmission among patients undergoing major colorectal surgery. Surg Endosc 2022 Feb;36(2):1584-1592. [doi: 10.1007/s00464-021-08449-3] [Medline: 33782756]
36. Fiore JJ, Castelino T, Pecorelli N, Niculiseanu P, Balvardi S, Hershorn O, et al. Ensuring early mobilization within an enhanced recovery program for colorectal surgery: a randomized controlled trial. Ann Surg 2017 Aug;266(2):223-231. [doi: 10.1097/SLA.0000000000002114] [Medline: 27997472]
37. Ni C, Wang Z, Huang Z, Zhou H, Fu L, Cai H, et al. Early enforced mobilization after liver resection: a prospective randomized controlled trial. Int J Surg 2018 Jun;54(Pt A):254-258 [FREE Full text] [doi: 10.1016/j.ijsu.2018.04.060] [Medline: 29753000]
38. Wolk S, Linke S, Bogner A, Sturm D, Meißner T, Müssle B, et al. Use of activity tracking in major visceral surgery-the enhanced perioperative mobilization trial: a randomized controlled trial. J Gastrointest Surg 2019 Jun 8;23(6):1218-1226. [doi: 10.1007/s11605-018-3998-0] [Medline: 30298422]
39. Waller GC, Kim TG, Perez S, Esper GJ, Srinivasan JK, Shaffer VO, et al. Comparing activity trackers with vs. without alarms to increase postoperative ambulation: a randomized control trial. Am Surg 2021 Jul 14;87(7):1093-1098. [doi: 10.1177/0003134820973364] [Medline: 33316165]
40. No JH, Kim K, Kim YB, Suh DH, Yang EJ, Hwang H, et al. Effects of an activity tracker with feedback on physical activity in women after midline laparotomy: a randomized controlled trial. J Obstet Gynaecol Res 2021 Jul 25;47(7):2544-2550. [doi: 10.1111/jog.14807] [Medline: 33899302]
41. Steffens D, Solomon MJ, Beckenkamp PR, Koh CE, Yeo D, Sandroussi C, Fit-4-Home Collaborators, et al. Individualised, targeted step count intervention following gastrointestinal cancer surgery: the Fit-4-Home randomised clinical trial. ANZ J Surg 2022 Apr 22;92(4):703-711. [doi: 10.1111/ans.17212] [Medline: 34553480]
42. Lim SE, Ibrahim K, Sayer AA, Roberts HC. Assessment of physical activity of hospitalised older adults: a systematic review. J Nutr Health Aging 2018 May 19;22(3):377-386. [doi: 10.1007/s12603-017-0931-2] [Medline: 29484351]
43. Peel N, Kuys S, Klein K. Gait speed as a measure in geriatric assessment in clinical settings: a systematic review. J Gerontol A Biol Sci Med Sci 2013 Jan;68(1):39-46. [doi: 10.1093/gerona/gls174] [Medline: 22923430]
44. Mudge AM, McRae P, McHugh K, Griffin L, Hitchen A, Walker J, et al. Poor mobility in hospitalized adults of all ages. J Hosp Med 2016 Apr 21;11(4):289-291. [doi: 10.1002/jhm.2536] [Medline: 26797978]
45. Clevenger KA, Pfeiffer KA, Mackintosh KA, McNarry MA, Brønd J, Arvidsson D, et al. Effect of sampling rate on acceleration and counts of hip- and wrist-worn ActiGraph accelerometers in children. Physiol Meas 2019 Sep 30;40(9):095008. [doi: 10.1088/1361-6579/ab444b] [Medline: 31518999]
46. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, et al. Reliability and validity of commercially available wearable devices for measuring steps, energy expenditure, and heart rate: systematic review. JMIR Mhealth Uhealth 2020 Sep 08;8(9):e18694 [FREE Full text] [doi: 10.2196/18694] [Medline: 32897239]
47. Mikkelsen MK, Berg-Beckhoff G, Frederiksen P, Horgan G, O'Driscoll R, Palmeira AL, et al. Estimating physical activity and sedentary behaviour in a free-living environment: a comparative study between Fitbit Charge 2 and Actigraph GT3X. PLoS One 2020 Jun 11;15(6):e0234426 [FREE Full text] [doi: 10.1371/journal.pone.0234426] [Medline: 32525912]
48. Rosenberger ME, Buman MP, Haskell WL, McConnell MV, Carstensen LL. Twenty-four hours of sleep, sedentary behavior, and physical activity with nine wearable devices. Med Sci Sports Exerc 2016 Mar;48(3):457-465 [FREE Full text] [doi: 10.1249/MSS.0000000000000778] [Medline: 26484953]

49. Tedesco S, Sica M, Ancillao A, Timmons S, Barton J, O'Flynn B. Accuracy of consumer-level and research-grade activity trackers in ambulatory settings in older adults. PLoS One 2019 May 21;14(5):e0216891 [FREE Full text] [doi: 10.1371/journal.pone.0216891] [Medline: 31112585]
50. Toth LP, Park S, Springer CM, Feyerabend MD, Steeves JA, Bassett DR. Video-recorded validation of wearable step counters under free-living conditions. Med Sci Sports Exerc 2018 Jun;50(6):1315-1322. [doi: 10.1249/MSS.0000000000001569] [Medline: 29381649]
51. Bassett DR, Toth LP, LaMunion SR, Crouter SE. Step counting: a review of measurement considerations and health-related applications. Sports Med 2017 Jul 22;47(7):1303-1315 [FREE Full text] [doi: 10.1007/s40279-016-0663-1] [Medline: 28005190]
52. Migueles JH, Cadenas-Sanchez C, Rowlands AV, Henriksson P, Shiroma EJ, Acosta FM, et al. Comparability of accelerometer signal aggregation metrics across placements and dominant wrist cut points for the assessment of physical activity in adults. Sci Rep 2019 Dec 03;9(1):18235 [FREE Full text] [doi: 10.1038/s41598-019-54267-y] [Medline: 31796778]
53. Anderson JL, Green AJ, Yoward LS, Hall HK. Validity and reliability of accelerometry in identification of lying, sitting, standing or purposeful activity in adult hospital inpatients recovering from acute or critical illness: a systematic review. Clin Rehabil 2018 Feb 14;32(2):233-242. [doi: 10.1177/0269215517724850] [Medline: 28805075]
54. Goldsack JC, Coravos A, Bakker JP, Bent B, Dowling AV, Fitzer-Attas C, et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). NPJ Digit Med 2020;3:55 [FREE Full text] [doi: 10.1038/s41746-020-0260-4] [Medline: 32337371]
55. Baldwin C, van Kessel G, Phillips A, Johnston K. Accelerometry shows inpatients with acute medical or surgical conditions spend little time upright and are highly sedentary: systematic review. Phys Ther 2017 Nov 01;97(11):1044-1065. [doi: 10.1093/ptj/pzx076] [Medline: 29077906]
56. Fazio S, Stocking J, Kuhn B, Doroy A, Blackmon E, Young HM, et al. How much do hospitalized adults move? A systematic review and meta-analysis. Appl Nurs Res 2020 Feb;51:151189 [FREE Full text] [doi: 10.1016/j.apnr.2019.151189] [Medline: 31672262]
57. Montoye AH, Moore RW, Bowles HR, Korycinski R, Pfeiffer KA. Reporting accelerometer methods in physical activity intervention studies: a systematic review and recommendations for authors. Br J Sports Med 2018 Dec 18;52(23):1507-1516. [doi: 10.1136/bjsports-2015-095947] [Medline: 27539504]

Abbreviations

ERAS: Enhanced Recovery After Surgery
MeSH: Medical Subject Headings
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
RoB 2: revised Cochrane risk-of-bias tool for randomized trials
RoBANS: risk-of-bias assessment tool for nonrandomized studies

Edited by T Leung, G Eysenbach; submitted 19.02.23; this is a non–peer-reviewed article;accepted 28.03.23; published 15.05.23.

Please cite as:
Fuchita M, Ridgeway KJ, Kimzey C, Melanson EL, Fernandez-Bustamante A
Accelerometer-Measured Inpatient Physical Activity and Associated Outcomes After Major Abdominal Surgery: Systematic Review
Interact J Med Res 2023;12:e46629
URL: https://www.i-jmr.org/2023/1/e46629
doi:10.2196/46629
PMID:37184924

Case Report

Acute Spontaneous Colonic Perforation in a Case of Newly Confirmed Scleroderma: Case Report

Glenn Goodwin[1], DO; Christian Ryckeley[1], MD; Davide Fox[2], MD; Michael Ashley[3], MD; Laurence Dubensky[1], MD; Mauricio Danckers[2], MD; Todd Slesinger[1], MD

[1]Emergency Department, Aventura Hospital and Medical Center, Aventura, FL, United States
[2]Department of Critical Care, Aventura Hospital and Medical Center, Aventura, FL, United States
[3]Department of Radiology, Aventura Hospital and Medical Center, Aventura, FL, United States

Corresponding Author:
Glenn Goodwin, DO
Emergency Department
Aventura Hospital and Medical Center
20900 Biscayne Blvd
Aventura, FL, 33180
United States
Phone: 1 3056923392
Email: ggoodwin954@gmail.com

Abstract

Scleroderma is a group of autoimmune diseases that principally affects the skin, blood vessels, muscles, and viscera. One of the more well-known subgroups of scleroderma is the limited cutaneous form of the multisystem connective tissue disorder known as CREST (calcinosis, Raynaud phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasis) syndrome. In this report, we present a case of a spontaneous colonic bowel perforation in a patient with incomplete features of CREST. Our patient underwent a complicated hospital course involving broad-spectrum antibiotic coverage, surgical hemicolectomy, and immunosuppressives. She was eventually discharged home with a return to functional baseline status after esophageal dysmotility confirmation via manometry. Physicians managing patients with scleroderma ensuing to an emergency department encounter must anticipate the multitude of complications that can occur, as was seen in our patient. The threshold for pursuing imaging and additional tests, in addition to admission, should be relatively low, given the extremely high rates of complications and mortality. Early multidisciplinary involvement with infectious disease, rheumatology, surgery, and other respective specialties is crucial for patient outcome optimization.

(Interact J Med Res 2023;12:e43295) doi:10.2196/43295

KEYWORDS

scleroderma; systemic sclerosis; spontaneous bowel perforation; CREST syndrome; calcinosis, Raynaud phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasis; multisystem connective tissue disorder; sclerosis; skin; dermatology; internal medicine; autoimmune; perforation; gastroenterology; esophagus; esophageal; connective tissue; emergency; gastrointestinal; case report

Introduction

Background

Scleroderma, sometimes more colloquially known as systemic sclerosis, is a group of autoimmune diseases that principally affects the skin, blood vessels, muscles, and viscera [1]. One of the more well-known subgroups of scleroderma is the limited cutaneous form of the multisystem connective tissue disorder known as CREST (calcinosis, Raynaud phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasis) syndrome [2]. Within the CREST syndrome spectrum of diseases exist numerous associated complications and conditions.

Objective

The focus of this paper will be to present a case of a spontaneous colonic bowel perforation in a patient with incomplete features of CREST. Currently, only a few case reports exist documenting this relatively rare but recognized complication. Typically, gastrointestinal (GI) complications seen in CREST syndrome involve the esophagus; however, evolving data are demonstrating that concomitant distal GI pathologies are extremely common [3-6], as exemplified by this case.

Ethical Considerations

We complied with all applicable laws and regulations concerning the privacy and security of patient personal information, including, but not limited to, the Health Insurance Portability and Accountability Act of 1996 and other US federal and state laws relating to the privacy and security of personally identifiable information. The patient provided her expressed and written consent for case report publication, with the family and coauthors present. Written consent was in accordance with the standardized hospital consent form. This case report was fully observational in nature and exempt from institutional review board approval.

Case Report

A 60-year-old woman with a past medical history of chronic constipation, diverticulosis, vagus nerve cardiac pacemaker use, heavy tobacco use, and incomplete CREST syndrome presented to the emergency department (ED) for sudden onset of diffuse abdominal tenderness that began while watching television on the Ukrainian-Russian war. The patient was recently displaced from Ukraine and endorsed experiencing severe emotional stress, culminating in the abdominal pain episode that brought her into the ED. Additionally, due to the war, her medical records could not be obtained, and she only had a scant recollection of her previous medical conditions, limiting her definitive medical history. The patient had reportedly been diagnosed with one of the scleroderma spectrum of diseases approximately 10 years prior to presentation but had not been on any pharmacologic interventions for it thus far.

Upon arriving at the ED, the patient's vital signs were significant for a blood pressure of 228/100 but otherwise normal with a pulse of 82 beats per minute, a respiratory rate of 14 breaths per minute, and an oxygen saturation level of 98% on room air. The patient's physical examination was significant for a midline sternotomy scar from prior open heart surgery along with severe and diffuse abdominal tenderness. Serum laboratory values revealed an elevated lactic acid level of 4.0 mmol/L, along with a normal complete blood count, electrolytes, and coagulation values. The patient's blood cultures subsequently grew *Bacteroides vulgatus*. The patient underwent a computed tomography scan of the abdomen and pelvis with intravenous contrast, which demonstrated a perforated descending colon along with adjacent intraperitoneal stool, free fluid and air, and pneumatosis (Figure 1). Other notable findings included moderate distention of the esophagus, likely secondary to reflux in addition to pneumobilia in the absence of a gallbladder.

The patient was immediately started on piperacillin-tazobactam in the ED followed by an emergent general surgery consultation. She underwent an exploratory laparotomy on the same day. The surgeons reported severe distention of the distal colon along with extensive adhesions. It was later revealed that the patient had undergone a previous exploratory laparotomy several years prior, which resulted in an open cholecystectomy and appendectomy.

The surgery concluded with lysis of adhesions, a partial left hemicolectomy and Hartman procedure, and end colostomy creation. The extensive GI findings precluded the complete closing of the initial midline entry incision, which was consequently left to heal by secondary closure. The patient was subsequently admitted to the intensive care unit (ICU).

