Interactive Journal of Medical Research

Impact Factor (2022): 2.0
Volume 13 (2024)  ISSN: 1929-073X

Contents

Reviews

Strategies to Alleviate the Burden Experienced by Informal Caregivers of Persons With Severe Mental Disorders in Low- and Middle-Income Countries: Scoping Review (e48587)
Olindah Silaule, Daleen Castelleijn, Fasloen Adams, Nokuthula Nkosi.

Digital Methods for the Spiritual and Mental Health of Generation Z: Scoping Review (e48929)
Susanna Park, Bridgette Do, Jaccelynn Yourell, Janice Hermer, Jennifer Huberty.

Original Papers

Improvements in Neoplasm Classification in the International Classification of Diseases, Eleventh Revision: Systematic Comparative Study With the Chinese Clinical Modification of the International Classification of Diseases, Tenth Revision (e52296)
Yicong Xu, Jingya Zhou, Hongxia Li, Dong Cai, Huanbing Zhu, Shengdong Pan.

Behavioral Insights from Vaccine Adoption in Nigeria: Cross-Sectional Survey Findings (e47817)
Sohail Agha, Ileanya Nsofor, Drew Bernard, Sarah Francis, Nandan Rao.

The Effect of Body Temperature Changes on the Course of Treatment in Patients With Pneumonia and Sepsis: Results of an Observational Study (e52590)

Evaluation of the Accuracy, Credibility, and Readability of Statin-Related Websites: Cross-Sectional Study (e42849)
Eunice Ling, Domenico de Pieri, Evenne Loh, Karen Scott, Stephen Li, Heather Medbury.

A Preliminary Checklist (METRICS) to Standardize the Design and Reporting of Studies on Generative Artificial Intelligence–Based Models in Health Care Education and Practice: Development Study Involving a Literature Review (e54704)
Malik Sallam, Muna Barakat, Mohammed Sallam.

Factors Associated With Worsened Mental Health of Health Care Workers in Canada During the COVID-19 Pandemic: Cross-Sectional Survey Study (e50064)
Ali AbdulHussein, Zahid Butt, Stanko Dimitrov, Brian Cozzarin.
A Web-Based Tool to Assess Social Inclusion and Support Care Planning in Mental Health Supported
Accommodation: Development and Preliminary Test Study (e45987)
Sharon Eager, Helen Killaspy, Joanna C, Gillian Mezey, Peter McPherson, Megan Downey, Georgina Thompson, Brynmor Lloyd-Evans. . . .

1 1 6
Strategies to Alleviate the Burden Experienced by Informal Caregivers of Persons With Severe Mental Disorders in Low- and Middle-Income Countries: Scoping Review

Olindah Silaule1, MSc; Daleen Casteleijn2, PhD; Fasloen Adams3, PhD; Nokuthula Gloria Nkosi4, PhD

1Department of Occupational Therapy, University of the Witwatersrand, Johannesburg, South Africa
2Department of Occupational Therapy, University of Pretoria, Pretoria, South Africa
3Division of Occupational Therapy, Stellenbosch University, Cape Town, South Africa
4Department of Nursing Education, University of the Witwatersrand, Johannesburg, South Africa

Corresponding Author:
Olindah Silaule, MSc
Department of Occupational Therapy
University of the Witwatersrand
7 York Road
Parktown
Johannesburg, 2193
South Africa
Phone: 27 0782035213
Email: osilaule@cartafrica.org

Abstract

Background: There is considerable evidence of the burden of care encountered by informal caregivers of persons with severe and enduring mental health conditions in low- and middle-income countries. Previous studies have highlighted the need to support these informal caregivers as key players in the care of these patients. To date, limited evidence exists on the extent and types of strategies for supporting these informal caregivers in low- and middle-income countries.

Objective: This scoping review aims to identify and describe the extent and type of evidence on the existing strategies for alleviating the burden of care among informal caregivers of persons with severe and enduring mental health conditions in low- and middle-income countries.

Methods: A systematic literature search was completed following the Joanna Briggs Institute methodology for scoping reviews. The participants, concept, and context framework was used to guide the search for literature sources across 5 databases: PubMed, MEDLINE, CINAHL, and PsycINFO for published literature and ProQuest for unpublished literature. This review included studies that reported on strategies for alleviating the burden of care among informal caregivers of persons with severe and enduring mental health conditions, with a focus on studies that evaluated or recommended caregiver interventions and support strategies in low- and middle-income countries. The search was limited to studies conducted between 2001 and 2021, and only papers written in English were considered for inclusion. Using the Covidence software (Veritas Health Innovation), 2 reviewers independently screened the papers, applied the inclusion and exclusion criteria, and met biweekly to discuss and resolve conflicts. The relevant studies and reported outcomes were summarized, organized, and analyzed descriptively using numeric summary analysis and deductive content analysis.

Results: Of the 18,342 studies identified, 44 (0.24%) met the inclusion criteria. The included studies were from 16 low- and middle-income countries in Asia, Africa, Europe, and South and North America. Most studies (21/44, 48%) were randomized controlled trials conducted in Asian countries. The identified strategies were grouped into 2 categories: implemented and recommended intervention strategies. Identified strategies included community-based interventions, psychoeducation interventions, support groups, cognitive behavioral therapy, spirituality-based interventions, and smartphone-based interventions. In addition, mindfulness and empowerment, collaborative interventions, standard care, financial and social support, counseling, occupation-based interventions, policy and legislature, and access to mental health care were identified. Psychoeducation and support group interventions were identified as common strategies for alleviating the burden of care among informal caregivers of persons with severe and enduring mental health conditions.
Conclusions: This review provides evidence on the types of implemented and recommended strategies for alleviating the burden of care among informal caregivers in low- and middle-income countries. Although psychoeducational interventions were the most preferred strategy for alleviating burden, their benefits were short-lived when compared with peer-led support groups.

International Registered Report Identifier (IRRID): RR2-10.2196/44268

KEYWORDS
severe mental disorders; informal caregivers; caregiver stress; caregiver support; low- and middle-income country; mobile phone

Introduction

Background Functional Implications of Severe and Enduring Mental Health Conditions

People with severe and enduring mental health conditions are unable to cope with the demands of everyday life and need care after being discharged from a health facility. Serious functional limitations frequently interfere with the ability of people with severe and enduring mental health conditions to perform essential roles such as being a worker, family member, or friend [1]. The lack of independence among people with severe and enduring mental health conditions in daily living activities and, in some cases, behavioral problems result in poor quality of life. Psychosis, bipolar mood, major depressive, anxiety, eating, and personality disorders are often classified as severe and enduring mental health conditions [2]. The amount of care needed varies greatly depending on the diagnosis, severity of symptoms, and level of independent functioning. As such, assistance is needed in areas of basic self-care, healthy eating, following daily routines, medication management and compliance, engagement in meaningful activities, and community integration. Therefore, informal caregivers are essential to fulfill the care needs of people with severe and enduring mental health conditions. This task becomes draining, and the caregiver burden should not be underestimated [3,4].

Informal Caregivers in Mental Health

Informal caregivers are people who deliver care without remuneration to persons with a chronic condition. The caregivers of persons with severe and enduring mental health conditions are most often family members or neighbors who assist with the care of the person with a mental health condition [5]. Dixon et al [6] noted that, often, the caregiver of a person with a severe and enduring mental health condition does not have a choice in being the caregiver, which is an additional burden of care for the caretaker. Informal caregivers’ roles include monitoring medication, being the contact person between the health provider and family, early identification of signs of relapse, taking care of daily tasks such as self-care, providing meals, and ensuring the safety of the person [7,8]. Informal caregivers provide emotional support when needed and deal with the challenging behavior of the person with a severe and enduring mental health condition, which may lead to police involvement [9]. Involving informal caregivers in the routine care and management of persons with severe and enduring mental health conditions has shown a positive influence on the course of the illness but only if the caregivers’ needs are addressed and they are supported in one way or another [5,6,10].

Caregiver Burden and Need for Support

The burden of care experienced by caregivers is classified into objective and subjective burden. Objective burden refers to the tangible impact of the demands of caring tasks and encompasses the practical and concrete aspects of caregiving that can be quantified or assessed externally [11,12]. Subjective burden refers to the emotional and psychological experiences, feelings, and perceptions of caregivers related to their caregiving role [8,13]. It focuses on how caregivers perceive the impact of their responsibilities on their well-being and mental health.

In low- and middle-income countries, objective burden can easily overshadow subjective burden, such as a lack of community mental health care services and clinics that are out of stock of medication for people with severe and enduring mental health conditions, causing relapse and often readmission [11]. Other concrete aspects of objective burden, such as long hours of being available to the care recipient, time lost from daily activities, and not being able to earn an income, have serious consequences for caregivers as they are unable to pursue their own goals in life and, accordingly, experience a lower quality of life [9,12]. Cultural beliefs about being cursed by ancestors [14] and stigma from health care professionals [15] further aggravate the caring duties of caregivers, precipitating objective burden. The objective burden of care is less reported in the literature than subjective feelings of burden. Many studies have reported the subjective burden of care, which highlights emotional distress and feelings of anxiety, depression, and sadness related to the challenges of caregiving and witnessing the struggles of their family member. Loss of freedom and autonomy has also been reported owing to the demand of caregiving [16]. Feelings of shame and social isolation stem from the stigma of mental illness and cultural beliefs about being cursed by ancestors. Caregivers can also experience role strain as they have to juggle responsibilities at home, which can cause fatigue and overall dissatisfaction in life [17]. Both dimensions of burden are crucial for understanding the challenges faced by family caregivers of individuals with severe and enduring mental health conditions and play a vital role in informing support strategies and interventions.

There is overwhelming evidence in the literature for the need to support caregivers, and several studies from high-income and low- and middle-income countries have indicated successful strategies for alleviating caregiver burden. Studies on informal caregivers of people with severe and enduring mental health conditions in low- and middle-income countries have also increased over the past 10 years, and the burden of caregivers has been well described [18-21]. Yerriah et al [16] reported on the extent of the burden of caregivers of persons with severe and enduring mental health conditions in low- and middle-income countries. Although psychoeducational interventions were the most preferred strategy for alleviating burden, their benefits were short-lived when compared with peer-led support groups.

KEYWORDS
severe mental disorders; informal caregivers; caregiver stress; caregiver support; low- and middle-income country; mobile phone

https://www.i-jmr.org/2024/1/e48587

Support Strategies for Caregivers

To date, a number of strategies to support caregivers have been investigated with varying results. In Turkey, family-to-family support programs have shown a positive impact on the burden of care [22]. A meta-analysis by Chen et al [23] showed some support for nonpharmacological interventions (mostly psychoeducation) for caregivers of persons with schizophrenia, but the authors reported a potential bias in the results because of the small sample size. In contrast, the systematic review and meta-analysis by Sin et al [24] did not support psychoeducation to improve compliance with treatment and prevent relapse in persons with psychosis, and they reported a lack of available data; thus, no meta-regressions could be conducted. Ewertzon and Hanson [18] conducted a narrative review and identified provision of knowledge, problem-solving, stress management, mutual support groups, and individual-support interventions as successful support interventions for caregivers. Finally, a systematic review by Napa et al [25] revealed that there is insufficient evidence of interventions for psychological distress and expressed emotions in families of persons who experienced first-episode psychosis.

There is limited evidence of web-based health care services and digital health technologies for supporting informal caregivers of individuals with severe and enduring mental health conditions [26]. Sin et al [27] developed an eHealth intervention called Carers for People with Psychosis e-support, but its effectiveness has not yet been investigated. Ploeg et al [28] conducted a rapid review of web-based interventions to improve general caregiver outcomes. More than half of the 17 included studies showed a positive outcome for decreased depressive symptoms, stress, and anxiety among caregivers. If virtual strategies can support caregivers in rural areas, they may be a feasible method to reach people in remote areas who have poor access to health care.

Informal caregivers in rural or remote areas face additional objective burdens such as poor access to services, lack of integration of mental health into community health services [29], and poor intersectoral collaboration. They often have to wait long hours before any support arrives, and in many cases, they have to deal with challenges with the limited resources they have available.

If the types of intervention strategies that could be relevant for low- and middle-income countries, how they were implemented, and the outcome that was achieved were mapped, it could guide health care workers to support informal caregivers on various levels and with various strategies. The availability of virtual support strategies could add another dimension of support to caregivers, which may lead to additional positive outcomes such as immediate support, available information, and contact with support groups. People with severe and enduring mental health conditions benefit as they are most likely to receive optimal care, and their relapse rate may decrease, which means less need for readmission in overextended mental health care wards or hospitals. In addition, people with severe and enduring mental health conditions may also experience a better quality of life if their carers are supported. Thus, it is essential to understand the strategies to alleviate informal caregiver burden and how these strategies should be implemented.

This scoping review aimed to map the strategies to alleviate the objective and subjective burden of informal caregivers of persons with severe and enduring mental health conditions in low- and middle-income countries. The objectives of this scoping review were to (1) identify the types of existing strategies (virtual and face-to-face) for alleviating the objective or subjective burden of care, (2) describe the characteristics of the identified strategies, and (3) list the positive outcomes that were achieved using the identified strategies.

Methods

Review Methodology

This scoping review followed the Joanna Briggs Institute methodology specific to scoping reviews [30]. An a priori protocol for this review has been published [31].

Review Question

The scoping review addressed two research questions related to the strategies for alleviating the burden of care among informal caregivers of persons with severe and enduring mental health conditions:

1. Which existing strategies are reported in the literature for alleviating the burden of informal caregivers of persons with severe and enduring mental health conditions in low- and middle-income countries?
2. What are the outcomes reported by the authors of the strategies for alleviating burden among informal caregivers?

Eligibility Criteria

Participants

Studies were included if participants were informal caregivers of persons with severe and enduring mental health conditions in low- and middle-income countries. Studies that focused on informal caregivers of patients diagnosed with Alzheimer disease and dementia were excluded. This included family, friends, neighbors, and community members who voluntarily provided care without any remuneration. Studies that included both informal caregivers and patients as participants were considered for inclusion, whereas those that reported only on patients were excluded.

Concept

Caregiver burden was defined as the physical, psychological, emotional, social, and financial stresses of providing care to a person with severe and enduring mental health conditions [32]. Severe and enduring mental health conditions in this scoping review included the schizophrenia spectrum and other psychotic disorders, bipolar and related disorders, mood or depressive disorders, and personality disorders. Studies that investigated strategies for alleviating caregiver burden were considered for inclusion. This also included studies that investigated the synonyms used to describe the burden of care among informal caregivers, such as caregiver strain, stress, and role fatigue.
Studies that reported on the burden of care without suggesting strategies for alleviating it were excluded.

**Context**

Studies from low- and middle-income countries, including countries in Africa, Asia, Latin America, and the Caribbean, were considered for inclusion. In addition, the inclusion criteria comprised studies from lower-middle-income countries using the World Bank classification of the economic status of a country at the time the study was conducted. The term developing country was included as it is similar to lower-middle-income country, and thus, countries in Africa, Asia, and Latin America (including the Caribbean) were eligible for inclusion.

**Types of Sources**

This review included published qualitative, quantitative, and mixed methods studies with all types of designs as well as unpublished studies, including dissertations and theses. Over the last decade, there has been a growing body of knowledge from studies outlining strategies for alleviating burden among informal caregivers in low- and middle-income countries. As a result, this review was only restricted to studies conducted between 2011 and 2021. Only papers written in English were included in this review.

**Search Strategy**

A 3-step search strategy process was used to identify relevant studies. The search strategy commenced with an initial limited search of PubMed, MEDLINE, CINAHL, and PsycINFO for published studies between 2011 and 2021. The keywords were as follows: “[informal caregiver/s OR caregiver/s] AND [caregiver burden OR caregiver stress] AND [support strategy/ies OR intervention/s] AND [severe mental disorder/s OR mental illness] AND [developing country/ies].”