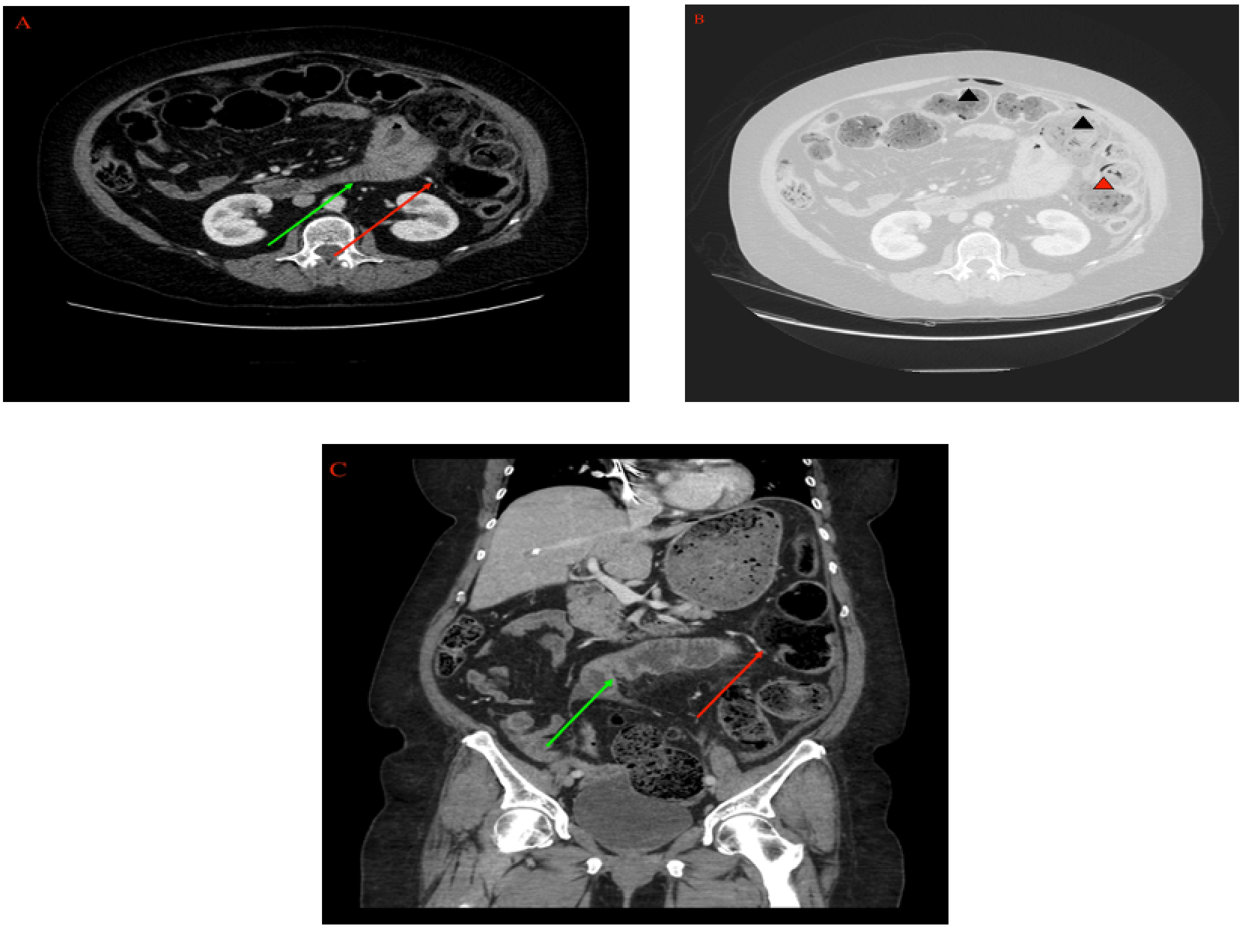
While in the ICU, the patient experienced numerous complications including *Candida albicans* fungemia, *Bacteroides vulgatus* bacteremia, and exacerbation of her preexisting sclerodactyly as demonstrated by the worsening of her finger stiffness and swelling. The patient was noted to have persistently swollen laryngeal and pharyngeal tissues, making intubations extremely difficult. After a long and complicated ICU course, she was eventually extubated on postoperative day 10 and was transferred out of the ICU on postoperative day 13.

The patient's blood infections were effectively treated with piperacillin-tazobactam, followed by meropenem and fluconazole, and then by micafungin. Serum laboratory values demonstrated complete resolution of all previously abnormal values.

While a biopsy is needed to definitively diagnose scleroderma, her scleroderma diagnosis was substantiated in the ICU with positive antinuclear and anticentromere antibodies along with an American College of Rheumatology/European League Against Rheumatism (ACR/EULAR) score of 15. The ACR/EULAR classification criteria are used to aid in the diagnosis of several rheumatological diseases, and scores of 9 or above in this validated scoring system are associated with scleroderma [6].

Additionally, the patient had baseline esophageal dysmotility, which was confirmed with manometry during her stay in the ICU. She was discharged home with outpatient rheumatology follow-up along with a daily steroid regimen soon after her transfer out of the ICU, with a general return to baseline functionality.

Figure 1. Systemic scleroderma presenting with spontaneous perforation: (A, B) axial and coronal images of contrast-enhanced computerized tomography at presentation demonstrating a localized perforation of the descending colon (solid arrows) with adjacent reactive small bowel wall thickening (green arrows); and (C) axial image in the lung window demonstrating the foci of free air (black arrowheads) in the anterior peritoneal cavity tracking from the perforated large bowel. Additional foci of pneumatosis coli are demonstrated in the adjacent descending colon (red arrowheads).



Discussion

Principal Findings

While up to 90% of patients with scleroderma are found to have some form of GI involvement, only 50% report symptom manifestation [7]. Notwithstanding that esophageal complications have been the most reported GI-related feature, colonic involvement is found almost as frequently, particularly in patients with abnormal esophageal manometry studies [3,7,8], as seen in our patient. Limited data exist relating to the correlation between the extent of disease and mortality; however, 1 study found that approximately 10% of deaths related to scleroderma were due to GI complications [9]. While many studies have demonstrated unfavorable patient outcomes relating to the lung and cardiac manifestations of the disease, GI involvement, particularly distal to the esophagus, portends poor survival [9,10]. While our patient's definitive cardiac condition could not be determined, her history of open-heart surgery and use of the vagus nerve pacemaker may have been related to her scleroderma. Myocardial and vascular compromise has been well documented in patients with scleroderma, possibly accounting for the additional surprising feature of our patient being persistently hypertensive, even while septic [8].

Oropharyngeal dysphagia and deglutination abnormalities are found in up to 25% of patients with scleroderma [10]. Our patient's exceptionally difficult airway may have reasonably been explained by these abnormalities. The patient was found to have extensive edema to the oropharynx and proximal larynx, refractory to high-dose steroids. Furthermore, the patient had repeated episodes of postextubation upper airway swelling, necessitating multiple reintubations.

In addition to the aforementioned complications, the patient also experienced several refractory and intractable infections. A possible explanation relates to findings from a study in 2022 by Kristofer et al [11], which explored the dysbiosis seen in many patients with scleroderma. The interplay between the hyperactive immune cells in the gut and the microbiome may have been responsible for many previously unexplained complications in patients with scleroderma [11]. Further compounding the interplay is the ubiquitous microvasculopathy and gut wall damage seen in scleroderma [8]. This patient's extensive and prolonged hospital course could possibly be attributed to her underlying pathology, especially when considering the high mortality and morbidity associated with surgically repaired bowel perforations [12]. Bowel perforations that require surgery in otherwise healthy patients have been found to have an overall mortality rate of 10% to 15% and a morbidity rate of 20% to 30% [12].

Conclusion

While much research exists regarding esophageal complications in patients with scleroderma, it is crucial for the clinician to consider the extremely high rates of extraesophageal GI involvement of the disease. This consideration may compel the emergency physician to have a lower threshold for additional testing and actions such as cross-sectional imaging, lab tests, and specialist consultations. This recommendation is particularly germane for patients who are not being treated for their disease, as was the case with our patient. Additionally, physicians managing patients with scleroderma ensuing to an ED encounter must anticipate the multitude of complications that can occur. Preemptively involving surgical and anesthesia teams early in the inpatient course is both reasonable and appropriate. Additionally, early multidisciplinary inpatient involvement with infectious disease and rheumatology is crucial for outcome optimization. While the focus of this paper was on the relatively rare complication of spontaneous intestinal perforation, the secondary objective was to illuminate the possibility of many others.

Acknowledgments

Yelixa Santos Roman and Safira Saint Forte were collaborators in this study.

Conflicts of Interest

None declared.

References

1. Barnes J, Mayes MD. Epidemiology of systemic sclerosis: incidence, prevalence, survival, risk factors, malignancy, and environmental triggers. Curr Opin Rheumatol 2012;24(2):165-170. [doi: 10.1097/bor.0b013e32834ff2e8]
2. Meyer O. CREST syndrome. Ann Med Interne (Paris) 2002 May;153(3):183-188. [Medline: 12218901]
3. Stupalkowska W, Mahler-Araujo B, Bennett J, Gourgiotis S. Spontaneous bowel perforation in the setting of colonic involvement with scleroderma. BMJ Case Rep 2020 Jan 05;13(1):e233220 [FREE Full text] [doi: 10.1136/bcr-2019-233220] [Medline: 31907219]
4. Richard N, Hudson M, Wang M, Gyger G, Proudman S, Stevens W, Canadian Scleroderma Research Group (CSRG), Australian Scleroderma Interest Group (ASIG), et al. Severe gastrointestinal disease in very early systemic sclerosis is associated with early mortality. Rheumatology (Oxford) 2019 Apr 01;58(4):636-644 [FREE Full text] [doi: 10.1093/rheumatology/key350] [Medline: 30517716]
5. McFarlane IM, Bhamra MS, Kreps A, Iqbal S, Al-Ani F, Saladini-Aponte C, et al. Gastrointestinal manifestations of systemic sclerosis. Rheumatology (Sunnyvale) 2018;8(1) [FREE Full text] [doi: 10.4172/2161-1149.1000235] [Medline: 30057856]
6. van den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, et al. 2013 classification criteria for systemic sclerosis: an American college of rheumatology/European league against rheumatism collaborative initiative. Ann Rheum Dis 2013 Nov 03;72(11):1747-1755 [FREE Full text] [doi: 10.1136/annrheumdis-2013-204424] [Medline: 24092682]
7. Thoua NM, Bunce C, Brough G, Forbes A, Emmanuel AV, Denton CP. Assessment of gastrointestinal symptoms in patients with systemic sclerosis in a UK tertiary referral centre. Rheumatology (Oxford) 2010 Sep 08;49(9):1770-1775. [doi: 10.1093/rheumatology/keq147] [Medline: 20530510]
8. Kumar S, Singh J, Rattan S, DiMarino AJ, Cohen S, Jimenez SA. Review article: pathogenesis and clinical manifestations of gastrointestinal involvement in systemic sclerosis. Aliment Pharmacol Ther 2017 Apr 09;45(7):883-898 [FREE Full text] [doi: 10.1111/apt.13963] [Medline: 28185291]
9. Altman RD, Medsger TA, Bloch DA, Michel BA. Predictors of survival in systemic sclerosis (scleroderma). Arthritis Rheum 1991 Apr;34(4):403-413 [FREE Full text] [doi: 10.1002/art.1780340405] [Medline: 1901491]
10. Bharadwaj S, Tandon P, Gohel T, Corrigan ML, Coughlin KL, Shatnawei A, et al. Gastrointestinal manifestations, malnutrition, and role of enteral and parenteral nutrition in patients with scleroderma. J Clin Gastroenterol 2015 Aug;49(7):559-564. [doi: 10.1097/MCG.0000000000000334] [Medline: 25992813]
11. Andréasson K, Lee SM, Lagishetty V, Wu M, Howlett N, English J, et al. Disease features and gastrointestinal microbial composition in patients with systemic sclerosis from two independent cohorts. ACR Open Rheumatol 2022 May 17;4(5):417-425 [FREE Full text] [doi: 10.1002/acr2.11387] [Medline: 35174673]
12. Shin R, Lee SM, Sohn B, Lee DW, Song I, Chai YJ, et al. Predictors of morbidity and mortality after surgery for intestinal perforation. Ann Coloproctol 2016 Dec;32(6):221-227 [FREE Full text] [doi: 10.3393/ac.2016.32.6.221] [Medline: 28119865]

Abbreviations

ACR/EULAR: American College of Rheumatology/European League Against Rheumatism
CREST: calcinosis, Raynaud phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasis
ED: emergency department
GI: gastrointestinal

ICU: intensive care unit

Edited by A Mavragani; submitted 07.10.22; peer-reviewed by S Ahmed, S Arya, N Nguon; comments to author 06.02.23; revised version received 06.02.23; accepted 24.02.23; published 08.03.23.

Please cite as:
Goodwin G, Ryckeley C, Fox D, Ashley M, Dubensky L, Danckers M, Slesinger T
Acute Spontaneous Colonic Perforation in a Case of Newly Confirmed Scleroderma: Case Report
Interact J Med Res 2023;12:e43295
URL: https://www.i-jmr.org/2023/1/e43295
doi:10.2196/43295
PMID:36862558

Case Report

Use of Wearable Devices for Peak Oxygen Consumption Measurement in Clinical Cardiology: Case Report and Literature Review

Gabriella Bayshtok[1,2], BA; Shmuel Tiosano[1,2,3], MPH, MD; Ariel Furer[3,4], MBA, MD

[1]Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel

[2]Arrow Program for Medical Research Education, Sheba Medical Center Tel Hashomer, Ramat Gan, Israel

[3]Leviev Heart Center, Sheba Medical Center, Ramat Gan, Israel

[4]Department of Military Medicine, Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel

Corresponding Author:
Ariel Furer, MBA, MD
Leviev Heart Center
Sheba Medical Center
Derech Sheba 2
Ramat Gan
Israel
Phone: 972 52 927 7372
Email: furera@gmail.com

Abstract

Background: Oxygen consumption is an important index to evaluate in cardiac patients, particularly those with heart failure, and is measured in the setting of advanced cardiopulmonary exercise testing. However, technological advances now allow for the estimation of this parameter in many consumer and medical-grade wearable devices, making it available for the medical provider at the initial evaluation of patients. We report a case of an apparently healthy male aged 40 years who presented for evaluation due to an Apple Watch (Apple Inc) notification of low cardiac fitness. This alert triggered a thorough workup, revealing a diagnosis of familial nonischemic cardiomyopathy with severely reduced left ventricular systolic function. While the use of wearable devices for the measurement of oxygen consumption and related parameters is promising, further studies are needed for validation.

Objective: The aim of this report is to investigate the potential utility of wearable devices as a screening and risk stratification tool for cardiac fitness for the general population and those with increased cardiovascular risk, particularly through the measurement of peak oxygen consumption (VO_2). We discuss the possible advantages of measuring oxygen consumption using wearables and propose its integration into routine patient evaluation and follow-up processes. With the current evidence and limitations, we encourage researchers and clinicians to explore bringing wearable devices into clinical practice.