The second search was refined with the assistance of the librarian at the University of the Witwatersrand, and additional terms were added: “[Carers or ‘Informal Carers’ or ‘Male Caregiver’ or Women Caregiver’ or ‘Family Caregivers’] AND [Caregiver* strain’ or ‘Caregiver *Exhaustion’ or [Caregiver* burnout’ or ‘Carer* burden’ or ‘Carer* stress’ or ‘Caretaker* role fatigue’ or ‘Burden of Caregiver*’ or ‘Caretaker* burden’ or ‘Caretaker* load’ or ‘Caregiver* Psychology’]

AND

“Strategy’ or ‘Strategies’ or ‘Intervention’ or ‘Procedure’ or ‘Programme*’ or ‘Management’ or ‘Protocol* Guidelines’ or ‘Guide’ or ‘Policy’ or ‘Policies’

AND

“Mental disorders’ or ‘Mental illness’ or ‘schizophrenia’ or ‘bipolar mood disorder’ or ‘Major depressive disorder’ or ‘Psychotic disorder’ or ‘Personality disorder’ or ‘Bipolar affective disorder’

AND

(developing OR (less* N1 developed) OR “under developed” OR underdeveloped OR “under served” OR underserved OR deprived OR poor* OR “middle income” OR (low* N1 income)) N1 (country* OR nation* OR population* OR world) OR ((developing OR (less* N1 developed) OR “under developed” OR underdeveloped OR “under served” OR underserved OR deprived OR poor* OR “middle income” OR (low* N1 income)) N1 (country* OR nation* OR population* OR world) OR ((developing OR (less* N1 developed) OR “under developed” OR underdeveloped

**Source of Evidence Selection**

Following the search, all identified citations were exported to the Mendeley reference manager and thereafter to a web-based software, Covidence (Veritas Health Innovation), for primary screening and data extraction from the selected articles. Duplicates were removed, and 2 reviewers (OS and DC) independently conducted the screening for titles, abstracts, and full texts using the inclusion and exclusion criteria. To enhance the reliability of the review results, the first 5 abstracts were screened, and the 2 reviewers compared the screening results and clarified conflicts. Following this, the criteria were revised to ensure a detailed description of the various types of burden. The remaining abstracts were screened, and comparisons were made after 10 screenings to manage conflicts. The 2 reviewers (OS and DC) proceeded to screen texts and extract data from the selected articles. Any disagreements regarding the inclusion of the studies were resolved through discussion between the 2 reviewers.

**Data Extraction**

The Covidence software has templates for data extraction. These templates were modified using the data extraction tables proposed by Peters et al [30]. To ensure the extraction of
relevant data to answer the review question, these templates were piloted by the reviewers before use. In total, 2 separate data extraction templates (Multimedia Appendix 1) were used to extract data from studies that implemented an intervention and from studies that recommended interventions and strategies for alleviating caregiver burden. Data related to the characteristics of the studies were extracted, including study title; study aims; citation details; population of interest; concept of interest; context of the study, including the country and type of setting; type of evidence sources; study approach and designs; and participant characteristics, such as age, gender, and diagnosis of the care recipients. In addition, information on the characteristics of the interventions was extracted. This included the intervention content (ie, type of intervention, intervention developer and deliverer, and type of burden targeted by the intervention). The intervention description included the duration of the intervention, number of sessions, and location for the intervention. The templates for extracting the data are available in Multimedia Appendix 1. The reviewers extracted the data independently. Web-based meetings were held biweekly to discuss and resolve any discrepancies in the extracted data. Regular comparisons were easy and quick to conduct and improved the consistency of the extracted data. After completing the data extraction, the 2 reviewers scanned the references of the included articles to ensure that no articles were missed. Included and excluded studies were reported in a PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) flow diagram (Figure 1).

**Figure 1.** PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.
Data Analysis and Presentation

To increase consistency in the data, this review followed a 3-step analysis process as proposed by Levac et al [33]. This included analyzing the data, reporting the results, and applying meaning to the results [33]. The first step was reviewing the extracted data and identifying the type of data to be analyzed, which were then grouped according to the objectives of the review. This was followed by identifying the type of analysis appropriate for the specific type of data and then analyzing the extracted data. A descriptive quantitative analysis was conducted in a Microsoft Excel (Microsoft Corp) spreadsheet to describe the characteristics of the studies. This included the overall number of studies, types of study design, years of publication, characteristics of the study populations, countries where the studies were conducted, and types of interventions. In addition, descriptive content analysis was conducted deductively using the NVivo software (Lumivero) to code the characteristics of the extracted data into overall categories [34]. Data were categorized into types of strategies—implemented and recommended intervention strategies—and the outcomes of the strategies aimed at alleviating the burden of informal caregivers as reported by the authors. To enhance the clarity of the emerging findings, the data were summarized and presented in graphs and tables.

Results

Study Inclusion

A total of 18,323 studies were identified from the databases using the keywords. In total, 19 additional studies were identified from the reference lists of the included studies, thereby bringing the total number of potentially relevant studies to 18,342. Subsequently, of the 18,342 studies, 256 (1.4%) duplicates were removed, thereby leaving 18,086 (98.6%) studies deemed relevant to the review based on the keywords. A total of 98.83% (17,875/18,086) of these studies were then excluded; this exclusion of numerous studies occurred because of their incongruence with the predefined inclusion criteria, such as instances in which the study did not pertain to a low- or middle-income country context or in which the primary emphasis lay on the patient rather than the caregiver. Therefore, 1.17% (211/18,086) of the studies were screened by title and abstract based on set inclusion and exclusion criteria. This led to the exclusion of 58.8% (124/211) of these studies from the review. A total of 41.2% (87/211) of the studies were then retrieved as full texts based on eligibility by screening their titles and abstracts. Of these 87 studies, 43 (49%) were excluded for various reasons: 31 (72%) were not from low- or middle-income countries; 3 (7%) did not specify the intervention strategy; 2 (5%) focused on the wrong patient population; and another 7 (16%) were excluded for being a clinical guideline, focusing on patients rather than caregivers, having misaligned outcomes and a study design within the exclusion criteria (eg, review articles), being out of the date range of this review, the full text being in a language other than English, and a lack of an abstract and full text. Finally, 44 studies were included in this scoping review. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram was used to present the selection of the studies (Figure 1). It took approximately 8 months, from September 2021 to June 2022, to carry out the study from conception to completion of the project.

Characteristics of the Eligible Studies

Summary of Studies and Evidence Type

The included studies were published between 2011 and 2021. Most studies (31/44, 70%) were published between 2016 and 2021, and 30% (13/44) were published between 2011 and 2016. The types of studies included peer-reviewed journal articles (40/44, 91%), research reports or theses (2/44, 5%), and opinion pieces (2/44, 5%). Of the 44 included studies, 29 (66%) reported on a specific implemented intervention strategy, and 15 (34%) outlined recommended intervention strategies aimed at alleviating the burden of informal caregivers of people with severe and enduring mental health conditions in low- and middle-income countries. The study designs included randomized controlled trials (21/44, 48%), quantitative studies (15/44, 34%), qualitative studies (3/44, 7%), mixed methods studies (3/44, 7%), and gray literature (2/44, 5%; Table 1).
Table 1. Characteristics of the included studies (N=44).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Studies, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Summary of studies</strong></td>
<td></td>
</tr>
<tr>
<td>Implemented strategy studies</td>
<td>29 (66)</td>
</tr>
<tr>
<td>Recommended strategy studies</td>
<td>15 (34)</td>
</tr>
<tr>
<td><strong>Type of evidence</strong></td>
<td></td>
</tr>
<tr>
<td>Journal articles</td>
<td>40 (91)</td>
</tr>
<tr>
<td>Research reports or theses</td>
<td>2 (5)</td>
</tr>
<tr>
<td>Opinion pieces</td>
<td>2 (5)</td>
</tr>
<tr>
<td><strong>Year of publication</strong></td>
<td></td>
</tr>
<tr>
<td>2011-2015</td>
<td>13 (30)</td>
</tr>
<tr>
<td>2016-2021</td>
<td>31 (70)</td>
</tr>
<tr>
<td><strong>Study design</strong></td>
<td></td>
</tr>
<tr>
<td>Randomized controlled trial</td>
<td>21 (48)</td>
</tr>
<tr>
<td>Quantitative design</td>
<td>15 (34)</td>
</tr>
<tr>
<td>Qualitative design</td>
<td>3 (7)</td>
</tr>
<tr>
<td>Mixed methods</td>
<td>3 (7)</td>
</tr>
<tr>
<td>Gray literature</td>
<td>2 (5)</td>
</tr>
</tbody>
</table>

**Context of the Studies**

A total of 44 studies identified from 16 low- and middle-income countries in Asia, Africa, Europe, and South and North America (Mexico) were included. The countries included China (12/44, 27%); Iran (10/44, 23%); India (5/44, 11%); Brazil (3/44, 7%); Greece (2/44, 5%); Ghana (2/44, 5%); and Thailand, Nepal, South Korea, Indonesia, Jordan, Mexico, Ethiopia, Turkey, Botswana, and mixed countries (including Brazil, Iran, Colombia and Mexico 1/44, 2% of the studies each). As shown in Table 2, majority of the studies with implemented strategies (24/29, 83%) and those with recommended strategies (9/15, 60%) were from Asian countries. Overall, 7% (2/29) of the studies each from Europe, South America, and North America implemented the strategies for alleviating burden and 13% (2/15) of the studies from South and North America recommended the strategies. Only 3% (1/29) of the studies conducted in Africa implemented the strategies and 20% (3/15) of the studies recommended the strategies. A study (1/15, 7%) from mixed countries recommended the strategies for alleviating burden.

Table 2. Continents of the recommended and implemented strategies (N=44).

<table>
<thead>
<tr>
<th>Continent</th>
<th>Implemented strategies, n (%)</th>
<th>Recommended strategies, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asia</td>
<td>24 (83)</td>
<td>9 (60)</td>
</tr>
<tr>
<td>Africa</td>
<td>1 (3)</td>
<td>3 (20)</td>
</tr>
<tr>
<td>Europe</td>
<td>2 (7)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>South and North America</td>
<td>2 (7)</td>
<td>2 (13)</td>
</tr>
<tr>
<td>Mixed countries</td>
<td>0 (0)</td>
<td>1 (7)</td>
</tr>
</tbody>
</table>

**Characteristics of the Study Populations**

The targeted populations in the included studies were mainly family caregivers (38/44, 86%), 7% (3/44) of the studies targeted informal carers, and some studies (3/44, 7%) did not specify their targeted population but rather identified them as caregivers. Most studies (33/44, 75%) included both male and female caregivers, and 25% (11/44) did not specify the sex of their sample. Similarly, most studies (24/44, 55%) included both male and female care recipients, only 2% (1/44) of the studies focused only on male care recipients, and 43% (19/44) did not specify the gender of the care recipients included in the studies. The diagnoses of the care recipients in the included studies were grouped into 6 categories: schizophrenia (17/44, 39%), mood disorders (4/44, 9%), multiple diagnoses including schizophrenia and mood disorders (5/44, 11%), mixed severe and enduring mental health conditions (5/44, 11%), other psychotic and chronic conditions (6/44, 14%), and unspecified mental disorders (7/44, 16%; Table 3).
Table 3. Characteristics of the study populations in the included studies (N = 44).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Studies, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Population of interest</strong></td>
<td></td>
</tr>
<tr>
<td>Family caregiver or caregivers</td>
<td>38 (86)</td>
</tr>
<tr>
<td>Informal carer or carers</td>
<td>3 (7)</td>
</tr>
<tr>
<td>Caregiver or caregivers</td>
<td>2 (5)</td>
</tr>
<tr>
<td>Carer or carers</td>
<td>1 (2)</td>
</tr>
<tr>
<td><strong>Concept of interest</strong></td>
<td></td>
</tr>
<tr>
<td>Caregiver burden</td>
<td>18 (41)</td>
</tr>
<tr>
<td>Family burden</td>
<td>16 (36)</td>
</tr>
<tr>
<td>Caregiver stress</td>
<td>9 (20)</td>
</tr>
<tr>
<td>Caregiver strain</td>
<td>1 (2)</td>
</tr>
<tr>
<td><strong>Sex of caregivers</strong></td>
<td></td>
</tr>
<tr>
<td>Mixed-gender groups (male and female)</td>
<td>33 (75)</td>
</tr>
<tr>
<td>Not specified</td>
<td>11 (25)</td>
</tr>
<tr>
<td><strong>Gender of care recipients</strong></td>
<td></td>
</tr>
<tr>
<td>Mixed-gender groups (male and female)</td>
<td>24 (55)</td>
</tr>
<tr>
<td>Male</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Not specified</td>
<td>19 (43)</td>
</tr>
<tr>
<td><strong>Diagnosis of care recipients</strong></td>
<td></td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>17 (39)</td>
</tr>
<tr>
<td>Mood disorders</td>
<td>4 (9)</td>
</tr>
<tr>
<td>Schizophrenia and mood disorders</td>
<td>5 (11)</td>
</tr>
<tr>
<td>Mixed diagnoses of severe mental disorder</td>
<td>5 (11)</td>
</tr>
<tr>
<td>Unspecified mental disorders</td>
<td>7 (16)</td>
</tr>
<tr>
<td>Other (drug dependency, psychotic disorders, and chronic conditions)</td>
<td>6 (14)</td>
</tr>
</tbody>
</table>

**Concept of Interest**

The concepts of interest reported in the studies included caregiver burden (18/44, 41%), family burden (16/44, 36%), caregiver stress (9/44, 20%), and caregiver strain (1/44, 2%; Table 3).

**Review Findings**

This section outlines the findings of the scoping review and includes the types of intervention strategies, characteristics of the intervention strategies, and author-reported outcomes from the included studies.

**Types of Intervention Strategies**

The various strategies for alleviating caregiver burden were grouped into 2 categories: implemented and recommended intervention strategies. The implemented intervention strategies emerged from studies that investigated the effects of a specific intervention to alleviate the burden of informal caregivers. The recommended interventions were identified from studies that investigated caregiver burden and outlined recommended intervention strategies to be used to alleviate the burden of informal caregivers of persons with severe and enduring mental health conditions. Most studies (29/44, 66%) reported on an implemented intervention strategy, and 34% (15/44) of the studies recommended strategies to alleviate caregiver burden.

Most of the studies and articles (33/44, 75%) originated in Asia, with 55% (24/44) implemented and 20% (9/44) recommended intervention strategies reported on, followed by Africa with 2% (1/44) implemented and 7% (3/44) recommended intervention strategies, South and North America with 5% (2/44) implemented and 5% (2/44) recommended intervention strategies, and Europe with 5% (2/44) implemented intervention strategies (Table 2). A study from mixed countries only reported recommended intervention strategies. The identified strategies were grouped into the following categories: community-based interventions, psychoeducation interventions, support groups, cognitive behavioral therapy (CBT), spirituality-based interventions, smartphone-based interventions, mindfulness and empowerment, collaborative interventions, standard care, financial and social support, counseling, occupation-based interventions, policy and legislature, and access to mental health care (Table 4 and Textbox 2). These categories are discussed in the following subsections.
Table 4. Description of the identified evaluated strategies for alleviating informal caregiver burden.

<table>
<thead>
<tr>
<th>Type of strategy</th>
<th>Evaluated interventions</th>
<th>Author-reported outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Community-based interventions</td>
<td>• CoMHIP&lt;sup&gt;a&lt;/sup&gt; • RESHAPE&lt;sup&gt;b&lt;/sup&gt;</td>
<td>• Positive outcomes (positive effects on family caregivers)</td>
</tr>
<tr>
<td></td>
<td>••</td>
<td></td>
</tr>
<tr>
<td>Psychoeducation interventions</td>
<td>• Psychoeducation program • Group psychoeducation • FLEP&lt;sup&gt;c&lt;/sup&gt; • Peer-assisted education</td>
<td>• Improved coping skills • Enhanced recovery • Improved QoL&lt;sup&gt;d&lt;/sup&gt; • Knowledge and skill acquisition • Reduced burden of care • Reduced anxiety and stress • Reduced psychological strain • Enhanced family functioning • Relapse prevention • Reduced prolonged admissions • Negative outcomes (no efficacy on burden of care, QoL, or self-esteem among BMD&lt;sup&gt;e&lt;/sup&gt; caregivers)</td>
</tr>
<tr>
<td></td>
<td>••</td>
<td></td>
</tr>
<tr>
<td>Support groups</td>
<td>• FPGP&lt;sup&gt;f&lt;/sup&gt; • Family-led mutual support program</td>
<td>• Knowledge and skill acquisition • Enhance help seeking • Alleviation of guilt • Improved family and patient functioning • Decreased demand on mental health services • Improved psychosocial health • Improved QoL</td>
</tr>
<tr>
<td></td>
<td>••</td>
<td></td>
</tr>
<tr>
<td>Cognitive behavioral therapy</td>
<td>• Emotional regulation training</td>
<td>• Reduced burden of care • Reduced anxiety and stress • Increased resilience</td>
</tr>
<tr>
<td></td>
<td>••</td>
<td></td>
</tr>
<tr>
<td>Spirituality-based intervention</td>
<td>• Spirituality-based program</td>
<td>• Reduced anxiety and stress</td>
</tr>
<tr>
<td></td>
<td>••</td>
<td></td>
</tr>
<tr>
<td>Guided self-help interventions</td>
<td>• Manual-guided PBS&lt;P&gt;&lt;sup&gt;g&lt;/sup&gt; • The Good Mood Guide</td>
<td>• Strengthened positive caregiving experience • Reduced negative caregiving experience • Increased access to support</td>
</tr>
<tr>
<td></td>
<td>••</td>
<td></td>
</tr>
<tr>
<td>Smartphone-based interventions</td>
<td>• MHapps&lt;sup&gt;h&lt;/sup&gt;</td>
<td>• None</td>
</tr>
<tr>
<td></td>
<td>••</td>
<td></td>
</tr>
<tr>
<td>Mindfulness and empowerment interventions</td>
<td>• IEP&lt;sup&gt;i&lt;/sup&gt; • MBSR&lt;sup&gt;j&lt;/sup&gt; program</td>
<td>• Caregiver empowerment • Improved QoL • Decreased depressive symptoms • Increased self-efficacy</td>
</tr>
<tr>
<td></td>
<td>••</td>
<td></td>
</tr>
<tr>
<td>Collaborative interventions</td>
<td>• Participatory care model</td>
<td>• Reduced burden of care • Increased resilience</td>
</tr>
<tr>
<td></td>
<td>••</td>
<td></td>
</tr>
<tr>
<td>Standard care</td>
<td>• Multimodal intervention (general medicine, psychiatry, psychology, family therapy, neuropsychological rehabilitation, and occupational therapy)</td>
<td>• Reduced burden of care • Increased social support</td>
</tr>
</tbody>
</table>

<sup>a</sup>CoMHIP: community mental health early intervention project.<br>
<sup>b</sup>RESHAPE: Reducing Stigma Among Healthcare Providers.<br>
<sup>c</sup>FLEP: Family Link Education Programme.<br>
<sup>d</sup>QoL: quality of life.<br>
<sup>e</sup>BMD: bipolar mood disorder.<br>
<sup>f</sup>FPGP: Family-led Peer Support Group Program.<br>
<sup>g</sup>PBS<P>: problem-solving–based self-learning program.<br>
<sup>h</sup>MHapps: mental health apps.<br>
<sup>i</sup>IEP: integrated empowerment program.<br>
<sup>j</sup>MBSR: mindfulness-based stress reduction.
**Textbox 2.** Description of the identified recommended strategies for alleviating informal caregiver burden.

<table>
<thead>
<tr>
<th>Community-based interventions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Strengthening existing primary health care system</td>
</tr>
<tr>
<td>• Establishment of day nursing or care</td>
</tr>
<tr>
<td>• Respite care</td>
</tr>
<tr>
<td>• Vocational training for patients</td>
</tr>
<tr>
<td>• Single-family caregiver organization</td>
</tr>
<tr>
<td>• Supported employment for patients</td>
</tr>
<tr>
<td>• Home visits</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Psychoeducation interventions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Group psychoeducation</td>
</tr>
<tr>
<td>• Family education</td>
</tr>
<tr>
<td>• Continuing education programs for effective practice</td>
</tr>
<tr>
<td>• Educational lectures</td>
</tr>
<tr>
<td>• Psychoeducation program</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Support groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Nongovernmental mental health–related support groups</td>
</tr>
<tr>
<td>• Family caregiver assistance programs</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cognitive behavioral therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Family therapy groups</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Spirituality-based intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Turning to traditional healers and spiritual leaders</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Smartphone-based interventions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Mental health apps</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mindfulness and empowerment interventions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Stress management training</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Collaborative interventions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Caregiver involvement in all program elements</td>
</tr>
<tr>
<td>• Pluralistic and ecological approach to service delivery</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Standard care</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Telepsychiatry</td>
</tr>
<tr>
<td>• Availability of emergency teams</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Financial support</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Disability grant</td>
</tr>
<tr>
<td>• Fee-free mental health services</td>
</tr>
<tr>
<td>• Medical insurance and free medication</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Social support</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Medical care social security</td>
</tr>
<tr>
<td>• Social resources</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Counseling</th>
</tr>
</thead>
</table>

https://www.i-jmr.org/2024/1/e48587

(page number not for citation purposes)
• Basic counseling
• Supportive psychotherapy

**Occupation-based interventions**
• Physical and leisure activities

**Policy and legislature**
• National health insurance schemes
• Advocacy for a strong mental health policy
• Integration of caregiver actions and interventions into national mental health care plans
• Fee-free mental health services

**Access to mental health care**
• Caregiver care inclusion in daily treatment facilities
• Access to therapeutic tools
• Availability of mental health units or departments
• Periodic health checks for caregivers (every 6 months on average)
• Improvement in institutional mental health care

**Psychoeducational Interventions**
The first strategy was psychoeducational interventions, which was identified as the most common strategy implemented and recommended to alleviate caregiver burden. A total of 27% (12/44) of the studies implemented 4 different psychoeducational interventions, including 58% (7/12) of these studies using a psychoeducation program [35-40], 17% (2/12) using group psychoeducation [41,42], 17% (2/12) using the Family Link Education Programme [43,44], and 8% (1/12) of the studies implementing peer-assisted education [45] (Table 4). Similarly to the implemented intervention strategies, the recommended psychoeducational interventions identified included psychoeducational programs, group psychoeducation, family psychoeducation, continuing education programs for effective practice, and educational lectures [7,46] (Textbox 2).

**Support Group Interventions**
The second strategy was support groups. In total, 7% (3/44) of the studies implemented support group strategies, including the Family-led Peer Support Group Program, family-led mutual support programs, and nongovernmental mental health–related support groups [47-49] (Table 4). The recommended support group strategies identified in this review included nongovernmental mental health–related support groups and family caregiver support programs [50-54] (Textbox 2).

**Community-Based Interventions**
Community-based interventions were identified as a third strategy to alleviate caregiver burden, and 5% (2/44) of the included studies reported on implementing a community mental health early intervention project and Reducing Stigma Among Healthcare Providers [55,56] as community-based interventions (Table 4). The recommended community-based interventions included strengthening existing primary health care systems, establishing day nursing or care for the care recipients, respite care, vocational rehabilitation training for patients, single-family caregiver organizations, supported employment for patients, and home visits [46,52,53,57] (Textbox 2).

**Guided Self-Help Interventions**
A total of 5% (2/44) of the studies implemented guided self-help interventions, namely, a manual-guided problem-solving–based self-learning program [58] and the Good Mood Guide [59] (Table 4). No recommended guided self-help interventions were identified in this review.

**Mindfulness and Empowerment Interventions**
Mindfulness and empowerment interventions were implemented in 5% (2/44) of the studies, and the programs implemented were a mindfulness-based stress reduction program and an integrated empowerment program [60,61] (Table 4). Stress management training was identified as a recommended mindfulness and empowerment intervention strategy in this review [7] (Textbox 2).

**CBT Interventions**
Emotional regulation training was the only implemented CBT strategy identified in 5% (2/44) of the included studies [62,63] (Table 4). The recommended cognitive behavioral strategy identified in this review was family therapy groups [64] (Textbox 2).

**Spirituality-Based Interventions**
In total, 2% (1/44) of the studies implemented a spirituality-based program for informal caregivers in which the comparison intervention was 2 standard group training sessions related to general mental disorders [65] (Table 4). Turning to traditional healers and spiritual leaders was identified as a recommended spirituality-based intervention for alleviating informal caregiver burden [52] (Textbox 2).
**Participatory Care Model**

In total, 2% (1/44) of the studies implemented a participatory care model as a collaborative intervention strategy [66] (Table 4). The recommended collaborative interventions were caregiver involvement in all program elements and a pluralistic and ecological approach to service delivery [56,67] (Textbox 2).

**Standard Care**

The implemented standard care interventions were multimodal interventions that encompassed general medicine, psychiatry, psychology, family therapy, neuropsychological rehabilitation, and occupational therapy with no comparison intervention [68] (Table 4). The recommended standard care interventions were telepsychiatry and ensuring the availability of emergency teams [53] (Textbox 2).

**Smartphone-Based Interventions**

Only 2% (1/44) of the studies implemented smartphone-based interventions using mental health apps to alleviate the burden of informal caregivers with no comparison intervention [67] (Table 4). Although the study indicated the use of the mental health apps with care recipients and their caregivers, the specific intervention offered was not described. No recommended smartphone-based interventions were identified in this review.

**Additional Strategies Recommended for Alleviating Caregiver Burden**

The additional recommended intervention strategies were financial support, including the provision of a disability grant, fee-free mental health services, medical insurance, and free medication for care recipients [7,57,68]. The recommended social support comprised medical care social security and social resources [52,54,56,57,69]. Recommended counseling interventions comprised basic counseling and supportive psychotherapy [64,69]. Occupation-based interventions included participation in physical and leisure activities [64]. Policy and legislature recommended interventions comprised the implementation of national health insurance schemes, advocacy for a strong mental health policy, integration of caregiver actions and interventions into national mental health care plans, and fee-free mental health services [53,62,68]. Finally, other recommended interventions were access to mental health care, which considers caregiver care inclusion in daily treatment facilities; access to therapeutic tools; availability of mental health units or departments; periodic health checks for caregivers (every 6 months on average); and improved institutional mental health care [35,52,54,67] (Textbox 2).

**Characteristics of the Implemented Intervention Strategies**

Most of the implemented intervention strategies (18/29, 62%) were targeted at alleviating both objective and subjective burden, and 38% (11/29) were exclusively aimed at alleviating subjective burden. The intervention developers were reported as being researchers based on evidence (13/29, 45%) and researchers with expert input (5/29, 17%). In total, 28% (8/29) were existing interventions implemented without any adaptations, and 10% (3/29) of the interventions did not specify the intervention developer. The intervention deliverers identified in the studies included trained peer facilitators (7/29, 24%), researchers (7/29, 24%), psychiatrists and nurses (5/29, 17%), multiple health care professionals (4/29, 14%), and psychologists (3/29, 10%), and 10% (3/29) of the interventions did not specify who delivered them.

The number of sessions offered in the implemented intervention strategies was between 6 and 8 sessions (13/29, 45%), ≤5 sessions (8/29, 28%), and ≥11 sessions (5/29, 17%), and 10% (3/44) of the interventions did not specify the number of sessions. The duration of the implemented interventions in the studies was 6 to 15 weeks (10/29, 34%), ≤5 weeks (7/29, 24%), and ≥16 weeks (4/29, 14%), and 28% (8/29) of the interventions did not specify the duration. Most interventions (9/29, 31%) were delivered once per week, some were offered twice a week (8/29, 28%), and some of the interventions (12/29, 41%) did not specify the frequency of their sessions. The length of the sessions in the implemented interventions ranged between 2 and 2.5 hours (8/29, 28%), 1 to 1.5 hours (6/29, 21%), and ≤1 hour (1/29, 3%), and 48% (14/29) of the interventions did not specify the length of their sessions (Table 5).
Table 5. Characteristics of the implemented strategies for alleviating informal caregiver burden (N=29).