Methods: The case was identified at Sheba Medical Center, and the patient's cardiac fitness was monitored through an Apple Watch Series 6. The patient underwent a comprehensive cardiac workup following his presentation. Subsequently, we searched the literature for articles relating to the clinical utility of peak VO_2 monitoring and available wearable devices.

Results: The Apple Watch data provided by the patient demonstrated reduced peak VO_2, a surrogate index for cardiac fitness, which improved after treatment initiation. A cardiological workup confirmed familial nonischemic cardiomyopathy with severely reduced left ventricular systolic function. A review of the literature revealed the potential clinical benefit of peak VO_2 monitoring in both cardiac and noncardiac scenarios. Additionally, several devices on the market were identified that could allow for accurate oxygen consumption measurement; however, future studies and approval by the Food and Drug Administration (FDA) are still necessary.

Conclusions: This case report highlights the potential utility of peak VO_2 measurements by wearable devices for early identification and screening of cardiac fitness for the general population and those at increased risk of cardiovascular disease. The integration of wearable devices into routine patient evaluation may allow for earlier presentation in the diagnostic workflow. Cardiac fitness can be serially measured using the wearable device, allowing for close monitoring of functional capacity parameters. Devices need to be used with caution, and further studies are warranted.

(Interact J Med Res 2023;12:e45504) doi:10.2196/45504

KEYWORDS

cardiac fitness; cardiac patient; cardiorespiratory fitness; CRF; clinical cardiology; oxygen consumption; peak VO2; smartwatch; wearable device

Introduction

Oxygen consumption has been measured and appreciated in clinical cardiology for decades, most commonly in advanced cardiopulmonary exercise tests (CPETs) in a complex laboratory setup. It is a parameter used to evaluate cardiorespiratory fitness (CRF), thus named peak oxygen consumption (VO_2), representing an individual's largest volume of oxygen extracted from inhaled air during efforts [1]. This measurement is affected by age, gender, genetics, underlying medical conditions, and physical activity, especially high-intensity training [2-4]. The routine use of this parameter in the setting of CPET is to assess changes in cardiac capacity following exertional physical activity and serve as a validated prognostic factor in cardiovascular patients. This parameter further allows for the evaluation of the interactions between the cardiac, musculoskeletal, respiratory, and vascular systems [5] and is of great importance due to its association with decreased all-cause mortality [6,7]. Table S1 in Multimedia Appendix 1 [1,8-14] summarizes the definitions and abbreviations of the key terms.

Wearable devices have gained tremendous popularity in recent years and were traditionally divided into consumer products and medical-grade devices. However, these distinctions are rapidly fading as top-selling consumer devices now provide validated and regulatory-approved medical-grade measurements of physiologic parameters, including heart rate and oxygen saturation, using a 1-lead electrocardiogram and photoplethysmography. Companies such as Apple, Garmin, Fitbit, and Samsung distribute wearable devices that allow for validated peak VO_2 measurements using proprietary algorithms with a combination of heart rate and pace data during exercise compared to a user's baseline. As of 2018, the Food and Drug Administration (FDA) categorizes the Apple Watch (Apple Inc) as a class II device for over-the-counter use of an electrocardiogram (ECG) and photoplethysmography. These tools allow for the identification of cardiac conditions such as atrial fibrillation [15,16]. In 2020, Apple released a new version of its wearable device that estimates submaximal oxygen consumption; however, further FDA approval and classification are still needed for this utility. During Apple's peak VO_2 assessment, various sensors such as a photoplethysmograph, gyroscope, accelerometer, barometer, and GPS are manipulated to maximize the algorithm to estimate peak VO_2 in the general population. Study participants completed CPETs with treadmills and cycle ergometers while wearing the Apple Watch Series 4, and linear projections estimated peak VO_2 using age-predicted maximum heart rates in the submaximal range (at least a 30% increase in heart rate). Algorithm predictions were compared with the average of all CPET measurements (at least 6) for each participant. A comparison between the 2 modalities found the Apple Watch to be valid (mean difference of 1.4, SD 4.7), reliable (intraclass correlation coefficient [ICC] of 0.86), and consistent (median of 1.2 and 90th percentile SD per participant of 2.6 on the Apple Watch for those with at least five estimates). The Apple Watch estimates peak VO_2 with an average error of less than 1 metabolic equivalent (MET) compared to CPETs. Apple Watch peak VO_2 measurements were found to be more reliable than submaximal treadmill tests (ICC of 0.87 vs 0.75), which are performed if a patient cannot complete a full CPET. Limitations to algorithm performance include pacemakers, medical conditions causing chronotropic incompetence, exercise intolerance, and arrhythmias. Apple supports the development of measuring cardiac fitness and addresses limitations such as rate-limiting medications through algorithm adjustment [17]. Ongoing clinical trials are underway to investigate the use of the Apple Watch to measure cardiopulmonary fitness in ambulatory heart failure patients [18]. Table S2 in Multimedia Appendix 1 [1,8-14] includes a list of other devices used for this purpose and details the evidence available for their validation.

The technological advancement of wearable devices, combined with their fast distribution and adoption by many, sets the stage for using this data in the preemptive screening of wide populations, including apparently healthy individuals. To demonstrate this concept (illustrated in Multimedia Appendix 2), we present the case of a young and previously fit individual without previous medical conditions who was alerted to a decline in his cardiac fitness index from his Apple Watch despite being asymptomatic. We describe the initial patient workup until the final diagnosis and then review the body of knowledge available investigating oxygen consumption and its clinical uses. Finally, we review current data on technologies and devices in this field and the potential clinical benefits of routinely using such technologies.

Methods

The Apple Watch Series 6 was used for cardiac fitness measurements. The data were provided by the patient from the Apple Health app.

Case

A male patient aged 40 years presented following a notification from his Apple Watch indicating low cardiac fitness. The patient had been using the Apple Watch since November 2020 and received the alert in October 2021 during a trip abroad. His device revealed a progressive decline in cardiac fitness index, as shown in Figure 1. The patient had no known previous medical history, hospitalizations, or medical treatment. He was fit and engaged in high-intensity workouts several times a week. Other than occasional palpitations, he reported being asymptomatic during the months before this alert. He reported no history of smoking or use of illicit substances and no family history of cardiovascular disease, including sudden death or ischemic disease. At the time of presentation, however, he reported that his sister had been undergoing a cardiac workup at the same time due to suspected peripartum cardiomyopathy following supraventricular tachycardia (SVT) after delivery. Her echocardiography exam suggested mild globally reduced left ventricle (LV) systolic function and an ejection fraction (EF) of 45%.

Figure 1. Screenshot of the patient's Apple Watch cardiac fitness index from November 2020 to October 2021 that triggered the initial cardiac evaluation.

Initial workup included a stress echocardiography exam demonstrating reduced global LV systolic function and an LV EF of 20% (visually estimated) without valvular malformations or regional wall abnormalities. Following these results, the patient was hospitalized and underwent a comprehensive workup, including repeated echocardiography, a 24-holter exam, and cardiac magnetic resonance imaging (MRI). Repeat echocardiography demonstrated a mildly dilated LV, severe diffuse global LV dysfunction with an EF of 23%, grade I diastolic dysfunction, a right ventricle with normal size and function, a normal left atrium, an aortic valve with minimal regurgitation, mild mitral regurgitation, and minimal tricuspid regurgitation with normal systolic pulmonary pressure. The rest of the exam was within normal limits. Cardiac MRI demonstrated a dilated LV with an LV end-diastolic diameter of 63 mm. There were no signs of late gadolinium enhancement. Normal T1 mapping was up to 1150 milliseconds and T2 mapping was up to 52 milliseconds (with normal values up to 50 milliseconds); the measured LV EF was 47%. The next step in the cardiomyopathy workup was genetic testing, which revealed a titin mutation (*TTN*, exon 326, c.86116C>T [p.Arg28706*]).

During hospitalization, the patient was started on a β-blocker, mineralocorticoid receptor antagonist, an angiotensin II receptor blocker, and a sodium-glucose transport protein 2 inhibitor. After hospitalization, he began a cardiac rehabilitation program with no reports of symptoms on exertion; the New York Heart Association classification was 1. Additional echocardiography in December 2021 demonstrated a LV EF of 35%. Figure 2 depicts an increase in the patient's cardiac fitness index from November 2021 to September 2022 after treatment initiation.

Figure 2. Screenshot of the patient's Apple Watch cardiac fitness index from November 2021 to September 2022 after treatment initiation.

Ethics Approval

The participant has provided written informed consent. All patient data has been deidentified.

Discussion

Overview

In recent years, we have witnessed growth in the accessibility of oxygen consumption measurement and its potential uses in clinical cardiology. Peak VO_2 helps establish a reference for a patient's cardiac fitness and provides insight into risk stratification for disease status, cardiac rehabilitation, and perioperative status. Oxygen consumption improvement has been shown to confer a survival benefit and minimize disease progression [19], and therefore, measurement of this parameter is useful for both one-time assessment and routine follow-up. The continuous feedback of peak VO_2 through wearable devices may allow for patient engagement in health monitoring, promote lifestyle changes, and guide clinical decision-making [20]. This case of an asymptomatic male with low peak VO_2 due to underlying cardiomyopathy demonstrates the advantages of wearable devices as a screening tool in healthy and at-risk adults, as well as the ability to monitor cardiac fitness after treatment initiation. Many other cardiac and noncardiac pathologies may be detected earlier due to their correlation to changes in peak VO_2. We discuss the various pathologies and clinical scenarios that relate to this index below.

Oxygen consumption has been shown to be decreased in various cardiac conditions, including heart failure (HF), ischemic heart disease (IHD), atrial fibrillation (AF), valvular disease, cardiomyopathy, hypertension, and other pathologies [21-23].

HF patients often experience decreased quality of life with restriction of basic activities due to dyspnea and exercise intolerance, with peak VO_2 being a well-studied estimator of functional capacity among this patient population. A study of chronic systolic HF patients showed that a 6% increase in peak VO_2 was associated with improved clinical outcomes, including all-cause mortality and hospitalizations [24]. Additional studies found that increased peak VO_2 due to systemic and skeletal muscle adaptations from high-intensity training is associated with the preservation of ejection fraction and prevention of LV remodeling [25]. From a hemodynamic standpoint, oxygen consumption has a strong linear correlation with cardiac output [26], making peak VO_2 tracking a surrogate marker for cardiac output estimation, which is essential among HF patients. Another study found that better cardiac fitness in midlife (median age 49 years) is associated with a decreased risk of developing HF and subsequent hospitalization later in life, independent of other cardiac or noncardiac risk factors (1 MET increase was associated with a 17% risk reduction) [27]. Interestingly, in a study with 63 patients with chronic AF who underwent cardioversion to sinus rhythm, peak VO_2 was monitored before the procedure and 1 month after. Peak VO_2 max after the procedure significantly increased, suggesting that low peak VO_2 accompanies AF, perhaps related to tachycardia-induced cardiomyopathy [28]. Hypertension also warrants oxygen consumption monitoring, as a prospective study reported that those in a lower VO_2 max group were almost 2 times more at risk of developing hypertension and that striving for a higher VO_2 max may be protective against hypertension [29].

Peak VO_2 measurement is also relevant in noncardiac diseases. Studies have shown that increased oxygen consumption is associated with a lower risk of developing metabolic syndromes such as diabetes through several mechanisms. First, exercise, as reflected by improved VO_2, builds muscle that uses and removes glucose from the blood, while other reports discuss the potential role of oxidative pathway regulation in mitochondria as a potential linker between oxygen consumption and metabolic risk [30,31]. CRF has been found to have an inverse relationship with metabolic syndrome in males and females, with waist circumference being the strongest predicting factor [32]. Additional studies report that improved CRF is as effective as statin therapy in lowering the mortality risk in patients with dyslipidemia [33]. The benefit of increased peak VO_2 can also be seen among patients with chronic obstructive pulmonary disease (COPD) and is associated with reduced all-cause mortality [1]. Further investigation of the relationship between peak VO_2 and other chronic medical conditions is warranted, as low peak VO_2 is associated with ongoing disease progression in patients with rheumatoid arthritis, as estimated by inflammatory markers and subjective assessment. Cardiovascular risk factors linked to rheumatoid arthritis, such as atherosclerosis and changes in fat and muscle distribution, were decreased among patients with improved aerobic capacity [34]. Moreover, it is hypothesized that a higher peak VO_2 is protective against brain pathologies such as stroke, depression, and dementia [35-37]. To investigate the relationship between cardiorespiratory fitness and stroke risk, a study examined 16,878 asymptomatic men aged between 40 and 87 years whose fitness was followed over 10 years, reporting that a higher VO_2 was associated with a 68% lower risk of stroke death when compared to those with a low VO_2, and a moderate VO_2 was associated with a 63% decrease [35]. As discussed here, there is a widespread need for oxygen consumption evaluation among pathologies extending from cardiology, especially because many conditions may not produce symptoms until later in the disease course.

Cardiac rehabilitation is perhaps the best proof of practice for following peak VO_2 to reduce all-cause mortality and hospital admissions. For patients with acute coronary syndrome (ACS) or surgical interventions, cardiac rehabilitation is imperative for secondary prevention. Future directions for rehabilitation intend to maximize the incorporation of exercise to achieve greater aerobic capacity, which can be monitored through changes in peak VO_2 [38]. Additionally, reports comment on the advantage of objective physical activity and prognostic factor monitoring for personalized feedback, which relies less on patient recollection [39]. Peak VO_2 also allows for assessing patients' readiness for cardiovascular and noncardiovascular surgery, with studies demonstrating its association with postoperative complications and mortality in procedures such as gastrointestinal and vascular surgeries, hepatic transplantation, lung tumor resection, and coronary artery bypass grafting [40]. Thus, this dynamic parameter may be considered in the preoperative assessment, leading to improved decision-making and outcomes. A unique role for peak VO_2 measurement is evaluating a patient's need or readiness for heart transplantation. A study with 181 HF patients reported that the actuarial 1- and 2-year survival of the 89 patients who achieved a VO_2 equal to or <50% of predicted peak VO_2 was 74% and 43%, respectively, compared with 98% and 90% in the 92 patients who achieved >50% predicted peak VO_2 [41]. Other studies have investigated VO_2's role in triaging for cardiac transplantation and proposed a peak VO_2 greater than 14 ml/min/kg as an appropriate cutoff value [42]. Further validation studies are needed to assess whether oxygen consumption measurement is a suitable marker for heart transplantation.