<table>
<thead>
<tr>
<th>Study</th>
<th>Year of publication</th>
<th>Type of intervention</th>
<th>Type of burden targeted by the intervention</th>
<th>Intervention developer</th>
<th>Intervention duration</th>
<th>Number of sessions</th>
<th>Duration of intervention</th>
<th>Frequency of sessions per week</th>
<th>Length of the sessions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ng et al [56]</td>
<td>2018</td>
<td>CoMHIP&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Subjective and objective</td>
<td>Not specified</td>
<td>Not specified</td>
<td>Not specified</td>
<td>Not specified</td>
<td>Not specified</td>
<td>Not specified</td>
</tr>
<tr>
<td>Rai et al [55]</td>
<td>2018</td>
<td>RESHAPE&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Subjective and objective burden</td>
<td>Developed as part of WHO's mhGAP&lt;sup&gt;d&lt;/sup&gt; and PRIME&lt;sup&gt;e&lt;/sup&gt;</td>
<td>TPO&lt;sup&gt;f&lt;/sup&gt; Nepal, a Nepali non-governmental organization</td>
<td>5 sessions</td>
<td>4 days</td>
<td>Not specified</td>
<td>Not specified</td>
</tr>
<tr>
<td>Rajai et al [45]</td>
<td>2021</td>
<td>Peer-assisted education</td>
<td>Subjective and objective burden</td>
<td>Researchers and approved by 3 faculty members of the Army College of Medical Sciences</td>
<td>Trained peer facilitators</td>
<td>6 sessions</td>
<td>3 weeks</td>
<td>Twice weekly</td>
<td>1 hour</td>
</tr>
<tr>
<td>Zhou et al [44]</td>
<td>2020</td>
<td>FLEP&lt;sup&gt;g&lt;/sup&gt;</td>
<td>Subjective and objective burden</td>
<td>FLEP: peer-led psychoeducation program developed from the stress and coping model by Pearlin et al [70]</td>
<td>Group facilitators who were peer specialists of experienced caregivers</td>
<td>8 sessions</td>
<td>8 weeks</td>
<td>Once weekly</td>
<td>2 hours</td>
</tr>
<tr>
<td>Dewi et al [37]</td>
<td>2019</td>
<td>Family psychoeducation and care decision without pasung</td>
<td>Subjective and objective burden</td>
<td>Researchers</td>
<td>Researchers</td>
<td>3 sessions</td>
<td>3 weeks</td>
<td>Not specified</td>
<td>35-40 minutes</td>
</tr>
<tr>
<td>Tabeleão et al [36]</td>
<td>2018</td>
<td>Psychoeducation</td>
<td>Subjective and objective burden</td>
<td>Not specified</td>
<td>10 psychologists</td>
<td>6 sessions</td>
<td>Not specified</td>
<td>Not specified</td>
<td>Not specified</td>
</tr>
<tr>
<td>Ntsayagae [40]</td>
<td>2017</td>
<td>Psychoeducation</td>
<td>Subjective and objective burden</td>
<td>Researcher based on existing literature</td>
<td>Researcher</td>
<td>2 sessions</td>
<td>Not specified</td>
<td>Not specified</td>
<td>Not specified</td>
</tr>
<tr>
<td>De Souza et al [39]</td>
<td>2016</td>
<td>Psychoeducation</td>
<td>Subjective and objective burden</td>
<td>Psychoeducational intervention for patients with BD&lt;sup&gt;h&lt;/sup&gt; by Colom et al [71] translated and adapted by Dell-Aglio et al [72]</td>
<td>Psychiatrist responsible for specific case</td>
<td>6 sessions</td>
<td>6 to 8 weeks</td>
<td>Twice weekly</td>
<td>Not specified</td>
</tr>
<tr>
<td>Kolostoumpis et al [38]</td>
<td>2015</td>
<td>Psychoeducation</td>
<td>Subjective burden</td>
<td>Adapted from the treatment protocol developed by Reinares et al [73] in the Barcelona Bipolar Disorders Program in Spain</td>
<td>Psychiatrist and psychologist</td>
<td>7 sessions</td>
<td>Not specified</td>
<td>Not specified</td>
<td>Not specified</td>
</tr>
<tr>
<td>Study</td>
<td>Year of publication</td>
<td>Type of intervention</td>
<td>Type of burden targeted by the intervention</td>
<td>Intervention deliverer</td>
<td>Intervention developer</td>
<td>Number of sessions</td>
<td>Duration of intervention</td>
<td>Frequency of sessions per week</td>
<td>Length of the sessions</td>
</tr>
<tr>
<td>-------</td>
<td>---------------------</td>
<td>----------------------</td>
<td>---------------------------------------------</td>
<td>-----------------------</td>
<td>-----------------------</td>
<td>--------------------</td>
<td>------------------------</td>
<td>-------------------------------</td>
<td>----------------------</td>
</tr>
<tr>
<td>Fallahi Khoshknab et al [41]</td>
<td>2014</td>
<td>Group psychoeducation</td>
<td>Subjective and objective burden</td>
<td>Researchers based on educational program of psychiatric nursing and psychiatric textbooks</td>
<td>Organized based on educational program of psychiatric nursing and psychiatric textbooks (Campbell [74]) and converted to understandable text for patients and families</td>
<td>4 sessions</td>
<td>4 weeks</td>
<td>Once weekly</td>
<td>2 hours</td>
</tr>
<tr>
<td>Navidian et al [42]</td>
<td>2012</td>
<td>Psychoeducation</td>
<td>Subjective burden</td>
<td>Researchers based on the families' needs and the existing literature</td>
<td>Psychiatrist and mental health nurse</td>
<td>4 sessions</td>
<td>4 weeks</td>
<td>Once weekly</td>
<td>2 hours</td>
</tr>
<tr>
<td>Sharif et al [75]</td>
<td>2012</td>
<td>Psychoeducation</td>
<td>Subjective and objective burden</td>
<td>Psychiatrist and psychiatric nurse based on the literature and needs of the families</td>
<td>Psychiatrist, nurse, and guest speakers</td>
<td>10 sessions</td>
<td>5 weeks</td>
<td>Not specified</td>
<td>1.5 hours</td>
</tr>
<tr>
<td>Chiu et al [43]</td>
<td>2011</td>
<td>FLEP</td>
<td>Subjective burden</td>
<td>Task force consisting of a mental health social worker, a recovered patient with editorial experience, and a caregiver based on available related local educational materials and the NAMI Family-to-Family program</td>
<td>Trainers who were themselves family members of people with SMI. They received training and a trainer manual.</td>
<td>8 sessions</td>
<td>Not specified</td>
<td>Once weekly</td>
<td>Not specified</td>
</tr>
<tr>
<td>Chien et al [48]</td>
<td>2018</td>
<td>Family-led support program</td>
<td>Subjective and objective burden</td>
<td>Family-led mutual support group—contents were based on similar program protocols and the researcher-developed family mutual support groups for psychotic disorders</td>
<td>Family-led mutual support group—co-led by 2 peer family caregivers along with a researcher and rehabilitation nurse</td>
<td>16 sessions</td>
<td>36 weeks</td>
<td>Twice weekly</td>
<td>2 hours</td>
</tr>
<tr>
<td>Chien et al [48]</td>
<td>2018</td>
<td>Family-led mutual support program</td>
<td>Subjective burden</td>
<td>6 experts on psychiatric rehabilitation (including psychiatrists, clinical psychologists, and nurse specialists)</td>
<td>Peer leader who received training from the researchers and worked closely with a group leader who was a trained psychiatric nurse</td>
<td>14 sessions</td>
<td>39 weeks</td>
<td>Twice weekly</td>
<td>2 hours</td>
</tr>
<tr>
<td>Study</td>
<td>Year of publication</td>
<td>Type of intervention</td>
<td>Type of burden targeted by the intervention</td>
<td>Intervention developer</td>
<td>Intervention duration</td>
<td>Number of sessions</td>
<td>Duration of intervention</td>
<td>Frequency of sessions per week</td>
<td>Length of the sessions</td>
</tr>
<tr>
<td>-------</td>
<td>---------------------</td>
<td>----------------------</td>
<td>-------------------------------------------</td>
<td>------------------------</td>
<td>----------------------</td>
<td>---------------------</td>
<td>-------------------------</td>
<td>------------------------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>Chien and Chan [47]</td>
<td>2013</td>
<td>FPGP</td>
<td>Subjective and objective burden</td>
<td>Content and format were based on previous programs conducted by Li and Arthur [76] in mainland China. Appropriateness of the content was rated by 7 experts, including psychiatrists, psychologists, and nursing specialists.</td>
<td>Peer support—peer leader supported by principal researcher</td>
<td>39 weeks</td>
<td>14 group sessions</td>
<td>Twice weekly</td>
<td>2 hours</td>
</tr>
<tr>
<td>Behrouian et al [62]</td>
<td>2021</td>
<td>Emotional regulation training</td>
<td>Subjective burden</td>
<td>On the basis of the Dialectical Behavior Therapy Skills Workbook and CBT principles</td>
<td>Clinical psychologist</td>
<td>8 weeks</td>
<td>8 sessions</td>
<td>Once weekly</td>
<td>Not specified</td>
</tr>
<tr>
<td>Behrouian et al [63]</td>
<td>2020</td>
<td>Emotion regulation training</td>
<td>Subjective and objective burden</td>
<td>Sessions were based on previous studies (Gratz and Gunderson [77]). Trainings were based on the Dialectical Behavior Therapy Skills Workbook (McKay et al [78]).</td>
<td>Clinical psychologist</td>
<td>8 weeks</td>
<td>8 sessions</td>
<td>Not specified</td>
<td>Not specified</td>
</tr>
<tr>
<td>Faghih and Pahalvanzadeh [79]</td>
<td>2019</td>
<td>CBT</td>
<td>Subjective burden</td>
<td>Researchers based on existing literature</td>
<td>Researcher</td>
<td>8 weeks</td>
<td>16 sessions</td>
<td>Twice weekly</td>
<td>1.5 hours</td>
</tr>
<tr>
<td>Khosravi et al [65]</td>
<td>2021</td>
<td>Spirituality-based program</td>
<td>Subjective burden</td>
<td>Developed by McCann et al [80] in Australia, and its Chinese translated version was validated and refined by the research team.</td>
<td>Psychiatric nurse</td>
<td>8 weeks</td>
<td>6 sessions</td>
<td>Twice weekly</td>
<td>1 hour</td>
</tr>
<tr>
<td>Chien et al [58]</td>
<td>2020</td>
<td>Manual-guided PBSP</td>
<td>Subjective and objective burden</td>
<td>The Good Mood Guide developed by Lifeline South Coast NSW [55], Australia</td>
<td>Researcher via telephone</td>
<td>8 weeks</td>
<td>8 modules</td>
<td>Once weekly</td>
<td>Not specified</td>
</tr>
<tr>
<td>Deb et al [67]</td>
<td>2018</td>
<td>Smartphone-based interventions (MHapps)</td>
<td>Subjective and objective burden</td>
<td></td>
<td>Psychiatrist</td>
<td>Not specified</td>
<td>Not specified</td>
<td>Not specified</td>
<td>Not specified</td>
</tr>
<tr>
<td>Study</td>
<td>Year of publication</td>
<td>Type of intervention</td>
<td>Type of burden targeted by the intervention</td>
<td>Intervention developer</td>
<td>Intervention duration</td>
<td>Number of sessions</td>
<td>Duration of intervention</td>
<td>Frequency of sessions per week</td>
<td>Length of the sessions</td>
</tr>
<tr>
<td>-------</td>
<td>---------------------</td>
<td>----------------------</td>
<td>---------------------------------------------</td>
<td>------------------------</td>
<td>----------------------</td>
<td>---------------------</td>
<td>-------------------------</td>
<td>-------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td>Hyun et al [61]</td>
<td>2018</td>
<td>IEP&lt;sup&gt;l&lt;/sup&gt;</td>
<td>Subjective burden</td>
<td>Developed by Hyun et al [61] for community-living PMIs&lt;sup&gt;g&lt;/sup&gt; based on the empowerment theories of Kanter [81] and McLean [82]</td>
<td>4 sessions</td>
<td>4 weeks</td>
<td>Once weekly</td>
<td>2 hours</td>
<td></td>
</tr>
<tr>
<td>Hou et al [60]</td>
<td>2014</td>
<td>MBSR&lt;sup&gt;2&lt;/sup&gt; program</td>
<td>Subjective burden</td>
<td>Not stated but assumed that the researchers modeled the program on the original MBSR by Kabat-Zinn [83]</td>
<td>8 sessions</td>
<td>8 weeks</td>
<td>Once weekly</td>
<td>2 hours, with a CD of 30-45 minutes for home practice</td>
<td></td>
</tr>
<tr>
<td>Zoladl et al [66]</td>
<td>2020</td>
<td>Participatory care model</td>
<td>Subjective burden</td>
<td>Researchers</td>
<td>8 sessions</td>
<td>12 weeks</td>
<td>Once weekly</td>
<td>1.5 hours</td>
<td></td>
</tr>
<tr>
<td>Ramirez et al [68]</td>
<td>2017</td>
<td>Standard care—included care from general medicine, psychiatry, psychology, family therapy, neuropsychological rehabilitation, and occupational therapy</td>
<td>Subjective and objective burden</td>
<td>Not stated</td>
<td>Medical officer, psychiatrist, psychologist, and occupational therapist</td>
<td>12-18 sessions</td>
<td>10 weeks</td>
<td>Not specified</td>
<td>Not specified</td>
</tr>
</tbody>
</table>

<sup>a</sup>CoMHIP: community mental health early intervention project.  
<sup>b</sup>RESHAPE: Reducing Stigma Among Health Care Providers.  
<sup>c</sup>WHO: World Health Organization.  
<sup>d</sup>mhGAP: Mental Health Gap Action Programme.  
<sup>e</sup>PRIME: Programme for Improving Mental Health Care.  
<sup>f</sup>TPO: Transcultural Psychosocial Organization.  
<sup>g</sup>FLEP: Family Link Education Programme.  
<sup>h</sup>BD: bipolar disorder.  
<sup>i</sup>NAMI: National Alliance on Mental Illness.  
<sup>j</sup>SMI: severe mental illness.  
<sup>k</sup>NGOMH: nongovernmental mental health.  
<sup>l</sup>FPGP: Family-led Peer Support Group Program.  
<sup>m</sup>CBT: cognitive behavioral therapy.  
<sup>n</sup>PBSP: problem-solving–based self-learning program.  
<sup>o</sup>NSW: New South Wales.  
<sup>p</sup>MHapps: mental health apps.  
<sup>q</sup>IEP: integrated empowerment program.  
<sup>r</sup>PMI: person with mental illness.  
<sup>s</sup>MBSR: mindfulness-based stress reduction.
Perceived Effectiveness of the Implemented Intervention Strategies

This review outlines the effectiveness of the implemented intervention strategies as reported by the authors who conducted the studies (Table 4). It should be noted that this scoping review included both published and gray literature, and therefore, no critical appraisal of the studies or meta-analyses were conducted. Therefore, care should be taken as these interpretations may be clouded by author bias [84].

The authors expressed the effectiveness of the implemented intervention strategies by highlighting whether the strategy resulted in a positive or negative outcome for the informal caregivers or their care recipients. Overall, the implemented strategies were reported to be effective in reducing the burden of care and improving the quality of life of informal caregivers. Psychoeducational intervention strategy outcomes included improved coping skills, improved quality of life, reduced anxiety and stress, reduced burden of care, reduced psychological strain, improved knowledge and skills in caregiving, enhanced family functioning, enhanced recovery, relapse prevention, and reduced prolonged admissions. Support group outcomes included alleviation of guilt, enhanced self-help-seeking behavior, improved quality of life, improved knowledge and skills in caregiving, improved psychosocial health, improved family and patient functioning, and decreased demand for mental health services. CBT outcomes included reduction in burden of care, anxiety, and stress and increased resilience of informal caregivers. Guided self-help outcomes included strengthened positive caregiving experiences, reduced negative caregiving experiences, and increased access to support. The reported outcomes for mindfulness and empowerment interventions included empowerment of caregivers, increased self-efficacy, improved quality of life, and decreased depressive symptoms. Collaborative intervention outcomes included reduced burden of care and increased resilience. The standard care outcomes reported included positive outcomes in caregiving, reduction in the burden of care, and increase in social support for caregivers.

Negative outcomes were only reported for the psychoeducational and mindfulness and empowerment intervention strategies. The reported negative outcomes of psychoeducational interventions were no improvement in the burden of care, quality of life, or self-esteem. The reported negative outcomes of mindfulness and empowerment interventions included short-lived improvement in anxiety, and the authors highlighted no improvement after 3 months of follow-up with the participants and no effect on perceived stress, quality of life, and self-compassion, indicating that the effects of the interventions were not sustainable.

Discussion

Principal Findings

This scoping review set out to map the literature on strategies for alleviating the burden of informal caregivers of persons with severe and enduring mental health conditions in low- and middle-income countries. The review identified types of strategies, strategy characteristics, and outcomes of the strategies as reported by the authors. The types of strategies identified were categorized as implemented and recommended intervention strategies and included community-based interventions, psychoeducational interventions, support groups, CBT, spirituality-based interventions, guided self-help, smartphone-based interventions, mindfulness and empowerment, collaborative interventions, standard care, financial and social support, counseling, occupation-based interventions, policy and legislature, and access to mental health.

Most of the implemented and recommended intervention studies (33/44, 75%) were conducted in Asian countries and targeted both subjective and objective burdens, and some studies (18/29, 62%) focused exclusively on subjective burden. This shows an increasing research interest in caregiver interventions in mental health research in Asian countries over the last decade. In contrast, the small number of studies from other low- and middle-income countries possibly confirms the limited research on support strategies for alleviating caregiver burden in mental health. Despite consensus on the high levels of burden of informal caregivers of persons with severe and enduring mental health conditions in these countries [56,85-87] and the importance of supporting caregivers, few studies (44/18,086, 0.24%) implemented and recommended intervention strategies for alleviating caregiver burden. This evidence gap raises significant concerns considering the scarcity of mental health care professionals and limited access to quality mental health services in many low- and middle-income countries. As a result, informal caregivers assume a vital role in the support and care of persons with severe and enduring mental health conditions in these regions [88,89]. Having limited strategies for alleviating the burden of informal caregivers may have dire consequences for the management of persons with severe and enduring mental health conditions in these countries.

The implemented and recommended intervention strategies were mainly focused on both male and female family caregivers. These findings are consistent with the literature as families have long been acknowledged as key stakeholders in the care and management of mental disorders [90-93]. Therefore, it is important to ensure that the strategies for alleviating caregiver burden are targeted specifically to this population, particularly in low- and middle-income countries, where occupying this role is often obligatory [94]. The fact that most implemented intervention strategies were aimed at both male and female caregivers is important to note. Although caregiving is identified as a female-oriented role, there is evidence that men also occupy the primary caregiver role and are also likely to experience high levels of caregiver burden [20,94]. This emphasizes the need to ensure that, where strategies for alleviating caregiver burden are implemented, they focus on all caregivers irrespective of their gender. The care recipients in these studies were mostly diagnosed with schizophrenia and mood disorders. This is consistent with the literature highlighting that depressive disorders, schizophrenia, and bipolar disorders are among the top 10 leading causes of disability in low- and middle-income countries [95].

The findings of this scoping review emphasize the need for evidence-based intervention strategies aimed at alleviating the
burden of informal caregivers in low- and middle-income countries. Most implemented intervention strategies were informed by evidence, in which researchers consulted existing literature and sought expert input to develop their interventions. Trained peer facilitators delivered most of the implemented interventions and were informal caregivers themselves. This is important to note as it aligns with task shifting, which is focused on transferring skills and responsibilities to local people with the aim of increasing access to mental health services in low- and middle-income countries where there is a shortage of human resources [96,97]. In addition, the findings of this review reveal that researchers, nurses, and psychiatrists offered some of the interventions, which may not be sustainable in low- and middle-income countries given the shortage of mental health professionals [95]. Most studies (13/29, 45%) offered 6 to 8 sessions over 6 to 8 weeks, and the sessions were facilitated once or twice a week for 1 to 2 and a half hours. Given the number and frequency of sessions, it may be useful to use peer facilitators to offer these interventions as most are already in the community, which will ensure sustainability in resource-constrained contexts in low- and middle-income countries.

Overall, the implemented intervention strategies were reported to have a positive effect on alleviating the burden of informal caregivers of people with severe and enduring mental health conditions in low- and middle-income countries. In some studies, these effects were reported for both informal caregivers and their care recipients. The authors reported the effectiveness of the strategies in their studies in terms of the help, benefit, and effect that a specific implemented intervention strategy had on the informal caregivers. Some compared their interventions mainly with standard psychiatric care and psychoeducation, which focused on providing caregivers with information on specific or general mental disorders. Psychoeducational interventions were frequently implemented intervention strategies and were identified as helpful, beneficial, and effective in reducing caregiver burden [35,38,42,75]. This type of intervention was reported to improve caregiver knowledge and skills to enable them to cope with the demands of caregiving. In addition, psychoeducational interventions were linked with positive patient outcomes as they were reported to enhance patient recovery, which subsequently reduced relapses and prolonged admissions. These interventions were commended for being simple, feasible, and cost-effective, making them the most preferred form of intervention strategy to address the burden of informal caregivers.

Similarly, implemented intervention strategies such as support groups, community-based interventions, guided self-help, mindfulness and empowerment, and CBT were reported to have a positive effect on the burden of informal caregivers in low- and middle-income countries. It is important to note that the family-led mutual support program, a peer-facilitated intervention, was reported to have long-term desirable benefits on the psychosocial health of both caregivers and their care recipients compared with the psychoeducational program and standard care offered by psychiatrists, clinical psychologists, and nursing specialists. Similarly, the manual-guided problem-solving–based self-learning program, a guided self-help intervention, was reported to have a superior treatment effect on caregiver burden, care recipients’ symptom severity, and the duration of rehospitalizations at the 6-month follow-up compared with a well-accepted family psychoeducation program. This implies that support groups and guided self-help, although not frequently implemented, should be considered beneficial strategies for alleviating caregiver burden. In addition, as self-directed and peer-evaluated interventions, these intervention strategies may be considered cost-effective in extending the services to informal caregivers, especially in low- and middle-income countries where a shortage of human resources affects the delivery and quality of mental health services.

Although not frequently implemented, intervention strategies such as spirituality-based, collaborative, and standard care interventions were also reported to reduce caregiver burden. Spirituality-based interventions were identified as an inexpensive and readily available resource for alleviating the burden of informal caregivers. Similarly, the participatory care model was reported to be an efficient and low-cost method for reducing caregiver burden and increasing caregiver resilience. Therefore, it is necessary to further explore the use of these interventions with informal caregivers of people with severe and enduring mental health conditions in low- and middle-income countries.

Despite their potential, the use of virtual interventions or telehealth strategies for informal caregivers remains underrepresented in the literature. This may be attributed to the relatively slow adoption of telehealth among informal caregivers. Notably, only 1 smartphone app was identified in this scoping review. However, telehealth holds promise as a viable approach for disseminating psychoeducational information, providing real-time support, and facilitating participation in virtual caregiver support groups, particularly in remote areas. The increasing proliferation of smartphones and internet access in Africa further emphasizes opportunities for the development of telehealth programs tailored to caregivers of individuals with severe and enduring mental health conditions. Although most of the implemented intervention strategies reported positive effects on caregiver burden, there were negative outcomes reported for some interventions. Hou et al [60] reported that the mindfulness-based stress reduction program led to a short-lived improvement in the stress and anxiety of the informal caregivers but that no improvement was noted after the 3-month follow-up. In addition, this intervention did not demonstrate a sustained effect on health-related quality of life, perceived stress, or self-compassion. These outcomes were attributed to the loss to follow-up and the fact that the measurement instruments could have been less sensitive to changes in quality of life and stress. Similarly, the 6-session individual psychoeducation intervention for informal caregivers of persons with bipolar mood disorder showed no effect on burden, quality of life, or self-esteem. This was attributed to the reduction in the number of sessions from 21 in the original instruction [71] to 6 in this study [35]. Another reason was that the sessions were conducted individually, whereas the literature highlights that multifamily intervention groups are effective [71,73,98].
Implications for Practice and Policy

The findings of this review provide evidence for the reported evaluated and recommended interventions having benefits in reducing the burden of care among informal caregivers in low- and middle-income countries. Previous studies conducted in low- and middle-income countries have since emphasized the urgent need to strengthen informal caregiver support regarding mental health. The need for information and skills in handling mental health care users, as well as the need for emotional and tangible support, has been highlighted in previous studies [42,99]. The findings of this review revealed that psychoeducation and support groups were highly used strategies for improving knowledge and skills as well as building support for informal caregivers in low- and middle-income countries. Furthermore, these strategies were reported to be beneficial and cost-effective, thereby making them a viable option for implementation in low- and middle-income countries, where limited access to mental health resources prevails [100]. The need to ensure that these strategies are offered on a continuous basis was highlighted in many studies, indicating the need for 6 to 10 sessions over a period of 6 to 10 weeks [7]. To ensure sustainability, training of peer facilitators to deliver these interventions may be realistic as this aligns with a task-shifting approach, which has long been advocated for as a cost-effective strategy for increasing access to mental health services in low- and middle-income countries [101]. It is interesting to note that peer support groups and guided self-help were reported to have long-term benefits on caregiver burden compared with psychoeducation; this highlights that using self-directed interventions is considered a practical option for alleviating informal caregiver burden. In low- and middle-income countries where a shortage of human resources prevails, it is important to consider such interventions as they empower informal caregivers to build support and take ownership of their health and well-being [35]. Spiritual-based interventions, although not frequently implemented, are important to note in the context of low- and middle-income countries as most informal caregivers have been identified as relying on spiritual and religious coping [102,103]. Recommended strategies such as the provision of financial and social support as well as policy and legislature strategies call for an urgent need for policies in mental health to shift focus toward integrating caregiver-oriented services into practice. In addition, these findings highlight the need to adopt an intersectoral approach [104] in which various sectors such as religious and spiritual organizations provide mental health services to extend their accessibility to informal caregivers.

Limitations

In alignment with Arksey and O’Malley, this review did not include a critical appraisal to ascertain the quality of the studies as the purpose of this scoping review was to map existing strategies for alleviating caregiver burden and report on the outcomes as stated by the authors. In their study, Woo et al [105] cautioned that the exclusion of critical appraisals in scoping reviews means that the review cannot ascertain the research gaps that it aims to address if the included studies are of poor methodological quality. Although this scoping review provides evidence on the existing strategies for alleviating caregiver burden, it is important that the suggested strategies be evaluated in the specific context to ensure their effectiveness before implementation in clinical practice. This step was omitted as the purpose of this review was to map the available intervention strategies and the outcomes reported by the authors in alleviating burden among informal caregivers in low- and middle-income countries. This review only included studies conducted in low- and middle-income countries, and only papers written in English were considered, thus reducing the extent and scope of the evidence on the strategies for alleviating burden among informal caregivers regarding mental health.

Conclusions

The findings of this scoping review provided the authors with categories they can use to develop semistructured interview guides to use when exploring the existing formal and informal community mental health services to alleviate the burden of informal caregivers in rural South Africa. The categories outline the different types of strategies that can be used to alleviate caregiver burden, in particular the types of strategies offered to informal caregivers and the intervention developers and deliverers in and outside the mental health care system.

Although most of the included studies (29/44, 66%) implemented these strategies, a few studies conducted in other low- and middle-income countries (15/44, 34%) recommended the use of these interventions to alleviate caregiver burden. Future studies from low- and middle-income countries in other continents, including Africa and South and North America, should address this gap in the research by evaluating these intervention strategies to alleviate the burden of informal caregivers.

Critical appraisal, which is used to ascertain the quality of the studies, was omitted as the purpose of this scoping review was to map existing strategies for alleviating caregiver burden and report on the outcomes as stated by the authors. Future studies should conduct quality appraisals to establish the effectiveness of these strategies in alleviating the burden of informal caregivers. Peer-facilitated support group interventions, although not frequently implemented, were identified as having long-term benefits compared with frequently implemented interventions such as psychoeducation. It is recommended that future research be directed at implementing and evaluating these interventions to alleviate burden in low- and middle-income countries.

Acknowledgments

The authors wish to acknowledge Mr Devind Peter, a librarian at the Wits Health Sciences Library, for assisting with the development of the search strategy and retrieval of the literature sources for this scoping review. This study was supported by the Consortium for Advanced Research Training in Africa. The Consortium for Advanced Research Training in Africa is jointly led by the African Population and Health Research Center and the University of the Witwatersrand and funded by the Carnegie
Corporation of New York (grant G-19-57145); Swedish International Development Cooperation Agency (grant 54100113); Uppsala Monitoring Center; Norwegian Agency for Development Cooperation; and Wellcome Trust (reference 107768/Z/15/Z) and UK Foreign, Commonwealth, and Development Office with support from the Developing Excellence in Leadership, Training, and Science in Africa program. The statements made and views expressed are solely the responsibility of the fellow.

Data Availability

All data generated or analyzed during this study are included in this published paper.

Authors’ Contributions

OS completed the literature search. OS and DC completed the screening of titles, abstracts, and full texts of the identified studies. In addition, they extracted data from all the included studies and completed the draft of the manuscript. FA and NGN reviewed the manuscript draft. OS incorporated the feedback from the authors for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Data extraction templates.

[DOCX File, 23 KB - ijmr_v13i1e48587_app1.docx ]

Multimedia Appendix 2

PRISMA-ScR Checklist.

[PDF File (Adobe PDF File), 156 KB - ijmr_v13i1e48587_app2.pdf ]

References


103. Modise TP, Mokgaola IO, Sehularo LA. Coping mechanisms used by the families of mental health care users in Mahikeng sub-district, North West province. Health SA 2021 Aug 16;26:1586 [FREE Full text] [doi: 10.4102/hsag.v26i0.1586] [Medline: 34522392]


Abbreviations

CBT: cognitive behavioral therapy
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
Digital Methods for the Spiritual and Mental Health of Generation Z: Scoping Review

Susanna Y Park1*, PhD; Bridgette Do2*, MPH, PhD; Jacqlyn Yourell2, MS, PhD; Janice Hermer3, MA; Jennifer Huberty2, PhD

1Skylight, Radiant Foundation, Salt Lake City, UT, United States
2Fit Minded Inc, Phoenix, AZ, United States
3Arizona State University, Tempe, AZ, United States
*these authors contributed equally

Corresponding Author:
Susanna Y Park, PhD
Skylight
Radiant Foundation
55 N 300 W Suite 800
Salt Lake City, UT, 84101
United States
Phone: 1 (720) 675 7228
Email: susanna@skylight.org

Abstract

Background: Generation Z (Gen Z) includes individuals born between 1995 and 2012. These individuals experience high rates of anxiety and depression. Most Gen Z individuals identify with being spiritual, and aspects from religion and spirituality can be integrated into mental health treatment and care as both are related to lower levels of depression. However, research on the spiritual and mental health of Gen Z is sparse. To date, there are no systematic or scoping reviews on digital methods to address the spiritual and mental health of Gen Z.

Objective: This scoping review aimed to describe the current state of digital methods to address spiritual and mental health among Gen Z, identify the knowledge gaps, and make suggestions for how to leverage digital spiritual and mental health interventions for Gen Z.

Methods: A comprehensive literature search was conducted in PubMed, Scopus, PsycInfo, CINAHL, Education Full Text, Google Scholar, SocIndex, and Sociological Abstracts. The inclusion criteria were as follows: (1) study population born between 1995 and 2012 (ie, Gen Z); (2) reporting on spiritual health or well-being, spirituality or religion, and mental health or well-being; (3) reporting on using digital methods; (4) publication in 1996 or beyond; (5) human subject research; (6) full text availability in English; (7) primary research study design; and (8) peer-reviewed article. Two authors screened articles and subsequently extracted data from the included articles to describe the available evidence.

Results: A total of 413 articles were screened at the title and abstract levels, of which 27 were further assessed with full text for eligibility. Five studies met the inclusion criteria, and data were extracted to summarize study characteristics and findings. The studies were performed across 4 different countries. There were 2 mixed-methods studies (South Africa and Canada), 2 cross-sectional studies (China and United States), and 1 randomized controlled trial (United States). Of these studies, only 2 discussed digital interventions (a text messaging–based intervention to improve spiritual and mental health, and a feasibility study for a mental health app). Other studies had a digital component with minor or unclear spiritual and mental health measures. Overall, there was a lack of consistency in how spiritual and mental health were measured.

Conclusions: Few studies have focused on assessing the spiritual and mental health of Gen Z in the digital context, and no research to date has examined a digital spiritual and mental health application among Gen Z. Research is needed to inform the development and evaluation of approaches to address the spiritual and mental health of Gen Z via digital means (eg, mobile apps).

KEYWORDS
Generation Z; Gen Z; spiritual health; digital mental health; spirituality

Introduction

The digital age is marked by widespread internet use and the ability to quickly communicate and find information online. Individuals born from 1995 to 2012 are considered “digital natives,” as they are the first generation to live in an age where technology and the internet are accessible at all times [1]. The Pew Research Center considers these individuals as part of Generation Z (“Gen Z”) and identifies the beginning of this generation to be in 1997 [1]. For the purpose of our research, we used the definition of Gen Z by Katz et al [2], who defined Gen Z as individuals born from 1995 to 2012, to account for those born when the World Wide Web made its public debut in 1995.

As a generation that grew up with technology, digital devices are familiar and seamlessly woven into the daily routines of Gen Z. It has been reported that 95% of Gen Z individuals have access to a smartphone [3]. In a study of 1000 Gen Z individuals aged 13 to 25 years, more than half spent 4 or more hours online compared with just 28% of all US adults who spent 4 or more hours online [4]. In a world where it is nearly impossible to socialize, work, and get an education without technology, Gen Z individuals are “always on,” and this is associated with higher rates of depression, attention deficit disorder or attention deficit hyperactivity disorder, and technology addiction [5]. Compared with other generations, Gen Z individuals spend more time alone or on digital communication platforms than engaging in in-person interactions. Between 2000 and 2015, Gen Z high school seniors spent an hour less on in-person social interactions compared with early millennials [6]. Adolescents who spend more time on social media than in in-person interactions are the loneliest compared with those who spend less time on social media. Moreover, between 2007 and 2018, there were great increases in the relative percentage rates for self-injury (47%), seriously considering suicide (76%), and suicide attempts (58%) among Gen Z [7].

The American Psychological Association (APA) states that Gen Z individuals are more likely to report mental health concerns (e.g., depression and anxiety) than previous generations [8]. In a 2022 survey of 1055 Gen Z adults, 1 out of 4 reported having more bad days than good within a 1-month time frame. More than 2 out of 5 (42%) had a diagnosed mental health condition, with more than a quarter of those being diagnosed during the COVID-19 pandemic (March 2020) or later. Anxiety and depression are the 2 largest mental health issues among Gen Z, with 9 out of 10 individuals diagnosed with a mental health condition having anxiety and 8 out of 10 having depression [3]. Notably, Gen Z individuals are the most comfortable discussing their mental health [3,8]. One third of Gen Z individuals report posting about their mental health on social media. They also attend therapy and are willing to pay out of pocket for mental health care and services [3]. Despite the comfort of Gen Z in talking about their mental health, there is a crucial need to address the high rates of anxiety, depression, and other health issues that they experience [8].

Spirituality may be an untapped resource to address the mental health crisis experienced by Gen Z today. While spirituality can serve as a component within organized religion, the 2 aspects are distinct. Religion is an organized belief or specific set of practices focusing on a higher power (i.e., Christian, Muslim, Buddhist, etc) [9]. Spirituality is a broader concept in which individuals seek connection to self, others, nature, and a sacred or higher being [10]. Individuals may identify with being either religious or spiritual, or both. Gen Z individuals do not necessarily identify with a particular religion or belief but instead practice spirituality. Only half of Gen Z individuals report turning to their faith for support in times of uncertainty [11], and they are more likely to engage in spiritual practices than religious practices [12]. In a study of 10,000 Gen Z individuals aged 13 to 25 years, 68% considered themselves religious and 77% considered themselves spiritual [13]. Gen Z individuals define spirituality as autonomous and faith unbundled, and it is inclusive of all faiths and practices [11,13].

Spirituality is related to several positive health and psychosocial outcomes, namely greater mental health [14]. A recently updated review of the literature on the relationship between spirituality and mental health found that greater spirituality was associated with lower depressive symptoms, lower suicidality, and lower substance abuse [14]. Gen Z individuals face some of the highest rates of mental health conditions (e.g., depression) [15,16]; thus, spirituality should be considered in addressing youth mental health today. In the aforementioned report by Singer [13], the majority of Gen Z individuals attributed their spiritual connection to their positive mental health state. Another aspect of mental health that is influenced by spirituality is quality of life among chronically and terminally ill patients. Palliative care patients who struggle with spirituality report poorer quality of life compared with those who feel stable with their spirituality [17]. Additionally, teens and young adults with cancer mention searching for meaning, hope, and life perspectives, even though they may not consider themselves as spiritual [18]. Interventions that promote spiritual well-being (one’s sense of purpose, meaning in life, and connection to something greater [19]) may be a powerful resource for improving mental health in Gen Z.

Research on digital mental health interventions and spirituality exist separately. Little is known about digital methods (e.g., mobile apps, text messaging, etc) that incorporate both spiritual and mental health among Gen Z. Scoping or systematic reviews on this topic are nonexistent, and research on this topic is very limited. Given that technology is woven into the daily lives of Gen Z, digital mobile apps that promote spirituality may offer a novel approach to supporting the mental health of Gen Z adolescents and young adults. Therefore, the purpose of this scoping review was to describe the current state of digital methods to address spiritual and mental health among Gen Z, identify the knowledge gaps, and make suggestions for how to leverage digital spiritual and mental health interventions for Gen Z.
Methods

Eligibility Criteria
The inclusion criteria for targeted articles were as follows: (1) study population born between 1995 and 2012 (ie, Gen Z); (2) reporting on spiritual health or well-being, spirituality or religion, and mental health or well-being; (3) reporting on using digital methods; (4) publication in 1996 or beyond; (5) human subject research; (6) full text availability in English; (7) primary research study design; and (8) peer-reviewed article.