As with this patient's case, most individuals with cardiomyopathy accompanied by decreased left ventricular systolic dysfunction are asymptomatic until the development of advanced disease [43]. It is difficult to estimate the number needed to screen to prevent cardiac events as the data concerning the validity of wearable peak VO_2 is preliminary and the true scope of the asymptomatic cardiac patients' burden is uncertain. However, serial peak VO_2 measurements have been shown to be advantageous in certain populations, such as those with congenital heart diseases. In a study with 1375 adult cyanotic and noncyanotic heart disease patients followed for 10 years while undergoing serial CPETs, the combination of peak VO_2 and heart rate reserve provided strong prognostic insight for mortality [44]. Another study in adults with a Fontan circulation who completed at least two maximal CPETs found that a decrease in peak VO_2 is a predictor of death or transplant, regardless of initial peak VO_2 [45]. These studies suggest that peak VO_2 measurements may be a useful prognostic marker for heart disease burden due to congenital heart defects, and perhaps in individuals similar to our patient after further studies are performed. We suggest there may be a place for using peak VO_2 data among both apparently healthy adults and those in high-risk populations as a first warning sign, triggering further workup. With most screening tools, there is a concern that false-positive data will cause unnecessary, costly, and, at times, hazardous workups. We believe that a more precise role for this tool will be only determined after a careful assessment is performed with robust scientific methodology and sufficient statistical power.

As reviewed here, a large body of evidence suggests the potential role of integrating peak VO_2 into various clinical scenarios. The option to access data routinely and continuously using a wearable device necessitates reconsidering peak VO_2's ideal place in clinical settings. Wearable devices can at least partially replace a patient's need to seek a CPET, which remains a logistically complex exam mostly limited to medical institutions, resulting in inconvenience for patients and poor accessibility. Reports recognize the potential of wearables for primary and secondary prevention but also identify barriers and limitations such as cost, data security, and false-positive results leading to unnecessary intervention and subsequent burden on the health care system [46]. The heterogeneity of methods used to collect data, along with the continuous device updates, similarly presents challenges for standardization [39]. Other reports have introduced a guide for integrating wearable devices into cardiovascular care with considerations for incorporating data into electronic medical records, staff education, index cutoffs, and frequency of review [47]. Further limitations of wearable devices include the inability to measure peak carbon dioxide production (VCO_2) and anaerobic threshold values, which serve as additional prognostic factors measured during CPETs. Peak VCO_2 is the amount of carbon dioxide exhaled from the body over time and reflects exercise capacity, whereas the anaerobic threshold is the point of substantially increased minute ventilation relative to VO_2 and represents a rise in lactic acid production. The anaerobic threshold suggests the functional capacity of HF patients, with a lower measurement indicating decreased capacity. It can also distinguish between cardiac and noncardiac conditions, with fatigue before the threshold point being less indicative of a cardiac condition [48].

We demonstrated that wearable devices could streamline the process from the onset of disease manifestation to therapeutic intervention. Indeed, a potential new trend that medical providers will need to address is the presentation of peak VO_2 or parallel indices by the patient through wearables at an early phase or as the trigger for a diagnostic workflow before other modalities, especially among HF patients. Moreover, routine peak VO_2 measurement in HF patients allows for another objective variable to assess, in addition to standard parameters such as weight, blood pressure, and heart rate, which addresses the issue of sparse, unbiased indices by broadening the opportunity for concrete data collection. Measuring this index continuously and remotely, which is not relevant for CPETs and other conventional methods that calculate peak VO_2, allows patients to present earlier in the usual diagnostic workflow. After diagnosis and treatment initiation, peak VO_2 can be monitored using the wearable device, as demonstrated in this patient's case. Health care providers and patients can be more informed about cardiac fitness during a treatment or rehabilitation regimen and accordingly adjust interventions while preventing clinical exacerbations. There is potential to improve clinical outcomes through enhanced monitoring of functional capacity parameters and cardiac fitness; however, devices should be used cautiously and only for adult populations in the appropriate clinical setting to prevent adverse use. Additional prospective studies are needed to enable this change while being supported with scientific evidence and an official classification as a medical-grade device. With the exponentially growing distribution of wearable devices, it may be less of a challenge to conduct these studies and obtain policy-changing evidence in a short period of time.

Conclusions

This report demonstrates the potential utility of routine peak VO_2 measurement for screening cardiac and noncardiac diseases to elucidate therapeutic options for patients in an accelerated timeframe and leverage widespread technology to initiate early interventions. Peak VO_2 provides a glimpse into risk stratification, rehabilitation, and perioperative care. Greater awareness of cardiorespiratory fitness among apparently healthy adults and those with high cardiovascular risk is needed and may result in impactful interventions.

Acknowledgments

The authors would like to thank Professor Michael Arad for assistance with data collection. This article was supported by grants from the Shalvi Fund for Medical Research and Innovation in Israel. All authors declared that they had insufficient or no funding to support open access publication of this manuscript, including from affiliated organizations or institutions, funding agencies, or other organizations. JMIR Publications provided article processing fee (APF) support for the publication of this article.

Data Availability

All data generated for this publication are included in the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Table S1: Definitions and abbreviations of key terms; Table S2: Comparison of devices on the market providing an estimation of oxygen consumption or surrogate indices.

[DOCX File , 19 KB - ijmr_v12i1e45504_app1.docx]

Multimedia Appendix 2

Graphical abstract. CPET: cardiopulmonary exercise testing; ECG: electrocardiogram; VO2: oxygen consumption.

[PNG File , 547 KB - ijmr_v12i1e45504_app2.png]

References

1. American Thoracic Society, American College of Chest Physicians. ATS/ACCP statement on cardiopulmonary exercise testing. Am J Respir Crit Care Med 2003;167(2):211-277 [FREE Full text] [doi: 10.1164/rccm.167.2.211] [Medline: 12524257]

2. Schutte NM, Nederend I, Hudziak JJ, Bartels M, de Geus EJC. Twin-sibling study and meta-analysis on the heritability of maximal oxygen consumption. Physiol Genomics 2016;48(3):210-219 [FREE Full text] [doi: 10.1152/physiolgenomics.00117.2015] [Medline: 26787216]
3. Zadro JR, Shirley D, Andrade TB, Scurrah KJ, Bauman A, Ferreira PH. The beneficial effects of physical activity: is it down to your genes? a systematic review and meta-analysis of twin and family studies. Sports Med Open 2017;3(1):4 [FREE Full text] [doi: 10.1186/s40798-016-0073-9] [Medline: 28074345]
4. Sultana RN, Sabag A, Keating SE, Johnson NA. The effect of low-volume high-intensity interval training on body composition and cardiorespiratory fitness: a systematic review and meta-analysis. Sports Med 2019;49(11):1687-1721. [doi: 10.1007/s40279-019-01167-w] [Medline: 31401727]
5. Stringer WW. Cardiopulmonary exercise testing: current applications. Expert Rev Respir Med 2010;4(2):179-188. [doi: 10.1586/ers.10.8] [Medline: 20406084]
6. Laukkanen JA, Zaccardi F, Khan H, Kurl S, Jae SY, Rauramaa R. Long-term change in cardiorespiratory fitness and all-cause mortality: a population-based follow-up study. Mayo Clin Proc 2016;91(9):1183-1188 [FREE Full text] [doi: 10.1016/j.mayocp.2016.05.014] [Medline: 27444976]
7. Mora S, Redberg RF, Cui Y, Whiteman MK, Flaws JA, Sharrett AR, et al. Ability of exercise testing to predict cardiovascular and all-cause death in asymptomatic women: a 20-year follow-up of the lipid research clinics prevalence study. JAMA 2003;290(12):1600-1607 [FREE Full text] [doi: 10.1001/jama.290.12.1600] [Medline: 14506119]
8. Automated Fitness Level (VO2max) Estimation with Heart Rate and Speed Data. Firstbeat Technologies Ltd. URL: https://assets.firstbeat.com/firstbeat/uploads/2017/06/white_paper_VO2max_30.6.2017.pdf [accessed 2023-06-23]
9. Kraft GL, Roberts RA. Validation of the garmin forerunner 920XT fitness watch VO2peak test. Int J Innov Educ Res 2017;5(2):63-69 [FREE Full text] [doi: 10.31686/ijier.vol5.iss2.619]
10. Helm MM, Carrier B, Davis DW, Cruz K, Barrios B, Navalta JW. Validation of the Garmin Fenix 6S Maximal Oxygen Consumption (VO2max) Estimate. Int J Exerc Sci 2021;14(1):29 [FREE Full text]
11. Klepin K, Wing D, Higgins M, Nichols J, Godino JG. Validity of cardiorespiratory fitness measured with fitbit compared to VO2max. Med Sci Sports Exerc 2019;51(11):2251-2256 [FREE Full text] [doi: 10.1249/MSS.0000000000002041] [Medline: 31107835]
12. FDA-Cleared Electrocardiogram Monitoring App is Available Starting Today on Galaxy Watch3 and Galaxy Watch Active2 in the New Samsung Health Monitor App. Samsung Newsroom U.S. 2021. URL: https://news.samsung.com/us/health-electrocardiogram-monitoring-app-ecg-galaxy-watch3-active2 [accessed 2023-06-23]
13. Shandhi MMH, Hersek S, Fan J, Sander E, De Marco T, Heller JA, et al. Wearable patch-based estimation of oxygen uptake and assessment of clinical status during cardiopulmonary exercise testing in patients with heart failure. J Card Fail 2020;26(11):948-958 Sawada S, Tanaka H, Funakoshi M, Shindo M, Kono S, Ishiko T [FREE Full text] [doi: 10.1016/j.cardfail.2020.05.014] [Medline: 32473379]
14. Shandhi MMH, Bartlett WH, Heller JA, Etemadi M, Young A, Plotz T, et al. Estimation of instantaneous oxygen uptake during exercise and daily activities using a wearable cardio-electromechanical and environmental sensor. IEEE J Biomed Health Inform 2021;25(3):634-646 [FREE Full text] [doi: 10.1109/JBHI.2020.3009903] [Medline: 32750964]
15. De Novo Classification request for ECG App. Food and Drug Administration. URL: https://www.accessdata.fda.gov/cdrh_docs/reviews/DEN180044.pdf [accessed 2023-06-23]
16. De Novo Classification Request for Irregular Rhythm Notification Feature. Food and Drug Administration. URL: https://www.accessdata.fda.gov/cdrh_docs/reviews/DEN180042.pdf [accessed 2023-06-23]
17. Using Apple Watch to Estimate Cardio Fitness with vo2 Max. Apple. URL: https://www.apple.com/healthcare/docs/site/Using_Apple_Watch_to_Estimate_Cardio_Fitness_with_VO2_max.pdf [accessed 2023-06-23]
18. Apple-CPET Ted Rogers Understanding Exacerbations of Heart Failure (TRUE-HF). ClinicalTrials.gov. 2022. URL: https://clinicaltrials.gov/ct2/show/NCT05008692 [accessed 2023-04-06]
19. Ross R, Blair SN, Arena R, Church TS, Després JP, Franklin BA, American Heart Association Physical Activity Committee of the Council on LifestyleCardiometabolic Health, Council on Clinical Cardiology, Council on Epidemiology Prevention, Council on Cardiovascular Stroke Nursing, Council on Functional Genomics Translational Biology, Stroke Council. Importance of assessing cardiorespiratory fitness in clinical practice: a case for fitness as a clinical vital sign: a scientific statement from the American Heart Association. Circulation 2016;134(24):e653-e699 [FREE Full text] [doi: 10.1161/CIR.0000000000000461] [Medline: 27881567]
20. Older PO, Levett DZH. Cardiopulmonary exercise testing and surgery. Ann Am Thorac Soc 2017;14(Supplement_1):S74-S83 [FREE Full text] [doi: 10.1513/AnnalsATS.201610-780FR] [Medline: 28511024]
21. Ueshima K, Myers J, Ribisl PM, Atwood JE, Morris CK, Kawaguchi T, et al. Hemodynamic determinants of exercise capacity in chronic atrial fibrillation. Am Heart J 1993;125(5 Pt 1):1301-1305. [doi: 10.1016/0002-8703(93)90998-o] [Medline: 8480581]
22. Scriven AJ, Lipkin DP, Fox KM, Poole-Wilson PA. Maximal oxygen uptake in severe aortic regurgitation: a different view of left ventricular function. Am Heart J 1990;120(4):902-909. [doi: 10.1016/0002-8703(90)90208-f] [Medline: 2220544]