Information Sources
Guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) extension for scoping reviews (PRISMA-ScR) [20], searches were conducted in PubMed, Scopus, PsycInfo (EBSCO), CINAHL Plus with Full Text (ProQuest), Education Full Text (H.W. Wilson), Google Scholar, SocIndex with Full Text, and Sociological Abstracts (ProQuest) on May 2 and 3, 2023, by a health sciences librarian (J Hermer) at Arizona State University. The same information sources were searched on September 21-26, 2023, and again on November 7-8, 2023, and were also included in the results. ERIC and Atla Religion Database were also searched, but they returned no results.

Search Strategy
The searches were optimized for each individual database but included a combination of keywords and subject headings for the following 4 categories: Generation Z or Gen Z; spirituality or religion; mental well-being; and mobile health, mHealth, and eHealth (Multimedia Appendix 1). Owing to the very limited nature of the results, only language filters were applied to ensure that all relevant literature was available to be screened. All records were imported into Zotero to check if any articles were retracted and then into Covidence systematic review software for deduplication and screening [21]. Apart from the initial search, 2 additional searches were conducted to ensure that all relevant literature was found and to add any recent literature from the following 6 months. A search of all databases was performed on September 21-26, 2023, using the additional keywords of “faith,” “transcenden*,” and “life purpose” or “existential needs,” and a final search was performed on November 7, 2023, using the additional keywords of “spiritu*” and “relig*.” The scoping review protocol was registered online in Open Science Framework (OSF) on May 16, 2023.

Selection Process
The selection process was completed entirely on Covidence. Prior to reviewing titles and abstracts, duplicate titles were eliminated by Covidence. Two authors (SYP and BD) screened all titles and abstracts independently and were blinded to each other’s decisions. Any disagreements were discussed between the authors SYP and BD, and agreed upon before full-text review. Agreement scores for article selection between the 2 authors were not logged in Covidence for the initial search; however, the agreement scores were 85% (81/95) and 90% (80/87) for reviewing titles and abstracts between the 2 authors for the September and November searches, respectively. For full-text article review, the author SYP independently reviewed half of the articles, and the author BD independently reviewed the remaining articles. The authors SYP and BD deliberated with each other, and with J Huberty and JY if there were any questions regarding inclusion based on the article eligibility criteria. Articles were excluded if the study population included Gen Z but did not explicitly distinguish Gen Z in the population sample and the results were not disaggregated by age. The final articles included in the review were agreed upon by all 4 authors.

Data Collection Process and Synthesis Methods
Prior to the search, all authors agreed on the following characteristics for data extraction and synthesis: title; authors; study country; study objectives; study design; data collection timeframe; recruitment methods; sample size; participant characteristics; description of digital methods; constructs related to religion, spirituality, or spiritual well-being; assessment or measure for religion, spirituality, or spiritual well-being; constructs related to mental health; assessment or measure for mental health; main findings; and study limitations. These characteristics were selected to ensure a detailed understanding of available literature as it relates to the goal of the scoping review. For articles included in the scoping review, the author SYP independently extracted data based on the a priori characteristics for half of the articles and the author BD independently extracted data from the remaining articles. Given that each article had data extracted by a single author, there were no agreement scores. After evaluating the included articles, study characteristics and main findings were summarized in a descriptive manner.

Some studies that initially appeared eligible for this review were ultimately excluded because they did not meet specific criteria. For instance, a cross-sectional study on the perceptions of 475 Gen Z individuals and young millennials and their use of a spiritual self-care app [22] was originally included when reviewing titles and abstracts. However, upon full-text review, we found that the study results did not distinguish between Gen Z individuals and young millennials, thus failing to meet our review’s criteria (ie, Gen Z only).

Results

Study Characteristics
Details on article selection are illustrated in a PRISMA diagram (Figure 1). Our search identified 824 articles from 8 databases based on the search terms. After removing 411 duplicates in Covidence, 413 articles were screened by title and abstract. After the initial screening based on the inclusion criteria, the full texts of the remaining 27 articles were screened. Review articles were not included in the final review, but references were screened to see if any additional literature was admissible. Ultimately, the scoping review included 5 articles. Of these 5 articles, 1 [23] was included in the review from the updated search conducted in September 2023 and 1 [24] was included from the updated search in November 2023. Characteristics and results of the studies are summarized in Table 1.
Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of article selection.

**Records identified from databases (n=824)**
- PubMed (n=186)
- Scopus (n=57)
- PsycInfo (n=63)
- CINAHL (n=28)
- Google Scholar (n=190)
- SociNDEX (n=11)
- Sociological Abstracts (n=300)
- Education Full Text (n=9)

**Duplicate records removed (n=411)**

**Titles and abstracts screened (n=413)**

**Records excluded (n=385)**

**Full-text articles sought for retrieval (n=28)**

**Articles not retrieved (full text unavailable) (n=1)**

**Full-text articles assessed for eligibility (n=27)**

**Articles excluded:**
- Wrong study population (ie, not Generation Z) (n=10)
- Did not examine spirituality/religion (n=2)
- No digital methods (n=1)
- Wrong study design (eg, review) (n=6)
- Not peer-reviewed (n=3)

**Articles included in scoping review (n=5)**
## Table 1. Characteristics and main findings of the included studies.

<table>
<thead>
<tr>
<th>Reference and study country</th>
<th>Study design</th>
<th>Sample characteristics</th>
<th>Digital aspect</th>
<th>Religion/spirituality measures</th>
<th>Mental health measures</th>
<th>Main findings</th>
</tr>
</thead>
</table>
| Mindu et al [25], 2023 (South Africa) | Mixed-methods study | 93 youth and young adults aged 16-24 years; 44% female | Assessed participants’ knowledge and preferences for a digital mental health app | Where do youth seek treatment or assistance when they have mental health problems (Response options: Visit a spiritual healer; Go to a church for prayers; Traditional medicine; Clinic/hospital; Visit a health care worker) | Mental health awareness (eg, taught about mental health and no prior education on mental health) |  • No participants had experience using a mental health app, but 99% indicated mental apps are important and can benefit youth.  
  • Religious and cultural beliefs were a barrier in the community (eg, substance abuse or misuse, posttraumatic stress disorder, depression, and anger) |
| Gao et al [26], 2021 (China) | Cross-sectional study | 1017 first-year college students (mean age 19 years); 77.8% female | Participants were recruited through an eHealth application to complete a survey | Religion (Response options: No religion; Buddhism; Christian; Others) | Generalized Anxiety Disorder-7 (GAD-7) and depression; Patient Health Questionnaire-9 (PHQ-9) [27,28] |  • 95.3% indicated having no religion.  
  • Belief in Christianity and in Buddhism were associated with greater anxiety. |
| Craig Rushing et al [29], 2021 (United States) | Randomized controlled trial | 833 American Indian or Alaska Native teenagers and young adults aged 15-24 years; 66.3% female | The 2 intervention arms included 3 SMS text messages per week for 8 weeks with information, role model videos, images, and engagement opportunities (eg, reply for more information, resource links, etc) | Rate your spiritual health (Response options: Excellent; Very good; Good; Fair) [30] | Rate your mental health (Response options: Excellent; Very good; Good; Fair) [30] |  • No significant differences between the 2 intervention arms within subjects.  
  • Mean scores of perceived health (physical, mental, and spiritual) significantly increased over time for both intervention arms.  
  • Those who reported better health also reported greater cultural resilience, identity, and cultural pride. |
| Reed et al [23], 2022 (United States) | Cross-sectional study | 349 American Indian or Alaska Native youth aged 15-24 years; 71.1% female | Assessed participants’ use of media technologies (ie, media types, frequency, and duration) and how they use media technologies (ie, online behaviors and activities) | Select the top 3 health topics from a list of 15, including spiritual health | Self-reported mental health (How good is your mental health? Response options: 4-point Likert-type scale; 4=Excellent and 1=Poor) |  • 53.5% of participants relied heavily on the internet to access health information.  
  • Nonsexual and gender minority participants reported better mental health than sexual and gender minority participants.  
  • The top 3 most important health topics were Native identity, mental health, and social justice and equality.  
  • Spiritual health was selected as the most important health topic 14% of the time. |
Overall, participants expressed high interest in using a digital app designed for individuals in South Africa. Qualitative results underscored the severe lack of mobile mental health apps among American Indian or Alaska Native teenagers and young adults. Participants were randomized to participate in 1 of 2 groups: (1) An 8-week intervention arm involving 3 SMS text messages per week designed to elevate and reaffirm Native voices in STEM (science, technology, engineering, and mathematics) and media production. Participants were in both arms, and both had messaging that included a combination of information, breathing techniques, and reflective awareness of thoughts and intentions, and a way of being that is expressed through self-compassion and compassion for other beings. The study authors acknowledged that while digital methods or interventions were implemented in the study, the results provided a unique Gen Z perspective on the usability of digital methods.

Study by Gao et al, 2021

Gao et al [26] conducted a cross-sectional study in China among first-year Chinese college students. Using a health management app (Residents e-Health), a questionnaire was distributed to examine depression and anxiety, and their associations with other health-related constructs, such as stress and nutrition. The authors examined participants’ religious beliefs (ie, Christian, Buddhism, other, and no religion) as a correlate of anxiety and depression. However, most of the sample (95%) did not have a religion. Belief in Christianity and belief in Buddhism were associated with greater anxiety in the sample. The study authors noted that the large number of participants ascribing to no religion limited the understanding of the relationship between religion and mental health. The overall conclusion of the study indicated that early lifestyle interventions assessing religion, as well as other demographic and behavioral factors, are important for understanding the factors contributing to mental health in Gen Z.

Study by Craig Rushing et al, 2021

Craig Rushing et al [29] conducted a randomized controlled trial in the United States to examine the efficacy of an mHealth intervention (BRAVE) for physical, mental, and spiritual health: resilience; self-esteem; and coping and help-seeking skills among American Indian or Alaska Native teenagers and young adults. Participants were randomized to participate in 1 of 2 groups: (1) An 8-week intervention arm involving 3 SMS text messages per week highlighting common coping strategies, preferred wellness strategies, help-seeking skills, and related protective factors such as cultural resilience, identity, and cultural pride; or (2) An 8-week control arm involving 3 SMS text messages per week designed to elevate and reaffirm Native voices in STEM (science, technology, engineering, and mathematics) and media production. Participants were in both arms, and both had messaging that included a combination of information, meditation, and reflective awareness of thoughts and intentions, and a way of being that is expressed through self-compassion and compassion for other beings.
role model videos, images, and opportunities of engagement (eg, reply for more information and links to access resources). The findings indicated that there were no significant differences between the 2 study arms within subjects, such that participants in the intervention arm did not report better outcomes than those in the control arm. However, mean scores of perceived health (ie, physical, mental, and spiritual) significantly increased over time in both arms. In addition, participants who reported better health also reported greater cultural resilience, identity, and cultural pride. The percentage of participants who used the resources and information in the intervention arm text messages also increased over 5 months. The study authors acknowledged some limitations, including high or favorable survey measure outcomes at baseline, only a 1-week break between receiving interventions in the arms, and control messaging likely being novel and helpful. Another limitation to note is the assessment of perceived health, which combined 3 separate survey measures assessing perceived physical, mental, and spiritual health. The combination of 3 different aspects of health makes it challenging to derive valid inferences regarding mental or spiritual health on their own. Overall, the intervention demonstrated improved health outcomes and underscores the acceptability of text messaging to promote and support well-being.

**Study by Reed et al, 2022**

Reed et al [23] conducted a cross-sectional study in the United States, in which 349 American Indian or Alaska Native youth (aged 15-24 years) were asked about the extent to which they use media technologies, how they use technologies, and their health priorities. Several trends were revealed. The majority of participants (64.7%) reported sending 1 to 50 text messages a day. Instagram was the most popular daily technology used, and 65.3% of participants reported using social media 3-7 hours per day. Participants also self-rated their mental health. The findings indicated that nonsexual and gender minority youth (56.7%) reported better mental health than sexual and gender minority youth (36.4%). To better understand important health topics, participants were asked to select their top 3 health topics from a list. The most popular topic selected was Native identity or cultural pride (73%), followed by mental health (57%) and social justice and equality (31%). Spiritual health was selected by 14% of youth in their top 3 health topics. While spiritual health was not among the top health topics selected, it is important to note that the range of options offered spanned across several categories (eg, social justice or inequality, alcohol and drug use or abuse, and the environment). Overall, the study authors concluded that building resources that foster cultural pride and positive identity must be included in any programs or technologies for addressing mental health among American Indian or Alaska Native youth.

**Study by Au-Yeung et al, 2023**

Au-Yeung et al [24] conducted a Haudenosaunee (Canada’s largest First Nations reserve, Six Nations of the Grand River) community-based study in which 5 Haudenosaunee youth (aged 11-16 years) tested the JoyPop mobile app (available in English and French) that is designed to promote resilience among youth. The app offers breathing exercises, mood tracking, journaling, personalizable social support, a 24-hour helpline, and games. The results indicated that participants had mixed ratings on their self-reported mental wellness, ranging from “fair” to “very good,” and that they used the app 1 to 3 times a day. Of the 5 participants, 4 were interviewed about their experiences using the app, perspectives of mental wellness, and characteristics of a “Good Mind” (an Indigenous concept on the physical, psychological, and spiritual journey that maintains balance and harmony in a person). Interview participants reported that the app was easy to use and esthetically pleasing. They also enjoyed all of the app’s features, with the exception of the Circle of Trust feature (ie, personalizable social support). Interview participants identified positivity and happiness, understanding emotions, acts of kindness, personal hobbies, and positive body language as important to their mental wellness. Participants also discussed important characteristics of a Good Mind, such as positivity, kindness, and connecting with nature. Overall, the app was favorable to the participants, but they suggested incorporating specific features like words in their own language and Indigenous visuals (eg, feathers and clan animals). While the app lacked explicit content on a Good Mind, the authors recommended incorporating concepts of a Good Mind to enhance its relatability to Haudenosaunee youth, given its cultural significance to the Haudenosaunee people. Although the study had a small sample size, the authors concluded that mobile health interventions can be beneficial to Indigenous youth, as mental health apps continue to be of interest and Indigenous cultures value the promotion of health and resilience. Further, Indigenous tribes across North America have unique perspectives, and pan-Indigenous resilience apps like JoyPop will need to be tailored to specific cultural contexts.

**Discussion**

**Overview**

The purpose of this scoping review was to describe the current state of digital methods to address spiritual and mental health among Gen Z, identify the knowledge gaps, and make suggestions for how to leverage digital spiritual and mental health interventions for Gen Z. A comprehensive literature search across 8 databases identified only 5 relevant studies, emphasizing the significant lack of published research on digital methods to address the spiritual and mental health of Gen Z. Among the 5 studies, only 2 discussed digital interventions, and of these studies, 1 examined a text messaging–based intervention to improve spiritual and mental health and reported improvements in spiritual and mental health over time [27], and 1 examined the feasibility of a mental health app [24]. The sparse available literature limits conclusions on the current state of digital methods to address spiritual and mental health, and warrants future research to address these gaps.

**Current State of Digital Methods for Spiritual and Mental Health**

Gen Z individuals are facing a mental health crisis as they experience high rates of depression and anxiety [8,16]. For example, one of the included studies reported that the prevalences of anxiety and depression among college freshmen were 40.3% and 45.3%, respectively [26]. Gen Z individuals also report feeling lonely, having low self-confidence, and being...
distressed about the future [3,5,12]. All 5 studies included in this review discussed the overall well-being of Gen Z and, to some extent, aspects of spirituality. The findings suggest that spirituality may play a role in the mental health of Gen Z and should be considered in the development and implementation of future digital applications that address mental health. The study by Mindu et al [25] particularly underscores the need for customizability in mental health applications considering that participants expressed that using digital platforms (ie, social media) conflicted with their religious and cultural beliefs. Thus, the introduction of applications that allow users to engage in practices and view content that aligns with their values and beliefs is a potential avenue for combating this barrier and could in turn strengthen one’s spirituality. However, research is needed to determine the acceptability of digital mental health applications that involve spiritual content.