23. Sawada S, Tanaka H, Funakoshi M, Shindo M, Kono S, Ishiko T. Five year prospective study on blood pressure and maximal oxygen uptake. Clin Exp Pharmacol Physiol 1993;20(7-8):483-487. [doi: 10.1111/j.1440-1681.1993.tb01729.x] [Medline: 8403528]
24. Swank AM, Horton J, Fleg JL, Fonarow GC, Keteyian S, Goldberg L, HF-ACTION Investigators. Modest increase in peak VO2 is related to better clinical outcomes in chronic heart failure patients: results from heart failure and a controlled trial to investigate outcomes of exercise training. Circ Heart Fail 2012;5(5):579-585 [FREE Full text] [doi: 10.1161/CIRCHEARTFAILURE.111.965186] [Medline: 22773109]
25. Strasser B, Burtscher M. Survival of the fittest: VOmax, a key predictor of longevity? Front Biosci (Landmark Ed) 2018;23(8):1505-1516 [FREE Full text] [doi: 10.2741/4657] [Medline: 29293447]
26. Weber KT, López-Antúnez L, Janicki JS. Cardiopulmonary exercise (CPX) testing. In: Cardiopulmonary exercise testing: Physiologic principles and clinical applications. Philadelphia: WB Saunders; 1986:151-167.
27. Pandey A, Patel M, Gao A, Willis BL, Das SR, Leonard D, et al. Changes in mid-life fitness predicts heart failure risk at a later age independent of interval development of cardiac and noncardiac risk factors: the Cooper Center Longitudinal Study. Am Heart J 2015;169(2):290-297.e1 [FREE Full text] [doi: 10.1016/j.ahj.2014.10.017] [Medline: 25641539]
28. Gosselink AT, Crijns HJ, van den Berg MP, van den Broek SA, Hillege H, Landsman ML, et al. Functional capacity before and after cardioversion of atrial fibrillation: a controlled study. Br Heart J 1994;72(2):161-166 [FREE Full text] [doi: 10.1136/hrt.72.2.161] [Medline: 7917690]
29. Sawada S, Tanaka H, Funakoshi M, Shindo M, Kono S, Ishiko T. Five year prospective study on blood pressure and maximal oxygen uptake. Clin Exp Pharmacol Physiol 1993;20(7-8):483-487. [doi: 10.1111/j.1440-1681.1993.tb01729.x] [Medline: 8403528]
30. Howlett RA, Gonzalez NC, Wagner HE, Fu Z, Britton SL, Koch LG, et al. Selected contribution: skeletal muscle capillarity and enzyme activity in rats selectively bred for running endurance. J Appl Physiol (1985) 2003;94(4):1682-1688 [FREE Full text] [doi: 10.1152/japplphysiol.00556.2002] [Medline: 12482771]
31. Thyfault JP, Rector RS, Uptergrove GM, Borengasser SJ, Morris EM, Wei Y, et al. Rats selectively bred for low aerobic capacity have reduced hepatic mitochondrial oxidative capacity and susceptibility to hepatic steatosis and injury. J Physiol 2009;587(Pt 8):1805-1816 [FREE Full text] [doi: 10.1113/jphysiol.2009.169060] [Medline: 19237421]
32. Earnest CP, Artero EG, Sui X, Lee DC, Church TS, Blair SN. Maximal estimated cardiorespiratory fitness, cardiometabolic risk factors, and metabolic syndrome in the aerobics center longitudinal study. Mayo Clin Proc 2013;88(3):259-270 [FREE Full text] [doi: 10.1016/j.mayocp.2012.11.006] [Medline: 23391253]
33. Kokkinos PF, Faselis C, Myers J, Panagiotakos D, Doumas M. Interactive effects of fitness and statin treatment on mortality risk in veterans with dyslipidaemia: a cohort study. Lancet 2013;381(9864):394-399. [doi: 10.1016/S0140-6736(12)61426-3] [Medline: 23199849]
34. Ångström L, Hörnberg K, Sundström B, Jonsson SW, Södergren A. Aerobic capacity is associated with disease activity and cardiovascular risk factors in early rheumatoid arthritis. Physiother Res Int 2020;25(3):e1833 [FREE Full text] [doi: 10.1002/pri.1833] [Medline: 31913553]
35. Lee CD, Blair SN. Cardiorespiratory fitness and stroke mortality in men. Med Sci Sports Exerc 2002;34(4):592-595 [FREE Full text] [doi: 10.1097/00005768-200204000-00005] [Medline: 11932565]
36. Sui X, Laditka JN, Church TS, Hardin JW, Chase N, Davis K, et al. Prospective study of cardiorespiratory fitness and depressive symptoms in women and men. J Psychiatr Res 2009;43(5):546-552 [FREE Full text] [doi: 10.1016/j.jpsychires.2008.08.002] [Medline: 18845305]
37. Defina LF, Willis BL, Radford NB, Gao A, Leonard D, Haskell WL, et al. The association between midlife cardiorespiratory fitness levels and later-life dementia: a cohort study. Ann Intern Med 2013;158(3):162-168 [FREE Full text] [doi: 10.7326/0003-4819-158-3-201302050-00005] [Medline: 23381040]
38. Snoek JA, Cramer MJM, Backx FJG. Cardiac rehabilitation: how much pain for the optimal gain? Neth Heart J 2013;21(3):135-137 [FREE Full text] [doi: 10.1007/s12471-013-0380-4] [Medline: 23408197]
39. Hammond-Haley M, Allen C, Han J, Patterson T, Marber M, Redwood S. Utility of wearable physical activity monitors in cardiovascular disease: a systematic review of 11 464 patients and recommendations for optimal use. Eur Heart J Digit Health 2021;2(2):231-243 [FREE Full text] [doi: 10.1093/ehjdh/ztab035] [Medline: 36712392]
40. Moran J, Wilson F, Guinan E, McCormick P, Hussey J, Moriarty J. Role of cardiopulmonary exercise testing as a risk-assessment method in patients undergoing intra-abdominal surgery: a systematic review. Br J Anaesth 2016;116(2):177-191 [FREE Full text] [doi: 10.1093/bja/aev454] [Medline: 26787788]
41. Stelken AM, Younis LT, Jennison SH, Miller DD, Miller LW, Shaw LJ, et al. Prognostic value of cardiopulmonary exercise testing using percent achieved of predicted peak oxygen uptake for patients with ischemic and dilated cardiomyopathy. J Am Coll Cardiol 1996;27(2):345-352 [FREE Full text] [doi: 10.1016/0735-1097(95)00464-5] [Medline: 8557904]
42. Mancini DM, Eisen H, Kussmaul W, Mull R, Edmunds LH, Wilson JR. Value of peak exercise oxygen consumption for optimal timing of cardiac transplantation in ambulatory patients with heart failure. Circulation 1991;83(3):778-786. [doi: 10.1161/01.cir.83.3.778] [Medline: 1999029]

43. Wang TJ, Levy D, Benjamin EJ, Vasan RS. The epidemiology of "asymptomatic" left ventricular systolic dysfunction: implications for screening. Ann Intern Med 2003;138(11):907-916. [doi: 10.7326/0003-4819-138-11-200306030-00012] [Medline: 12779301]
44. Inuzuka R, Diller GP, Borgia F, Benson L, Tay ELW, Alonso-Gonzalez R, et al. Comprehensive use of cardiopulmonary exercise testing identifies adults with congenital heart disease at increased mortality risk in the medium term. Circulation 2012;125(2):250-259 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.111.058719] [Medline: 22147905]
45. Cunningham JW, Nathan AS, Rhodes J, Shafer K, Landzberg MJ, Opotowsky AR. Decline in peak oxygen consumption over time predicts death or transplantation in adults with a Fontan circulation. Am Heart J 2017;189:184-192. [doi: 10.1016/j.ahj.2017.04.009] [Medline: 28625375]
46. Jensen MT, Treskes RW, Caiani EG, Casado-Arroyo R, Cowie MR, Dilaveris P, et al. ESC working group on e-cardiology position paper: use of commercially available wearable technology for heart rate and activity tracking in primary and secondary cardiovascular prevention-in collaboration with the European Heart Rhythm Association, European Association of Preventive Cardiology, Association of Cardiovascular Nursing and Allied Professionals, Patient Forum, and the Digital Health Committee. Eur Heart J Digit Health 2021;2(1):49-59 [FREE Full text] [doi: 10.1093/ehjdh/ztab011] [Medline: 36711174]
47. Bayoumy K, Gaber M, Elshafeey A, Mhaimeed O, Dineen EH, Marvel FA, et al. Smart wearable devices in cardiovascular care: where we are and how to move forward. Nat Rev Cardiol 2021;18(8):581-599 [FREE Full text] [doi: 10.1038/s41569-021-00522-7] [Medline: 33664502]
48. McElroy PA, Janicki JS, Weber KT. Cardiopulmonary exercise testing in congestive heart failure. Am J Cardiol 1988;62(2):35A-40A [FREE Full text] [doi: 10.1016/s0002-9149(88)80083-3] [Medline: 3133938]

Abbreviations

ACS: acute coronary syndrome
AF: atrial fibrillation
COPD: chronic obstructive pulmonary disease
CPET: cardiopulmonary exercise test
CRF: cardiorespiratory fitness
ECG: electrocardiogram
EF: ejection fraction
FDA: Food and Drug Administration
HF: heart failure
ICC: intraclass correlation coefficient
IHD: ischemic heart disease
LV: left ventricle
MET: metabolic equivalent
MRI: magnetic resonance imaging
SVT: supraventricular tachycardia
VCO_2: carbon dioxide production
VO_2: oxygen consumption

Edited by A Mavragani; submitted 07.01.23; peer-reviewed by Y Moayedi, X Gu; comments to author 27.03.23; revised version received 16.04.23; accepted 11.05.23; published 15.08.23.

Please cite as:
Bayshtok G, Tiosano S, Furer A
Use of Wearable Devices for Peak Oxygen Consumption Measurement in Clinical Cardiology: Case Report and Literature Review
Interact J Med Res 2023;12:e45504
URL: https://www.i-jmr.org/2023/1/e45504
doi:10.2196/45504
PMID:37581915

Case Report

Rapidly Deteriorating Degenerative Cervical Myelopathy Following Ventricular Shunt Revision for Hydrocephalus: Case Report

Tanzil Rujeedawa[1], BA; Oliver Mowforth[1], MA, MB BChir, MSt, MRCS; Mark Kotter[1], MD, MPhil, PhD; Benjamin Davies[1], BSc, MPhil, MRCS

Division of Neurosurgery, Department of Clinical Neurosciences, University of Cambridge, Cambridge, United Kingdom

Corresponding Author:
Benjamin Davies, BSc, MPhil, MRCS
Division of Neurosurgery
Department of Clinical Neurosciences
University of Cambridge
The Old Schools, Trinity Ln
Cambridge, CB2 1TN
United Kingdom
Phone: 1 44 01223 337733
Email: benjamin.davies4@nhs.net

Abstract

A female patient in her early 40s presented with a several-month history of gait unsteadiness and dragging her left leg. She had a background of congenital hydrocephalus, treated with a ventriculoatrial shunt. On examination, she had increased tone and brisk reflexes in the lower limbs and a positive Hoffmann sign. A computed tomography (CT) scan and shunt series x-rays identified hydrocephalus secondary to a disconnected shunt. Magnetic resonance imaging (MRI) of her cervical spine was also performed as part of the workup for her presenting symptoms and demonstrated features compatible with degenerative cervical myelopathy (DCM). The patient subsequently underwent a shunt revision. Following the operation, her walking and hand function deteriorated over a period of several weeks. She consequently underwent an anterior cervical decompression and fusion for DCM, which partially improved her symptoms. The sequence of events suggests that the shunt surgery may have precipitated a worsening of the DCM. Possible explanations include spinal cord injury related to neck extension or hypoperfusion during intubation and general anesthesia or the loss of cerebrospinal fluid cushioning following the reinstitution of effective cerebrospinal fluid shunting. Surgeons should be alert to this possibility and offer prompt surgical intervention for DCM if required.

(Interact J Med Res 2023;12:e48222) doi:10.2196/48222

KEYWORDS

cervical myelopathy; ossification of posterior longitudinal ligament; spondylosis; disk herniation; stenosis; spine; spinal; neck; disk; myelopathy; case; cervical; woman; women; ligament; gait

Introduction

Degenerative cervical myelopathy (DCM) is the umbrella term for a range of chronic spinal injuries caused by cervical stenosis due to degenerative or congenital pathology [1,2]. DCM presents with motor or sensory dysfunction in the upper or lower limbs, such as the loss of dexterity, paresthesia and imbalance, pain, and bladder and bowel dysfunction [3]. We report the case of a patient with worsening DCM following the treatment of coexisting hydrocephalus.

Case Report

A previously independent, self-employed female patient in her early 40s with congenital hydrocephalus presented with a dragging left leg and an abnormal gait for several months. There were no upper limb symptoms. On examination, there was increased tone and brisk reflexes in the lower limbs and a positive Hoffmann sign. Power and sensation were normal. She had an intracranial ventriculoatrial shunt, first inserted at 8 months old with 3 subsequent revisions. It was last revised when she was 15 years old. She is a married housekeeper with depression and anxiety, has never smoked, and does not drink alcohol.

Magnetic resonance imaging (MRI) of her cervical spine demonstrated cord compression at C3/4 and C5/6 with T2 signal hyperintensity within the spinal cord (Figure 1), confirming a diagnosis of DCM. However, the shunt series x-rays also demonstrated the disconnection of the distal shunt catheter at the level of the external auditory meatus (Figure 2 C). A computed tomography (CT) head scan demonstrated that the lateral and third ventricles were minimally enlarged compared to the most recent CT scan performed less than 1 year earlier and were dilated compared to a previous scan 13 years earlier (Figures 2 A and B). There was no periventricular interstitial oedema or sulcal effacement.

Based on the above investigations, diagnoses of hydrocephalus secondary to shunt dysfunction and DCM were made. The shunt revision was prioritized, taking place within the next 3 weeks. The ventriculoatrial shunt was replaced with a ventriculoperitoneal shunt in an uncomplicated procedure. Figure 3 illustrates the changes in blood pressure during the procedure. It is unclear whether precautions were taken with the cervical spine during intubation, such as fiberoptic intubation. A postoperative CT scan confirmed satisfactory ventricular catheter placement and a reduction in ventricular size. The patient was referred to spinal surgery for the management of her DCM following discharge.

While awaiting an outpatient appointment, the patient experienced a progressive deterioration in her mobility, becoming unable to stand or walk without assistance due to unsteadiness and struggling to use her hands. This led to 2 falls. The patient attended the emergency department. On examination, she had a spastic tetraparesis with urinary urgency and frequency, alongside C7 numbness. Her Modified Japanese Orthopaedic Association (mJOA) score was 9 (2+4+1+2). A CT head scan excluded shunt malfunction. She underwent anterior cervical decompression and fusion (C3/4 and C5/6), with good postoperative recovery. Postoperative imaging is demonstrated in Figure 4. At her 12-month follow up, her mJOA score had recovered to 12 (4+4+2+2). She was able to walk independently, fasten the buttons on her clothes, and peel vegetables but was unable to return to full-time employment. This was unchanged at 24 months. A summary of the timeline of events is illustrated in Figure 5.

Figure 1. Investigations supporting a diagnosis of degenerative cervical myelopathy: (A) sagittal T2 MRI; (B) C3/4 axial T2 MRI; and (C) C5/6 axial T2 MRI. The MRI was performed 1 month before shunt revision. MRI: magnetic resonance imaging.