Nearly every facet of the lives of Gen Z involves technology, for example, using computers for school or work, using mobile apps to order food, and using social media and texting to communicate with friends. Despite the widespread use of technology among Gen Z and the increasing number of research studies employing digital methods to test and deliver mental health programs and interventions [31], only 1 of the studies included in this review examined the effects of a digital intervention on spiritual and mental health [27] and I assessed the feasibility of a mental health mobile app [24]. The lack of digital methods or interventions to address the spiritual and mental health of Gen Z warrants the development of accessible ways for Gen Z to practice spiritual and mental self-care. Gen Z individuals spend more time (4 or more hours daily) on social media compared with other generations [3], and poor mental health is often attributed to social media use [32]. However, social media is not necessarily harmful and is instead dependent upon what Gen Z individuals do and see online, their pre-existing strengths or vulnerabilities, and the environment in which they are raised (eg, parental monitoring) [32,33]. Digital methods for spiritual wellness that are specifically targeted for Gen Z and built to empower Gen Z to practice self-care and build healthy coping mechanisms may benefit the mental health of Gen Z. Results from this review illustrate the acceptability of a mobile intervention promoting spiritual and mental health [24,29], and the interest Gen Z individuals have toward using digital methods to address their spiritual and mental health [25]. However, given that little is known about this topic, more research is needed to truly grasp the feasibility and efficacy of digital approaches to address spiritual and mental health with regard to the well-being of Gen Z.

Gaps in the Literature
Published literature or research about the spiritual and mental health of Gen Z is limited, reporting mostly mental health statistics rather than examining determinants of mental health or interventions for improving mental health (eg, spirituality) [1,33]. Research to date on spirituality in Gen Z has only been performed by faith-based organizations or by nonacademic research institutions and has mainly focused on comparing the views of Gen Z on religion and spirituality to the views of other generations [11,12,34]. For example, the majority of Gen Z individuals (77%) identify as spiritual, preferring to ascribe to a set of values from various beliefs [13]. While we know that the majority of Gen Z individuals identify as spiritual, empirical research that specifically examines their spiritual practices and preferences, overall spiritual well-being, and associations with mental and physical health outcomes is warranted. Among the 5 studies included in this review, 3 examined spiritual health. The randomized controlled trial [29] asked participants to rate their spiritual health using a combined measure for overall health that included physical, mental, and spiritual health, limiting the ability to assess spiritual well-being explicitly among these participants. The cross-sectional survey [23] asked participants to select the top 3 health topics important to them (eg, mental health, spiritual health, and Native identity), but the study did not examine any associations between spiritual health and mental health. Spiritual health was not defined for participants in either study. The mixed-methods study in Canada [24] interviewed participants who used an app designed to promote resilience in youth with regard to a Good Mind, an Indigenous concept on the physical, psychological, and spiritual journey that maintains balance and harmony in a person. However, this concept is specific to the Haudenosaunee people; therefore, the findings may not be applicable to the perspectives on spirituality of other young communities. In addition, only 1 article administered a valid measure to assess mental health, spirituality, or spiritual well-being (ie, Generalized Anxiety Disorder-7 [GAD-7]) [26]. The validity and reliability of instruments used to measure mental health and spirituality or spiritual well-being in the study samples were not reported in any of the included articles. Identified gaps offer research opportunities to comprehensively examine spiritual and mental health among Gen Z.

Suggestions for Leveraging Digital Spiritual and Mental Health Interventions for Gen Z
Gen Z individuals are remarkably familiar with navigating digital spaces and integrating spirituality into their lives, and are the most comfortable talking about their mental health compared with other generations [3]. Gen Z individuals who have a spiritual connection have better perceptions about their mental health and believe their spiritual health contributes to their mental health, compared with those who do not have a spiritual connection [11]. Since Gen Z individuals primarily consume information through technology, digital interventions are a promising method to teach and facilitate various practices in spiritual and mental self-care. For example, the digital intervention content of Craig Rushing et al [29] included wellness strategies, such as self-care and goal setting, which resulted in improved perceived health (eg, physical, mental, and spiritual) for participants. In the study by Mindu et al [25], digital mobile interventions through social media were concluded to be potentially useful to increase mental health literacy and knowledge of resources. Additionally, the study by Gao et al [26] suggested that early interventions that target the lifestyle behaviors of Gen Z (ie, smoking) can improve their depression and anxiety. A spiritual self-care app tailored for Gen Z, for example, may allow Gen Z to engage with full autonomy and convenience. Rather than requiring Gen Z to seek out places to practice religious or spiritual beliefs, a spiritual self-care app can be available wherever they are and
whenever they want. Digital spiritual self-care interventions should also consider incorporating topics that are deemed important to Gen Z, such as cultural relevance, inclusivity, social justice, and nature [23,24].

Mobile apps grant users the ability to customize their experience, which vastly differs from traditional means of practicing religion that ascribe to a predetermined set of beliefs, values, and practices that are often fixed [13]. Access to technology enables Gen Z individuals to autonomously decide which aspects of different spiritual beliefs and practices they resonate with. Most digital apps include features that allow users to customize their experience and the content they engage with. Thus, a mental health app that includes various components of spiritual practices that users can choose from may provide a new way for young people to tend to their spiritual and mental health. Based on this review, research examining the feasibility, acceptability, and efficacy of digital health tools specifically targeting spiritual and mental health among Gen Z is absent, indicating the need for research in this area. Researchers, companies, and nonprofit organizations can leverage existing digital spaces that Gen Z individuals frequently use (eg, Instagram and TikTok) to garner feedback on what they might desire in a digital spiritual and mental health intervention. For example, 2 of the studies [24,29] incorporated feedback from Indigenous youth on existing digital methods (eg, SMS text messaging and JoyPop) to inform the development of their digital content. Spiritual well-being apps, for example, can be used to target Gen Z and deliver evidence-based content that integrates spirituality and mental health. These studies also emphasize the importance of culturally relevant interventions that can speak to diverse cultural backgrounds and beliefs among Gen Z.

**Strengths and Limitations of This Review**

There are multiple strengths of this review. First, this scoping review indicates that there is a limited knowledge base surrounding digital methods for addressing the spiritual and mental health of Gen Z and summarizes the current state of the literature on this topic. This is the first scoping review to address this topic and highlights a crucial gap in supporting young people’s mental health. Additionally, a librarian was consulted and involved in the search process to bolster the rigor and accuracy of the review. Covidence was used to minimize human error in screening eligible articles. Along with these strengths, there are some limitations in this review. First, a scoping review limits the objective understanding of a topic such that quantitative results cannot be compiled to determine effect sizes across studies. Second, database searches are not uniform and require nuanced search methodologies, which can result in relevant studies being missed. Third, manuscripts during the full-text screening stage were split between the authors SYP and BD, and thus, manuscripts were assessed by a single rater. Finally, research on this topic is severely lacking, which limits the number of articles included in this review and impacts the ability to construct a cohesive narrative or draw definitive conclusions about the state of the field. This underscores the necessity of this review to highlight the gaps and urge further investigation.

**Scientific Contribution**

The goal of this scoping review was to assess the extent to which research on digital methods to address the mental and spiritual health of Gen Z has been conducted. The scoping review revealed a lack of available research on spirituality and mental health. Specifically, there is a dearth of studies on the use of digital methods to deliver spiritual well-being for mental health in Gen Z. Despite an increase in mental health concerns among Gen Z and the growing body of evidence on the beneficial effects of spiritual self-care on mental health, few published articles touch on this topic. The findings from this review highlight the opportunity for addressing the mental and spiritual health of Gen Z through digital methods (eg, mobile apps). The use of digital methods to address mental health is a growing area of research; however, spirituality and spiritual self-care have received little attention. There is potential for researchers to examine spiritual self-care, which is delivered through digital methods, and its impact on populations experiencing significant mental health problems. Overall, the scoping review underscores the need for future research to examine the acceptability and feasibility of digital approaches to address spiritual and mental health among Gen Z.

**Conclusion**

This scoping review underscores the dearth of research surrounding digital methods to address spiritual and mental health among Gen Z. Considering that digital methods to address aspects of mental health are increasingly popular and effective [35,36], research is needed to examine digital platforms that address spiritual and mental health. This is especially pertinent for Gen Z individuals as they have some of the greatest rates of mental health issues and are the most digitally savvy generation to date, and most indicate that they are spiritual. Leveraging spirituality as a way to address mental health among Gen Z via digital means offers a novel and relevant approach for addressing the mental health crisis impacting young people today.

**Acknowledgments**

The authors would like to thank Clare Beatty for the organization of this project and Justin Jones for inspiring us to write this paper.

**Data Availability**

The full review process, including the search terms and databases used, is available as a supplementary file.
Authors' Contributions
SYP and BD wrote and reviewed this paper. J Hermer searched and curated articles for review. JY reviewed and edited the manuscript. J Huberty supervised the review.

Conflicts of Interest
SYP is currently the Behavioral Research Scientist for the Radiant Foundation and does not receive incentives for the outcomes of the research. J Huberty is the Chief Science Officer for the Radiant Foundation’s Skylight app. J Huberty was hired by the Radiant Foundation to lead scientific strategy and mentor the scientists for Skylight, but does not receive incentives for the outcomes of the research. BD and JY are employees of J Huberty and are independent from the Radiant Foundation leadership. The role of SYP and J Huberty is to ensure the quality of science regarding the Skylight app. J Hermer is a librarian at Arizona State University and is not associated with the Radiant Foundation or with J Huberty. The authors do not have stock in Skylight and receive no financial incentives from the sales of Skylight. The authors do not receive any financial benefit from the outcomes of this study.

Multimedia Appendix 1
Databases and search terms used for review.
[DOCX File, 23 KB - ijmr_v13i1e48929_app1.docx ]

Multimedia Appendix 2
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[PDF File (Adobe PDF File), 239 KB - ijmr_v13i1e48929_app2.pdf ]

References


33. Gen Z: Generation Z

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Abbreviations

Gen Z: Generation Z

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
©Susanna Y Park, Bridgette Do, Jacqlyn Yourell, Janice Hermer, Jennifer Huberty. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 06.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.
Improvements in Neoplasm Classification in the International Classification of Diseases, Eleventh Revision: Systematic Comparative Study With the Chinese Clinical Modification of the International Classification of Diseases, Tenth Revision

Yicong Xu\textsuperscript{1,*}, MPH; Jingya Zhou\textsuperscript{2,3,*}, MPH; Hongxia Li\textsuperscript{4}, MD; Dong Cai\textsuperscript{1}, BA; Huanbing Zhu\textsuperscript{5,*}, MD, PhD; Shengdong Pan\textsuperscript{6,*}, MPA, LLM

\textsuperscript{1}Medical Records Room, Department of Medical Administration, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China
\textsuperscript{2}Department of Medical Records, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China
\textsuperscript{3}Collaborating Center for the WHO Family of International Classifications in China, Beijing, China
\textsuperscript{4}Department of Gastroenterology, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China
\textsuperscript{5}Department of Hepatobiliary and Pancreatic Surgery, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China
\textsuperscript{6}Department of Medical Administration, The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China

*these authors contributed equally

Abstract

Background: The International Classification of Diseases, Eleventh Revision (ICD-11) improved neoplasm classification.
Objective: We aimed to study the alterations in the ICD-11 compared to the Chinese Clinical Modification of the International Classification of Diseases, Tenth Revision (ICD-10-CCM) for neoplasm classification and to provide evidence supporting the transition to the ICD-11.
Methods: We downloaded public data files from the World Health Organization and the National Health Commission of the People’s Republic of China. The ICD-10-CCM neoplasm codes were manually recoded with the ICD-11 coding tool, and an ICD-10-CCM/ICD-11 mapping table was generated. The existing files and the ICD-10-CCM/ICD-11 mapping table were used to compare the coding, classification, and expression features of neoplasms between the ICD-10-CCM and ICD-11.
Results: The ICD-11 coding structure for neoplasms has dramatically changed. It provides advantages in coding granularity, coding capacity, and expression flexibility. In total, 27.4\% (207/755) of ICD-10 codes and 38\% (1359/3576) of ICD-10-CCM codes underwent grouping changes, which was a significantly different change ($\chi^2_1=30.3; P<.001$). Notably, 67.8\% (2424/3576) of ICD-10-CCM codes could be fully represented by ICD-11 codes. Another 7\% (252/3576) could be fully described by uniform resource identifiers. The ICD-11 had a significant difference in expression ability among the 4 ICD-10-CCM groups ($\chi^2_3=93.7; P<.001$), as well as a considerable difference between the changed and unchanged groups ($\chi^2_1=74.7; P<.001$). Expression ability negatively correlated with grouping changes ($r=-.144; P<.001$). In the ICD-10-CCM/ICD-11 mapping table, 60.5\% (2164/3576) of codes were postcoordinated. The top 3 postcoordinated results were specific anatomy (1907/3576, 53.3\%), histopathology (201/3576, 5.6\%), and alternative severity 2 (70/3576, 2\%). The expression ability of postcoordination was not fully reflected.
Conclusions: The ICD-11 includes many improvements in neoplasm classification, especially the new coding system, improved expression ability, and good semantic interoperability. The transition to the ICD-11 will inevitably bring challenges for clinicians, coders, policy makers and IT technicians, and many preparations will be necessary.


KEYWORDS
Chinese Clinical Modification of the International Classification of Diseases, Tenth Revision; ICD-10; ICD-10-CCM; ICD-11; improvement; International Classification of Diseases, Eleventh Revision; International Classification of Diseases, Tenth Revision; International Classification of Diseases; neoplasm; transition

Introduction
The World Health Organization (WHO) adopted the International Classification of Diseases, Tenth Revision (ICD-10) in May 1990 [1]. The ICD-10 has been widely used in over 120 countries over the past 30 years. In multiple countries, expansions of the ICD-10, such as the Chinese Clinical Modification of the International Classification of Diseases, Tenth Revision (ICD-10-CCM), based on the second edition [2], have been developed to meet specific requirements. The ICD-10-CCM has been used for national performance assessment and medical insurance payment in public hospitals. Given the advances in medical knowledge and health information, the WHO started working on the International Classification of Diseases, Eleventh Revision (ICD-11) in 2007 [3], and it was expected to be implemented by WHO member countries starting in January 2022.

The most critical changes in the ICD-11 were the establishment of a semantic knowledge base and the reconstruction of the coding system. It introduced the Foundation Component, the Common Ontology, and linearization, as well as the new concepts of precoordination and postcoordination [3-5]. The chapter on neoplasms has also changed. It adds crucial morphology to precoordination, having the highest proportion (98.9%) and the most dimensions (3.5) of postcoordination. Cancer has been one of the top 3 causes of death in China since 2005 [6]; thus, the coding system of the ICD-11 will significantly impact cause of death reporting, cancer registration, and disease diagnosis records. Stakeholders need to have a good understanding of the classification of neoplasms in the ICD-11.

This study analyzes the changes between the ICD-10-CCM and ICD-11 in terms of coding features, classification features, and expression features in neoplasm classification and hopes to provide evidence supporting the transition in China.

Methods
Ethical Considerations
According to the Measures for Ethical Review of Human Life Science and Medical Research issued by the National Health Commission of the People's Republic of China, this study utilized public data and did not involve human subjects, and thus, the requirement of ethical permission was waived. All examples were constructed and neither correspond to real clinical cases nor to any datasets.

Materials
The following 3 existing, publicly available files were used in this paper: (1) ICD-10/ICD-11 mapping tables [7], (2) ICD-11 simple tabulation [8], and (3) the second revision of the ICD-10-CCM [9].

The first 2 files were downloaded from the WHO website, and the ICD-10-CCM was released by the National Health Commission of the People’s Republic of China.

Research Methods
The ICD-10-CCM neoplasm codes were manually recoded by the ICD-11 coding tool [10] based on the International Classification of Diseases, Eleventh Revision, Mortality and Morbidity Statistics (ICD-11 MMS) codes to generate an ICD-10-CCM/ICD-11 mapping table for neoplasms. The 3 existing files and the ICD-10-CCM/ICD-11 mapping table were used to analyze ICD-11 features, namely, the coding structure, coding capacity, grouping changes, expression ability, expression flexibility, and the expression of postcoordination in the neoplasm classification.

Mapping ICD-10-CCM to ICD-11 Codes
Due to the homology of the morphology section between the ICD-10 and ICD-11, mapping was performed for only the topography codes. To ensure the accuracy of the results, manual recoding was independently implemented by 2 authors who both had more than 10 years of coding experience and had received ICD-11 training. Inconsistent results were resolved by consulting a senior coder from the Collaborating Center for the WHO Family of International Classifications in China.

Statistical Standards
Leaf Codes
The ICD-10, ICD-10-CCM, and ICD-11 MMS codes that can be used at the lowest level are called leaf codes. All statistical analyses were based on the leaf codes.

Coding Capacity
This capacity involves the number of leaf codes that can be used in actual coding.