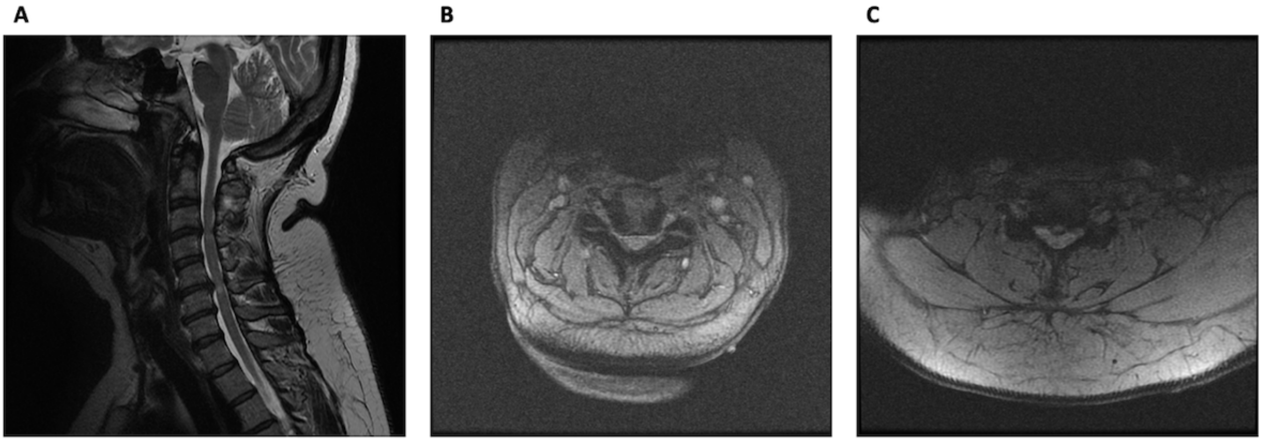


Figure 2. Investigations supporting hydrocephalus and shunt dysfunction: (A) CT scan from 13 years before the revision; (B) CT scan 1 month before the revision; and (C) shunt series x-ray 1 month before the revision. CT: computed tomography.

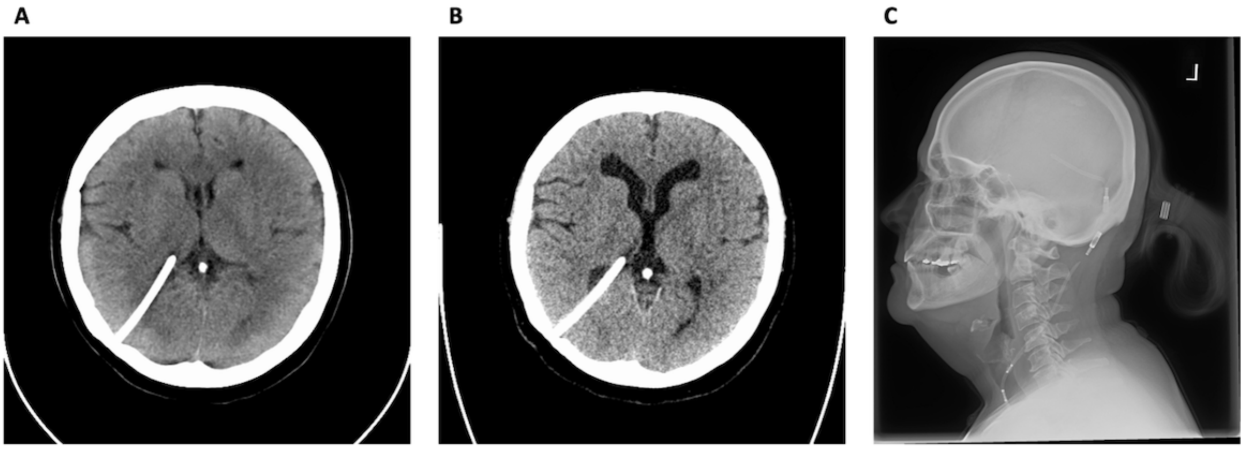


Figure 3. Graph demonstrating the changes in noninvasive blood pressure (NIBP), mean NIBP, and pulse over the duration of the procedure.

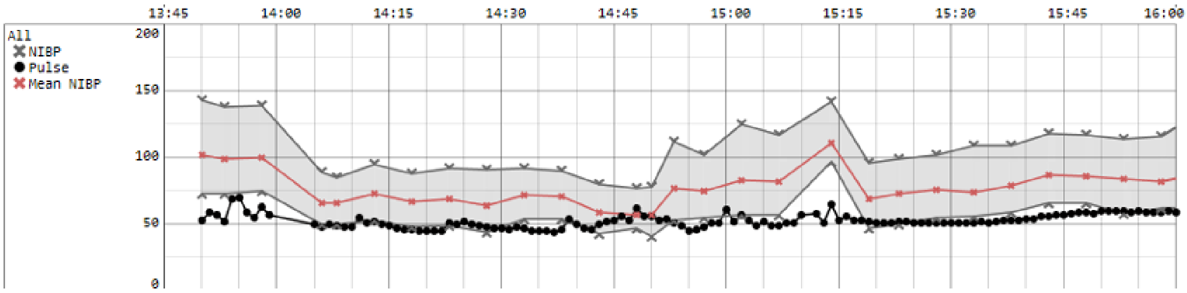


Figure 4. Postoperative x-ray of the cervical spine.

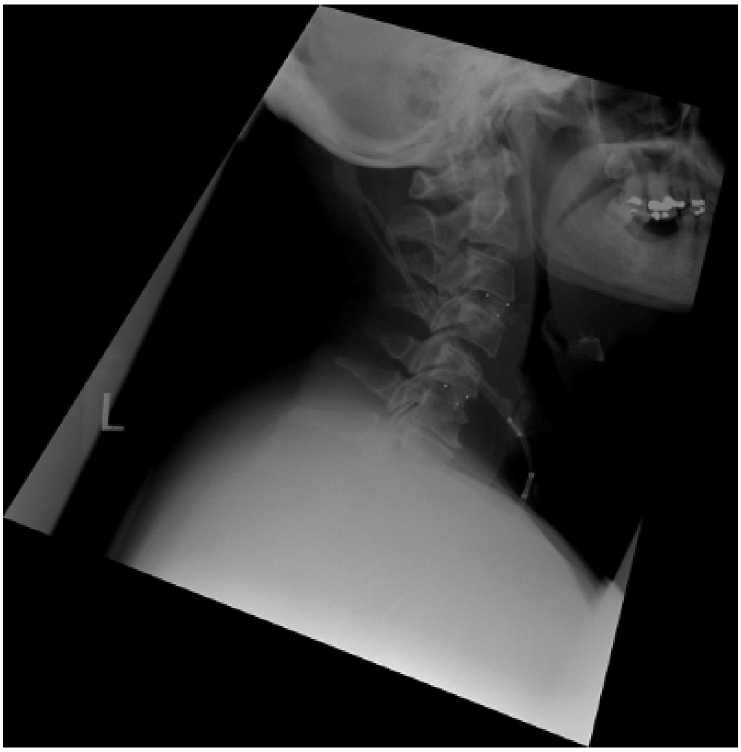
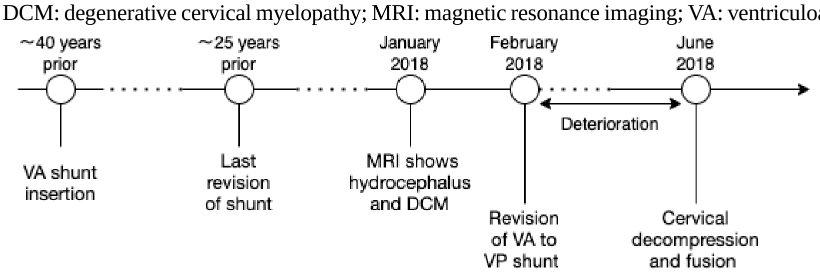


Figure 5. Timeline of events. DCM: degenerative cervical myelopathy; MRI: magnetic resonance imaging; VA: ventriculoatrial; VP: ventriculoperitoneal.



Ethical Considerations

Ethics review board assessment was not required. The patient was given a copy of the manuscript and has consented to publication.

Discussion

The case describes a patient with both hydrocephalus and DCM. Based on her clinical presentation, the management of the hydrocephalus was prioritized. This coincided with a rapid deterioration in the DCM. Given that DCM is a chronic and progressive condition, it is possible that this was coincidental. However, in early, mild stages of DCM, rapid deterioration is unusual without a trigger. This raises the question as to whether the treatment of hydrocephalus may have inadvertently precipitated a worsening of the DCM. However, it should be noted that other neurological pathology such as hydrocephalus may confound the use of mJOA score as an assessment of DCM severity.

We propose 2 potential mechanisms to explain how shunt surgery might have triggered a deterioration in DCM. The first mechanism is due to general anesthesia, either through neck manipulation for intubation or spinal cord hypoperfusion. Ordinarily, the neck is hyperextended to facilitate intubation. This increases the loading on the spinal cord via stretch but also reduces the spinal canal diameter. For an individual with a normal cervical spine, this has no consequence [4]. However, in patients with cervical stenosis, this could lead to further injury [4] and hypoperfusion [5-7]. Furthermore, anesthetic agents commonly precipitate reductions in blood pressure, which is often most profound during induction and intubation. Chronic hypoperfusion is considered a key feature of DCM, particularly at more advanced stages [8]. Clinical series have shown that DCM can be associated with hypertension that resolves following surgical treatment [9]. This is hypothesized to represent autoregulation [9,10]. Therefore, falls in systolic blood pressure could potentially contribute to the worsening of DCM secondary to hypoperfusion. Generally, anesthetic precautions such as intubation in the neutral position and arterial blood pressure control are taken in patients with DCM; it is unclear how much precaution was taken in this case.

An alternative hypothesis is that the elevated cerebrospinal pressure and volume was protective and that its diversion exacerbated the loading mechanism driving spinal cord injury.

At this stage, which theory or combination of theories explains the deterioration remains uncertain. The perioperative anesthetic management raises concerns but would not explain a body of (albeit low quality) evidence describing similar problems following cerebrospinal diversion, nor importantly a delayed and progressive deterioration of her DCM in the weeks to months after discharge following shunt surgery.

Surgical outcomes for DCM are strongly influenced by baseline disability. Put simply, the goal is to offer surgery when the benefits are known to outweigh the risks, but before there is irreversible damage. While the etiology remains uncertain, it is therefore impossible to suggest whether a different course of action should have been followed. The isolated lower limb presentation of imbalance could very reasonably be associated with the radiologically demonstrated shunt dysfunction. Even in the absence of an alternative diagnosis, for mild levels of impairment, DCM guidelines would suggest surveillance in the first instance. Therefore, in a patient with congenital hydrocephalus for whom shunt revisions had been required and with radiologically confirmed shunt dysfunction, a shunt revision would be the priority. The learning point at this stage is therefore to be aware of the potential risk of DCM deterioration, in order to intervene promptly if necessary.

Conclusion

This case serves as a reminder that the goal in DCM is to offer surgery when the benefits are known to outweigh the risks, but before there is irreversible damage. While the etiology remains uncertain, given that surgical outcomes for DCM are strongly influenced by baseline disability and symptom duration, surgeons should be alert to the possibility of other surgery, and in particular cerebrospinal diversion, being associated with worsening DCM.

Acknowledgments

BD is supported by a National Institute for Health Research (NIHR) Clinical Doctoral Fellowship. OM is supported by an Academic Clinical Fellowship at the University of Cambridge. The views expressed in this publication are those of the authors and not necessarily those of the National Health Service (NHS), NIHR, or Department of Health.

Authors' Contributions

TR, OM, and BD contributed to manuscript drafting and preparation. OM and BD contributed to conceptualization. MK contributed to manuscript review.

Conflicts of Interest

None declared.

References

1. Davies BM, Khan DZ, Barzangi K, Ali A, Mowforth OD, Nouri A, et al. We choose to call it 'degenerative cervical myelopathy': findings of AO Spine RECODE-DCM, an international and multi-stakeholder partnership to agree a standard unifying term and definition for a disease. Global Spine J 2022 Jun 29:21925682221111780 [FREE Full text] [doi: 10.1177/21925682221111780] [Medline: 35769029]
2. Davies BM, Mowforth OD, Smith EK, Kotter MR. Degenerative cervical myelopathy. BMJ 2018 Feb 22;360:k186 [FREE Full text] [doi: 10.1136/bmj.k186] [Medline: 29472200]
3. Davies BM, McHugh M, Elgheriani A, Kolias AG, Tetreault LA, Hutchinson PJA, et al. Reported outcome measures in degenerative cervical myelopathy: a systematic review. PLoS One 2016 Aug 2;11(8):e0157263 [FREE Full text] [doi: 10.1371/journal.pone.0157263] [Medline: 27482710]
4. Gadomski BC, Hindman BJ, Page MI, Dexter F, Puttlitz CM. Intubation biomechanics: clinical implications of computational modeling of intervertebral motion and spinal cord strain during tracheal intubation in an intact cervical spine. Anesthesiology 2021 Dec 01;135(6):1055-1065 [FREE Full text] [doi: 10.1097/ALN.0000000000004024] [Medline: 34731240]
5. Delamarter RB, Sherman J, Carr JB. Pathophysiology of spinal cord injury. recovery after immediate and delayed decompression. J Bone Joint Surg Am 1995 Jul;77(7):1042-1049 [FREE Full text] [doi: 10.2106/00004623-199507000-00010] [Medline: 7608226]

6. Durga P, Sahu BP. Neurological deterioration during intubation in cervical spine disorders. Indian J Anaesth 2014 Nov;58(6):684-692 [FREE Full text] [doi: 10.4103/0019-5049.147132] [Medline: 25624530]
7. Dinsmore J, Bacon RC, Hollway TE. The effect of increasing degrees of spinal flexion on cerebrospinal fluid pressure. Anaesthesia 1998 May;53(5):431-434 [FREE Full text] [doi: 10.1046/j.1365-2044.1998.00333.x] [Medline: 9659014]
8. Davies BM, Mowforth O, Gharooni AA, Tetreault L, Nouri A, Dhillon RS, et al. A new framework for investigating the biological basis of degenerative cervical myelopathy [AO Spine RECODE-DCM Research Priority Number 5]: mechanical stress, vulnerability and time. Global Spine J 2022 Feb;12(1_suppl):78S-96S [FREE Full text] [doi: 10.1177/21925682211057546] [Medline: 35174728]
9. Li P, Wei Z, Zhang H, Zhang K, Li J. Effects of decompressive operation on cardiac autonomic regulation in patients with cervical spondylotic myelopathy: analysis of blood pressure, heart rate, and heart rate variability. Eur Spine J 2019 Aug;28(8):1864-1871 [FREE Full text] [doi: 10.1007/s00586-019-05972-9] [Medline: 31011802]
10. Liu H, Wang H, Wu L, Wang S, Yang Z, Ma R, et al. Effects of decompressive cervical surgery on blood pressure in cervical spondylosis patients with hypertension: a time series cohort study. BMC Surg 2016 Jan 06;16:2 [FREE Full text] [doi: 10.1186/s12893-015-0117-y] [Medline: 26738624]

Abbreviations

CT: computed tomography
DCM: degenerative cervical myelopathy
mJOA: Modified Japanese Orthopaedic Association
MRI: magnetic resonance imaging

Edited by T de Azevedo Cardoso; submitted 16.04.23; peer-reviewed by A Nouri, BD Elder; comments to author 13.06.23; revised version received 06.07.23; accepted 24.07.23; published 28.08.23.

Please cite as:
Rujeedawa T, Mowforth O, Kotter M, Davies B
Rapidly Deteriorating Degenerative Cervical Myelopathy Following Ventricular Shunt Revision for Hydrocephalus: Case Report
Interact J Med Res 2023;12:e48222
URL: https://www.i-jmr.org/2023/1/e48222
doi:10.2196/48222
PMID:37639306

Case Report

Periorbital Necrotizing Fasciitis: Case Presentation

Ryan S Huang[1], MSc; Nikhil S Patil[2], MD; Yasser Khan[3], MD

[1]Temerty Faculty of Medicine, University of Toronto, Toronto, ON, Canada

[2]Michael DeGroote School of Medicine, McMaster University, Hamilton, ON, Canada

[3]Department of Ophthalmology, McMaster University, Hamilton, ON, Canada

Corresponding Author:
Ryan S Huang, MSc
Temerty Faculty of Medicine
University of Toronto
1 King's College Cir
Toronto, ON, M5S 1A8
Canada
Phone: 1 4038911818
Email: ry.huang@mail.utoronto.ca

Abstract

Necrotizing fasciitis (NF) is an aggressive and potentially life-threatening infection of the superficial fascia and surrounding skin, fat, fascia, muscle, and other soft tissue structures. Here, we outline the rare case of a 26-year-old man with a periorbital Streptococcus pyogenes A NF infection. Our case report underscores a unique instance of periorbital NF, distinctively presenting without any predisposing risk factors, shedding light on its presentation, treatment, and pathophysiology.

(Interact J Med Res 2023;12:e52507) doi:10.2196/52507

KEYWORDS

periorbital necrotizing fasciitis; Streptococcus pyogenes A; skin infection; soft tissue infection; dermatology infection; skin reaction; periorbital; necrotizing fasciitis; necrotizing; necrosis; case report; case reports; fasciitis; fatal; life-threatening; fascia; soft tissue; infection; pathology; pathophysiology; periorbital; eye; orbital; orbit; muscle; bacteria; bacterial; Streptococcus; inflammation; tissue; tissues

Introduction

Necrotizing fasciitis (NF) or necrotizing soft tissue infection is a rare and severe infection of the skin, muscle, subcutaneous tissue, and underlying fascia. In the United States, approximately 0.4 in every 100,000 people per year are affected by NF, while it is as common as 1 in every 100,000 people in other countries [1]. NF is characterized by rapidly progressing soft tissue necrosis invading the superficial and deep fascia, spreading along the fascial plane, and if not treated immediately, it eventually spreads into the systemic circulation [2]. The infection is most commonly associated with trauma involving a laceration that introduces bacteria into the wound site [3,4]. The infectious agents for NF are variable, but it is most commonly a result of a group A β-hemolytic streptococcal (*Streptococcus pyogenes*) infection [5]. NF is a rare condition, especially in urban settings, and most often occurs in the abdomen, perineum, arms, or legs. It is exceedingly rare in the head and neck regions; however, NF is known to cross anatomical planes [3,4]. Periorbital NF infections are generally localized to the eyelid and penetrate from the skin to the tarsus plate fascia. It can be identified with specific features such as necrosis with purulent discharge, blistering, and nonspecific features such as localized tenderness, severe pain, and erythema at the site of infection [5]. The spread of infection beyond the eyelid to the eyebrow, down the cheek, or to the nose is a major indicator of severe NF. If left untreated, periorbital NF complications can lead to blindness, neuralgias, and even death [5]. To the best of our knowledge, there are few cases documenting the successful treatment of periorbital NF in the literature. Here, we present the successful treatment of a unique case of a young patient with periorbital NF without any preexisting risk factors.

Case Presentation

A 26-year-old male patient was referred to us for left upper lid periorbital bruising after a bicycle fall onto asphalt 3 days prior. He had no previous medical history or risk factors for NF such as intravenous drug use, immunosuppression medications, advanced age, obesity, or malnutrition. The patient had left eye (OS) swelling with discharge, difficulty opening the eye, and moderate to severe pain. Preseptal purulent discharge and tissue emphysema were observed (Figure 1). Inflammatory changes extended to the left temporalis, masseter, and buccinator muscles, along with enlargement of the left parotid gland and intraparotid lymph nodes. An orbital computed tomography demonstrated extensive preseptal cellulitis, along with preorbital gas and fluid accumulation. The discharge was drained and cultured, revealing a *S pyogenes* A infection. The patient was treated with intravenous Tazocin and vancomycin immediately, leading to a decrease in inflammation over the following day (Figure 2). He was scheduled for emergency surgical debridement. During the procedure, the oculoplastics team found extensive necrosis of the left upper lid skin, orbicularis oculi, levator palpebrae superioris muscles, tarsus, surrounding fat, and neurovasculature. No necrosis was seen past the septum. After the debridement, the tissue was washed with Proviodine and treated with antibiotics. The tissue was then packed and dressed using Tobradex and bacitracin-soaked gauze. After 3 days of the surgery, the patient had no pain, erythema, ulcerations, or necrotic tissue. Granulation tissue had started to develop, but there was still a considerable amount of swelling restricting palpebral movement. At 1-week follow-up, the swelling had improved substantially, and the function of the levator palpebrae superioris and orbicularis oculi was regained.

Figure 1. Patient presentation. The patient had left eye (OS) swelling with difficulty opening the eye and pain. Preseptal purulent discharge and tissue emphysema were observed.

Figure 2. Posttreatment day 1. The patient improved the following day after treatment with intravenous Tazocin and vancomycin.

Discussion

Our case report presents the successful treatment of a rare case of periorbital NF in a patient with no risk factors, manifesting as extensive preseptal cellulitis, along with preorbital gas and fluid accumulation. Patients with an NF infection often have predisposing factors including diabetes mellitus, peripheral vascular disease, chronic alcoholism, malignancy, or immunosuppression [6]. Periorbital NF can be a result of a variety of pathogens including *Staphylococcus aureus*, *Cryptococcus neoformans*, *Pseudomonas aeruginosa*, and *Clostridium perfringens*. The most common pathogenic infection is *S pyogenes* A. NF can present in 1 of 2 types: type I or type II. Type I NF is a polymicrobial infection, with both aerobic and anaerobic bacteria, enterobacterium, and gram-negative bacilli. Type II NF is a result of a streptococcal infection and is the most prevalent causative agent of periorbital NF [6].

The pathogenesis of the disease is mediated by bacterial toxins, such as M protein and hyaluronic acid capsules from *S pyogenes*, causing the evasion of leukocyte phagocytosis and humoral immune surveillance [7]. Other toxins are responsible for toxin-induced intravascular platelet aggregation, preventing leukocytes from reaching the site of infection, which in turn creates a hypoxic state decreasing polymorphonuclear neutrophil function. While the bacterial toxins traverse the fascia, they induce platelet aggregation to thrombose vessels supplying the skin, allowing the toxin-induced necrosis to proliferate. As a result of the dermal hypoxic conditions, hemorrhagic bulla and a purple hue form. It is important to note that preseptal infections have an uncompromised septum. Therefore, only superficial structures and a moderate debridement are needed to remove the necrosis with most of the orbicularis intact, as shown in our case. However, in postseptal infections, the septum is compromised and therefore the infection can spread into the orbit and can result in proptosis. This can lead to postseptal disease where the infection can spread through the optic nerve to the brain and debridement may not be enough, requiring enucleation of the eye.

While our case report primarily focused on the successful treatment and manifestations of periorbital NF, the long-term prognosis and implications for patients deserve emphasis. In the absence of risk factors, it is paramount for clinicians to maintain a high index of suspicion for NF, especially given its rarity and potentially rapid progression. Patients without risk factors can present with more subtle and atypical symptoms, delaying diagnosis, and potentially worsening outcomes. The long-term implications for such patients are multifaceted. On a physiological level, even after successful treatment, these patients might be at risk for residual scarring, functional impairment, and potential cosmetic concerns. Moreover, due to the unexpected nature of the disease in these individuals, there might be psychological implications, including anxiety and concerns about recurrence. Regular follow-ups with these patients are crucial, not only to monitor and manage potential complications but also to provide emotional and psychological support. Early rehabilitation, counseling, and patient education can play a pivotal role in optimizing their quality-of-life after treatment. In conclusion, the success of our described treatment approach not only underscores its viability for similar cases but also sets a precedent for potential therapeutic strategies in managing periorbital NF. Future research focusing on novel biomarkers for NF may offer insights into prevention and early detection strategies, especially in patients without conventional risk factors.

Data Availability

The data supporting the findings of this study are available on request from the corresponding author, YK. The data are not publicly available to protect the privacy of research participants.

Conflicts of Interest

None declared.

References

1. Khalid M, Junejo S, Mir F. Invasive community acquired methicillin-resistant staphylococcal aureus (CA-MRSA) infections in children. J Coll Physicians Surg Pak 2018;28(9):S174-S177. [doi: 10.29271/jcpsp.2018.09.S174] [Medline: 30173689]
2. Shindo ML, Nalbone VP, Dougherty WR. Necrotizing fasciitis of the face. Laryngoscope 1997;107(8):1071-1079. [doi: 10.1097/00005537-199708000-00013] [Medline: 9261011]
3. Bisno AL, Stevens DL. Streptococcal infections of skin and soft tissues. N Engl J Med 1996;334(4):240-245. [doi: 10.1056/NEJM199601253340407] [Medline: 8532002]
4. Kronish JW, McLeish WM. Eyelid necrosis and periorbital necrotizing fasciitis. report of a case and review of the literature. Ophthalmology 1991;98(1):92-98 [FREE Full text] [doi: 10.1016/s0161-6420(91)32334-0] [Medline: 2023741]
5. Lazzeri D, Lazzeri S, Figus M, Tascini C, Bocci G, Colizzi L, et al. Periorbital necrotising fasciitis. Br J Ophthalmol 2010;94(12):1577-1585. [doi: 10.1136/bjo.2009.167486] [Medline: 19897473]
6. Tessier JM, Sanders J, Sartelli M, Ulrych J, De Simone B, Grabowski J, et al. Necrotizing soft tissue infections: a focused review of pathophysiology, diagnosis, operative management, antimicrobial therapy, and pediatrics. Surg Infect (Larchmt) 2020;21(2):81-93. [doi: 10.1089/sur.2019.219] [Medline: 31584343]
7. Georgiou A, Haque SA, Henderson H, Woollard A. Necrotizing fasciitis of the periorbital region: from presentation to reconstructive journey. Eur J Plast Surg 2021;44(4):531-536 [FREE Full text] [doi: 10.1007/s00238-020-01743-8] [Medline: 32952307]

Abbreviations

NF: necrotizing fasciitis
OS: left eye

Edited by G Eysenbach, T de Azevedo Cardoso; submitted 06.09.23; peer-reviewed by A Hasan, H Henderson, D Rusia, S Arya, V Prajap, K Wong Riff, M Rothschild; comments to author 25.10.23; revised version received 27.10.23; accepted 10.11.23; published 28.11.23.

Please cite as:
Huang RS, Patil NS, Khan Y
Periorbital Necrotizing Fasciitis: Case Presentation
Interact J Med Res 2023;12:e52507
URL: https://www.i-jmr.org/2023/1/e52507
doi:10.2196/52507
PMID:37971729

Research Letter

Three-Dimensional Virtual Reconstructions of Shoulder Movements Using Computed Tomography Images: Model Development

Yu-Hee Kim[1], PhD; In Park[2], MD, PhD; Soo Buem Cho[3], MD, PhD; Seoyon Yang[4], MD, PhD; Il Kim[5], BSc; Kyong-Ha Lee[6], PhD; Kwangnam Choi[6], PhD; Seung-Ho Han[7,8], MD, PhD

[1]Advanced Biomedical Research Institute, Ewha Womans University Seoul Hospital, Seoul, Republic of Korea
[2]Department of Orthopedic Surgery, Ewha Womans University College of Medicine, Seoul, Republic of Korea
[3]Department of Radiology, Ewha Womans University College of Medicine, Seoul, Republic of Korea
[4]Department of Rehabilitation Medicine, Ewha Womans University College of Medicine, Seoul, Republic of Korea
[5]SurgicalMind Inc, Gwangju, Republic of Korea
[6]Division of National S&T Data, Korea Institute of Science and Technology Information, Daejeon, Republic of Korea
[7]Department of Anatomy, Ewha Womans University College of Medicine, Seoul, Republic of Korea
[8]Ewha Medical Academy, Ewha Womans University Medical Center, Seoul, Republic of Korea

Corresponding Author:
Seung-Ho Han, MD, PhD
Department of Anatomy
Ewha Womans University College of Medicine
25 Magokdong-ro 2-gil
Gangseo-gu
Seoul, 07804
Republic of Korea
Phone: 82 2 6986 2601
Email: sanford@ewha.ac.kr

(Interact J Med Res 2023;12:e48381) doi:10.2196/48381

KEYWORDS

human digital twin; musculoskeletal twin; shoulder movement; visualization application; digital twin; musculoskeletal; visualization; movement; joint; shoulder; tomography; development; animation; animated; anatomy; anatomical; digital health; representation; simulation; virtual

Introduction

The demand for digital reconstruction of the human body, the "human digital twin," is increasing. Areas where human digital twins can be used include supporting disease diagnoses and predicting various treatment outcomes [1,2]. However, accurate digitalization of the structure and function of the human body is required, and a digital twin of the whole skeleton will provide a broader understanding of the human body. The Digital Korean Project, which was conducted from 2003 to 2007, established a database of Korean skeletons that included physical properties [3] and is being used in various research fields [4,5]. Digital representations of the human skeleton can be constructed at a higher level by including movement. Recently, Hernigou et al [6] proposed a method to include an ankle motion axis in digital twins. Although this method can be useful, one drawback of this study was that muscle movement was not included. This is because human skeletal movements cannot be evaluated except for the muscles. Therefore, in this study, we aimed to develop a visualization application of human shoulder movements involving bones and muscles.

Methods

Overview

Computed tomography (CT) images of the shoulders obtained at Ewha Womans University Seoul Hospital were used for 3D modeling. CT images in the DICOM file format were converted to an STL file format and then imported into the Blender program (Blender Institute). The size and volume of each bone and muscle involved in shoulder movement were measured and used for image visualization and 3D modeling. For the virtual reconstruction of shoulder movement, the Controller function was used to crudely model bone and muscle movement, and the Shapekey function was used to further adjust muscle contraction, relaxation, and rotation.

The Unity engine (Unity Technologies, San Francisco, CA) was used to develop a real-time 3D visualization application of shoulder movements. Reconstructed 3D resources and shoulder movement animations were exported from the Blender program in the FilmBoX file format and imported as Unity assets. A separate collision box was generated to avoid slowdown during the real-time selection of bones and muscles in polygonal units. Unity cameras were customized to display 3D structures and animations from different viewpoints. Interactive visibility controls for each bone and muscle were designated, and animation sequences were programmed for each motion.

Ethical Considerations

This study received ethics approval from Ewha Womans University Seoul Hospital (ethics approval No. 2023-02-037).

Results

To build a 3D model of the shoulders, 3 bones (scapula, clavicle, and humerus) were constructed to accommodate 6 shoulder movements (flexion, extension, abduction, adduction, internal rotation, and external rotation), and then, 9 muscles (coracobrachialis, deltoid, infraspinatus, latissimus dorsi, pectoralis majors, subscapularis, supraspinatus, teres major, and teres minor) that are primarily involved in each movement were built [7]. The reconstructed bones and muscles have been anatomically validated and are shown in Figure 1. These were then used to construct animations of muscle movements (Multimedia Appendix 1). The alignment and rotation angles of bones in the normal range of shoulder movement were implemented in the animation based on a kinesiology textbook [8].

Next, we developed a visualization application using the digital reconstruction of shoulder movements. The user interface of the application is shown in Figure 2A. The following functions were programmed in the application: present the name of the structure where the mouse is located; change the visibility of bones and muscles (Figure 2B); play/stop shoulder movement animation through the animation bar (Figure 2C); and move, rotate, or zoom in/out of the entire structure (Figure 2D). Using this application, we demonstrated a 3D real-time visualization of each structure and movement of interest.

Figure 1. Three-dimensional reconstruction of muscles involved in shoulder movement. The muscles involved in shoulder movement, which were manually segmented using computed tomography DICOM files, were reconstructed in 3D and placed on the bones.

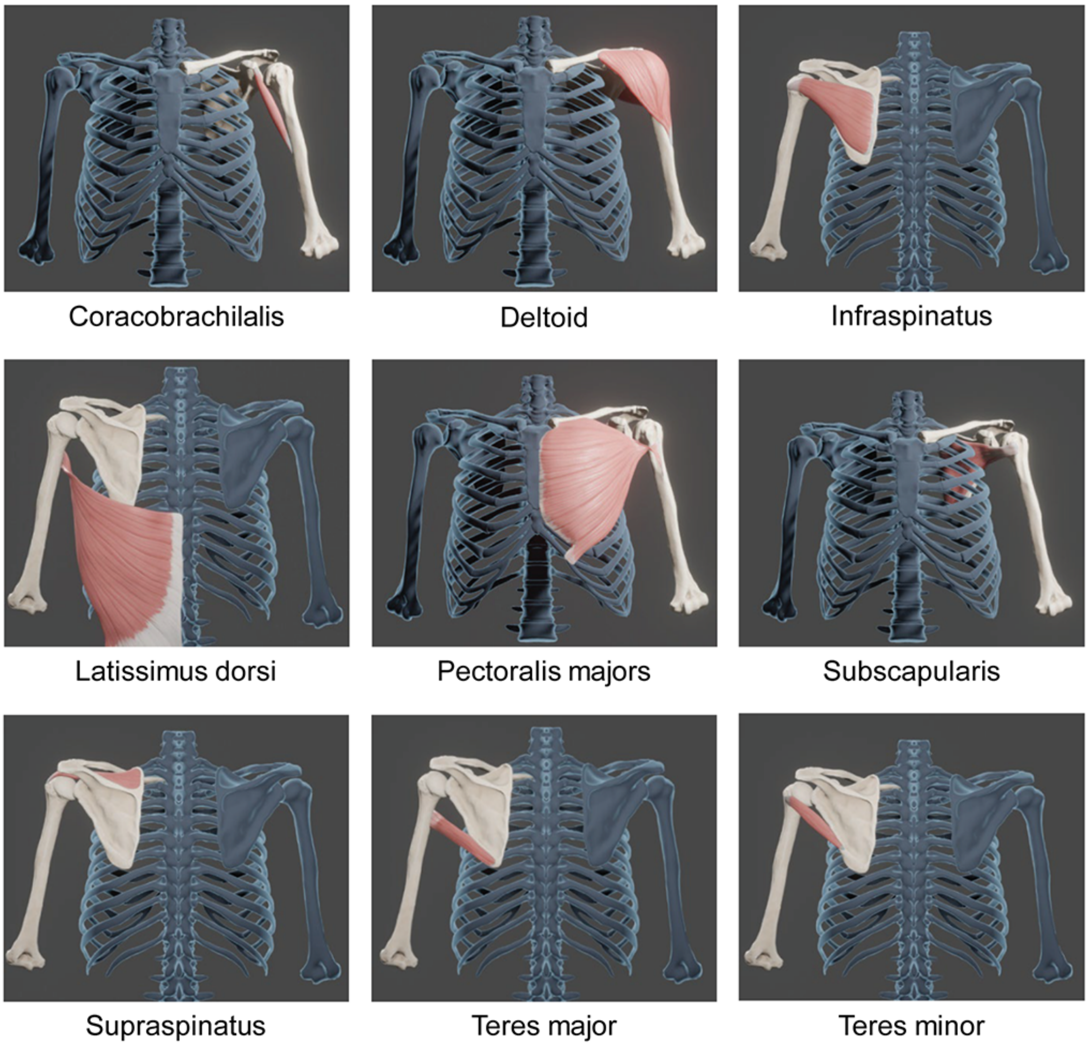
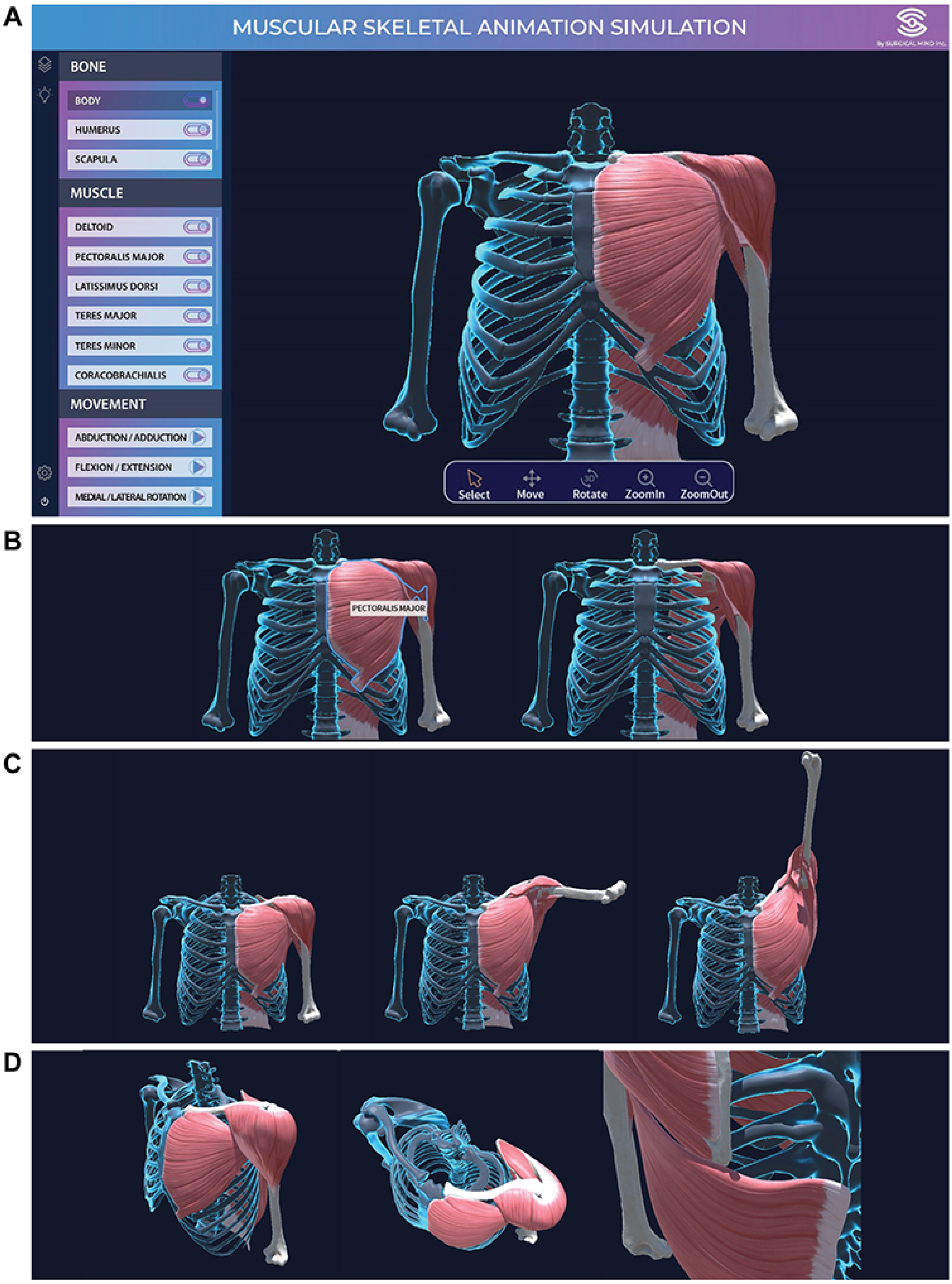


Figure 2. Application for visualization of shoulder movement. (A) A visualization application for real-time simulation of shoulder movement was constructed. Functions included in the application are shown. (B) Present the structure name and change the visibility; (C) play the movement animation; and (D) move, rotate, or zoom in/out of the structure.



Discussion

The ultimate goal of a human digital twin is the development of a comprehensive whole-body digital twin that includes the body structure and movement. This will enable the modeling of various diseases at the systemic level without limitation to a single organ or stationary postures. In this study, we used CT images to build a 3D digital representation of the shoulders comprising bones and muscles, and developed an application for visualizing shoulder movements. To the best of our knowledge, this is the first report implementing an application for the visualization of shoulder movements involving bones and agonistic muscles.

Future research will lead to the development of an algorithm to import patient-specific CT data and present it in our 3D visualization application. This includes an automated system that can identify normal and abnormal ranges of motion and can be used for the development of medical applications in arthroplasty or rehabilitation. Furthermore, it will be possible to develop an evolving digital twin that can continuously reflect and update changes that occur during the human life cycle.

Acknowledgments

The authors would like to thank Hyung-Seok Choi and Soo-Hyun Chae, medical illustrators at Ewha Medical Academy, for their work in creating illustration content. This research was funded by the Korea Institute of Science and Technology Information (grant K-22-L01-C03).

Authors' Contributions

KHL, KC, and SHH conceptualized the study. YHK, IP, SBC, SY, and IK curated the data. SHH acquired the funding. YHK and IK conducted the investigation. KHL, KC, IK, and SHH developed the methodology. YHK and SHH were the project administrators. IP and SBC were involved with the resources. SHH and KC supervised the study. IP and SY performed the validation. YHK wrote the original draft. YHK, IP, SBC, SY, and SHH reviewed and edited the paper.

Conflicts of Interest

IK is the founder and chief executive officer of SurgicalMind Inc. The remaining authors have no actual or potential conflicts of interest to declare.

Multimedia Appendix 1
Animation of muscle contraction, relaxation, and rotation.
[MP4 File (MP4 Video), 180657 KB - ijmr_v12i1e48381_app1.mp4]

References

1. Qin J, Wu J. Realizing the potential of computer-assisted surgery by embedding digital twin technology. JMIR Med Inform 2022 Nov 08;10(11):e35138 [FREE Full text] [doi: 10.2196/35138] [Medline: 36346669]
2. Fagherazzi G. Deep digital phenotyping and digital twins for precision health: time to dig deeper. J Med Internet Res 2020 Mar 03;22(3):e16770 [FREE Full text] [doi: 10.2196/16770] [Medline: 32130138]
3. Digital Korean. URL: http://dk.kisti.re.kr/ [accessed 2023-09-29]
4. Kwak DS, Surendran S, Pengatteeri YH, Park SE, Choi KN, Gopinathan P, et al. Morphometry of the proximal tibia to design the tibial component of total knee arthroplasty for the Korean population. Knee 2007 Aug;14(4):295-300. [doi: 10.1016/j.knee.2007.05.004] [Medline: 17600719]
5. Kim M, Kwak D, Park C, Park S, Oh S, Lee S, et al. Quantitative anatomy of the endplate of the middle and lower cervical vertebrae in Koreans. Spine (Phila Pa 1976) 2007 Jun 15;32(14):E376-E381. [doi: 10.1097/BRS.0b013e318067e384] [Medline: 17572609]
6. Hernigou P, Olejnik R, Safar A, Martinov S, Hernigou J, Ferre B. Digital twins, artificial intelligence, and machine learning technology to identify a real personalized motion axis of the tibiotalar joint for robotics in total ankle arthroplasty. Int Orthop 2021 Sep;45(9):2209-2217. [doi: 10.1007/s00264-021-05175-2] [Medline: 34351462]
7. Bindal VD. Textbook of Kinesiology. First edition. New Delhi, India: Jaypee; 2018.
8. Neumann DA. Kinesiology of the Musculoskeletal System: Foundations for Physical Rehabilitation. First edition. St. Louis, MO: Mosby; 2002.

Abbreviations

CT: computed tomography

Edited by T Leung, T de Azevedo Cardoso; submitted 21.04.23; peer-reviewed by J Wu, M Brown, V Girishan Prabhu; comments to author 16.07.23; revised version received 15.09.23; accepted 26.09.23; published 05.10.23.

Please cite as:
Kim YH, Park I, Cho SB, Yang S, Kim I, Lee KH, Choi K, Han SH
Three-Dimensional Virtual Reconstructions of Shoulder Movements Using Computed Tomography Images: Model Development
Interact J Med Res 2023;12:e48381
URL: https://www.i-jmr.org/2023/1/e48381
doi:10.2196/48381
PMID:37796554

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org