Grouping Changes
Grouping was based on the block structure of the ICD. The equivalent groups included ICD-10 group 1 (malignant neoplasms) and ICD-11 group 3 (malignant neoplasms, except...
for lymphoid, hematopoietic, central nervous system, or related tissues), ICD-10 group 2 (in situ neoplasms) and ICD-11 group 4 (in situ neoplasms, except for lymphoid, hematopoietic, central nervous system, or related tissues), and ICD-10 group 3 (benign neoplasms) and ICD-11 group 5 (benign neoplasms, except for lymphoid, hematopoietic, central nervous system, or related tissues). If an ICD-10/ICD-10-CCM leaf code was not classified into the equivalent ICD-11 group, it was considered to have undergone a grouping change.

Expression Ability

For each ICD-10-CCM code, we identified the best-matching ICD-11 MMS leaf code. When all the clinical details in the diagnosis were expressed without redundant information, the code was considered fully represented. This study also defined the synonyms contained in the Foundation Component as a full representation.

Statistical Analysis

All data were analyzed with SPSS (version 25.0, IBM). The changes in coding capacity, groups, and expression ability between the ICD-10 and ICD-11 were described as rates and percentages. The chi-square test was used to determine the difference in grouping changes when mapped to the ICD-11 between the ICD-10 and ICD-10-CCM. The chi-square test was also used to analyze the difference in full expression ability among the 4 ICD-10-CCM groups, and the Bonferroni method was used for pairwise comparisons. The chi-square test was used to analyze the difference in full expression ability between the changed and unchanged groups mapped to the ICD-11. The $\phi$ correlation coefficient was used to analyze the correlation between expression ability and grouping changes. Statistical significance was set at $P<.05$.

Results

Coding Features

Coding Structure

Chapter 2 in both the ICD-10 and ICD-11 addresses neoplasms, and the codes in both versions are alphanumeric but have different structures. The ICD-10 codes for neoplasms consist of topography codes and morphology codes. The topography codes range from C00 to D48. Except for leaf codes, the coding range, categories, and subcategories of the ICD-10-CCM are the same as those of the ICD-10. The ICD-10 morphology codes consist of 5 digits. The first 4 digits identify the histological type of the neoplasm, and the fifth digit, following a slash (/), indicates its behavior. In the ICD-10-CCM, the morphology codes consist of 6 digits, with a fifth number before the slash.

In contrast, the coding structure of the ICD-11 MMS codes has dramatically changed. It is composed of stem codes and extension codes that are connected by an ampersand (&). The precoordinated stem codes consist of sites and essential morphology types. The newly added chapter on extension codes addresses other morphology types and greater site specificity [11], as well as stage, grading, laterality, and the diagnostic method. These codes can be used for postcoordination. The stem codes range from 2A00 to 2F9Z. The extension codes for morphology are 6-digit codes composed of letters and numbers, starting with the letter X. Examples of the ICD-10 and ICD-11 complete neoplasm codes are shown in Table 1.

Table 1. Examples of the International Classification of Diseases, Tenth Revision (ICD-10), Chinese Clinical Modification of the International Classification of Diseases, Tenth Revision (ICD-10-CCM), and International Classification of Diseases, Eleventh Revision (ICD-11) complete neoplasm codes.

<table>
<thead>
<tr>
<th>Neoplasm</th>
<th>ICD-10</th>
<th>ICD-10-CCM</th>
<th>ICD-11</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adenocarcinoma of the common bile duct, stage III, diagnosis confirmed by histology</td>
<td>C24.0</td>
<td>8140/3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>C24.003</td>
<td>81400/3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2C15.0</td>
<td>&amp;XS6H</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>&amp;XY9Q</td>
</tr>
</tbody>
</table>

In the ICD-10, C24.0 represents a malignant neoplasm of the extrahepatic bile duct, and 8140/3 represents adenocarcinoma. In the ICD-10-CCM, C24.003 represents a malignant neoplasm of the common bile duct, and 81400/3 represents adenocarcinoma. In the ICD-11, 2C15.0 represents adenocarcinoma of the distal bile duct, XS6H represents stage III, and XY9Q represents a diagnosis confirmed by histology. We used the the ICD-11 coding tool website [10].

Coding Capacity

Categories, subcategories, and leaf codes can be used for statistics, but only leaf codes can be used for actual coding. In the ICD-10, chapter 2 includes 759 leaf codes [7]. In the ICD-10-CCM, the topography codes are basically expanded by refining the sites, with 3634 leaf codes for neoplasms. Taking malignant neoplasm of the nasal cavity (C30.0) as an example, the ICD-10-CCM contains 5 additional leaf codes, such as a malignant neoplasm of nasal cartilages (C30.001) and a malignant neoplasm of the nasal concha (C30.002).

In the ICD-11, chapter 2 includes 1037 leaf codes [8]. Compared with the ICD-10, the number of ICD-11 leaf codes was expanded by 36.6%, which is conducive to better granularity of statistics and classification.

Classification Feature

The ICD-11 has readjusted the neoplasm groups. In the ICD-10, chapter 2 was divided into 4 groups: malignant neoplasms, in situ neoplasms, benign neoplasms, and neoplasms of uncertain or unknown behavior. In the ICD-11, chapter 2 was increased to 7 groups (Table 2). The newly added ICD-11 group 1 includes all neoplasms of the brain and central nervous system, regardless of behavior, and the newly added ICD-11 group 2 includes all hematopoietic and lymphoid tissues. The ICD-10 group 4 was split into 2 separate groups: ICD-11 group 6 (neoplasms of
uncertain behavior) and ICD-11 group 7 (neoplasms of unknown behavior).

Table 2. Grouping changes between the International Classification of Diseases, Tenth Revision (ICD-10) and the International Classification of Diseases, Eleventh Revision (ICD-11).

<table>
<thead>
<tr>
<th>Classification system</th>
<th>ICD-11</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
<th>Group 5</th>
<th>Group 6</th>
<th>Group 7</th>
<th>Other chapters</th>
<th>No mapping</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICD-10 codes, n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 1</td>
<td>23a</td>
<td>91a</td>
<td>352</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>8a</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Group 2</td>
<td>N/A</td>
<td>N/A</td>
<td>57</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 3</td>
<td>N/A</td>
<td>N/A</td>
<td>1a</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>1a</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 4</td>
<td>11a</td>
<td>18a</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>48a</td>
<td>51a</td>
<td>2a</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>ICD-10-CCM codes, n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 1</td>
<td>93a</td>
<td>499a</td>
<td>1252</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>17a</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Group 2</td>
<td>N/A</td>
<td>N/A</td>
<td>169</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 3</td>
<td>118a</td>
<td>N/A</td>
<td>N/A</td>
<td>1a</td>
<td>N/A</td>
<td>796</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 4</td>
<td>85a</td>
<td>60a</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>316a</td>
<td>149a</td>
<td>18a</td>
<td></td>
<td>55</td>
</tr>
</tbody>
</table>

*a*Indicates grouping changes. The ICD-10 and ICD-10-CCM no mapping codes were not included in the statistical analysis of this study. Hence, 755 ICD-10 codes and 3576 ICD-10-CCM codes were used for percentage and chi-square analyses.

In total, 27.4% (207/755) of ICD-10 codes underwent grouping changes. Among them, 150 codes were migrated by 1 group, 56 by 2 groups, and one by 3 groups. In the ICD-10-CCM, 38% (1359/3576) leaf codes underwent grouping changes (Table 2). A chi-square test revealed significant differences in grouping changes between the ICD-10 and ICD-10-CCM ($\chi^2_F=30.3; P<.001$).

### Expression Features

#### Expression Ability

Because of the classification changes, 58 of the 3634 ICD-10-CCM codes for neoplasms could not be recoded. In total, 3576 codes were included in the manual recoding study. The results of 2 separate recodings showed that 6% (213/3576) codes were mapped inconsistently. A total of 32 stem codes were inconsistent, and 181 extension codes differed. The 213 codes were all identified after consultation with the senior coder from the Collaborating Center for the WHO Family of International Classifications in China.

The final results showed that 16.6% (594/3576) codes were fully represented without postcoordination, 51.2% (1830/3576) codes were fully characterized with postcoordination, and the remaining 32.2% (1152/3576) codes were only partially described (Table 3). In addition, 7% (252/3576) codes were fully represented when using uniform resource identifiers (URIs).

### Table 3. Comparison of the International Classification of Diseases, Eleventh Revision (ICD-11) expression ability among the 4 Chinese Clinical Modification of the International Classification of Diseases, Tenth Revision (ICD-10-CCM) groups.

<table>
<thead>
<tr>
<th>ICD-10-CCM group</th>
<th>Total codes, N</th>
<th>Full representation</th>
<th>Coded without PC (b), n</th>
<th>Coded with PC (a), n</th>
<th>Partial representation</th>
<th>Codes (n1=N), %</th>
<th>Codes (n2=N), %</th>
<th>Codes (n1=n2), %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>1861</td>
<td>935</td>
<td>406</td>
<td>1341</td>
<td>72.1</td>
<td>520</td>
<td>27.9</td>
<td></td>
</tr>
<tr>
<td>Group 2</td>
<td>169</td>
<td>106</td>
<td>44</td>
<td>150</td>
<td>88.8</td>
<td>19</td>
<td>11.2</td>
<td></td>
</tr>
<tr>
<td>Group 3</td>
<td>918</td>
<td>501</td>
<td>73</td>
<td>574</td>
<td>62.5</td>
<td>344</td>
<td>37.5</td>
<td></td>
</tr>
<tr>
<td>Group 4</td>
<td>628</td>
<td>288</td>
<td>71</td>
<td>359</td>
<td>57.2</td>
<td>269</td>
<td>42.8</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>3576</td>
<td>1830</td>
<td>594</td>
<td>2424</td>
<td>67.8</td>
<td>1152</td>
<td>32.2</td>
<td></td>
</tr>
</tbody>
</table>

*PC: postcoordination.

Based on the 4 ICD-10-CCM groups, ICD-11 significantly differed in expression ability ($\chi^2_F=93.7; P<.001$). The Bonferroni method showed that the ICD-11 had the most robust expression ability in the ICD-10-CCM group 2, followed by the ICD-10-CCM group 1, and there was no significant difference between the remaining 2 groups. Based on whether grouping changes occurred during mapping, 73.1% (1620/2217) ICD-10-CCM codes in the unchanged group.
and 59.2% (804/1359) ICD-10-CCM codes in the changed group were fully expressed, and the difference was significant ($\chi^2 = 74.7; P < .001$). Expression ability had a negative correlation with grouping changes ($r = -.144; P < .001$).

**Expression Flexibility**

The expression flexibility of the ICD-11 is reflected in many aspects. For instance, in the ICD-10, the subcategory “.” generally describes overlapping neoplasm sites, while in the ICD-11, several methods are used. Specifically, the ICD-11 uses multiple extension codes. Sometimes, only 1 extension code is used, such as XA4YW8 (overlapping sites of the esophagus). Occasionally, the ICD-11 describes this condition through stem codes, such as 2B71.0 (adenocarcinoma of the esophago gastric junction). ICD-11 classifications can also be represented through URIs, for instance, http://id.who.int/icd/entity/419755630 (Kaposi sarcoma of multiple organs).

Expression flexibility is also reflected in the additional option of postcoordination, which can meet the different requirements of most hospitals for clinical phenotype mining. For example, a patient with hepatocellular carcinoma in the left lobe of the liver that causes chronic intermittent cancer pain and tumor anemia would be coded as 2C12.02&XA5766/MG30.10&XT5Gi/3A71.0. Postcoordination fully expresses the clinical phenotypes and demonstrates the relationship between diseases and clinical phenotypes.

**Expression of Postcoordination**

According to the ICD-10-CCM/ICD-11 mapping table, 60.5% (2164/3576) of codes had postcoordination, and the average amount of postcoordination per code was approximately 0.7. Regarding the type of postcoordination, the proportion of specific anatomy (1907/3576, 53.3%) was the highest, followed by histopathology (201/3576, 5.6%), and the remaining dimensions were basically less than 2% (Table 4).

<table>
<thead>
<tr>
<th>Number and dimensions of postcoordination</th>
<th>Codes, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1412 (39.5)</td>
</tr>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Specific anatomy</td>
<td>1750 (48.9)</td>
</tr>
<tr>
<td>Histopathology</td>
<td>176 (4.9)</td>
</tr>
<tr>
<td>Alternative severity 2</td>
<td>60 (1.7)</td>
</tr>
<tr>
<td>Laterality</td>
<td>5 (0.1)</td>
</tr>
<tr>
<td>Course of the condition</td>
<td>3 (0.1)</td>
</tr>
<tr>
<td>Stem codes</td>
<td>3 (0.1)</td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Specific anatomy/specific anatomy</td>
<td>118 (3.3)</td>
</tr>
<tr>
<td>Specific anatomy/histopathology</td>
<td>17 (0.5)</td>
</tr>
<tr>
<td>Specific anatomy/laterality</td>
<td>8 (0.2)</td>
</tr>
<tr>
<td>Histopathology/alternative severity 2</td>
<td>8 (0.2)</td>
</tr>
<tr>
<td>Alternative severity 2/stem codes</td>
<td>2 (0.1)</td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Specific anatomy/specific anatomy/specific anatomy</td>
<td>11 (0.3)</td>
</tr>
<tr>
<td>Specific anatomy/specific anatomy/laterality</td>
<td>3 (0.1)</td>
</tr>
</tbody>
</table>

**Discussion**

**Principal Findings**

The different coding structure of the ICD-11 provides advantages in terms of coding granularity, coding capacity, and expression flexibility. According to the mapping tables, the grouping changes between the ICD-10 and ICD-10-CCM differed ($\chi^2 = 30.3; P < .001$). Meanwhile, neither the ICD-10 group 2 nor the ICD-10-CCM group 2 exhibited group migration when mapped to the ICD-11. In this study, ICD-11 expression ability (67.8%) was slightly higher than that in other studies (60%) [12,13]. Among the 4 ICD-10-CCM groups, there were significant differences in expression ability ($\chi^2 = 93.7; P < .001$), with ICD-10-CCM group 2 having the highest expression ability. The expression ability negatively correlated with grouping changes ($r = -.144; P < .001$).

The expression ability of the ICD-11 was still underestimated. There were many reasons why 1152 ICD-10-CCM codes did not have full representation, some of which can be avoided in actual cases. First, of the 296 ICD-10-CCM codes mapped to the ICD-11 group 1, only 26 (8.8%) codes were fully expressed.
which is far below the overall level. Among them, 217 codes cannot be fully expressed because most of the stem codes in ICD-11 group 1 do not include behavior. However, in coding actual cases with morphological types, some stem codes, including behavior, would be used instead of the residual category. Therefore, in actual coding, the expression ability of this group would be higher than that determined in this study.

Second, 714 codes could not be fully expressed due to site classification. Fortunately, 109 codes detailing specific sites may be resolved by refining the value set of extension codes. Examples include the frenulum of the upper lip, the ileocecal valve, and the rectouterine recess. However, there were also some codes for which the classification was different, such as peripheral nervous system neoplasms. This condition could be solved by coding the actual cases. In addition, some words, such as canceration of the gastric stump and cervical stump, need to be addressed.

Third, URIs can supplement the function of the ICD-11 MMS codes, especially for some diseases that cannot be identified by the ICD-11 codes. For instance, URIs (http://id.who.int/icd/entity/1595913346) make classic Kaposi sarcoma classifiable. The new coding system can also make other diseases identifiable [14-16], such as chronic pain and rare diseases. In summary, the ICD-11 has advantages in terms of actual coding and can address the expression needs of neoplasms.

The ICD-11 vs the ICD-10

Compared to the ICD-10, the ICD-11 has undergone significant changes from design to use. First, the design purpose of the ICD-11 has changed. For more than a century, the ICD has been the basis for comparable statistics on causes of mortality and morbidity between places and over time. As a statistically friendly classification system, when the ICD-10 is used for clinical term records and diagnosis-related grouping, it cannot meet practical needs, resulting in various expanded versions of the WHO ICD-10 in multiple countries, which often leads to inconsistent statistical standards for data. In contrast, the ICD-11 is a clinically friendly classification system that meets diverse goals beyond mere health statistics, including clinical term records, patient safety and quality, reimbursement, decision support, and more. The ICD-11 is entirely digital, terminology is coded with the coding tool and application programming interface, and it has a semantic knowledge base; these features are beneficial for standardized data collection. In short, the ICD-11 has benefits in terms of obtaining statistics as well as multiaxial coding, coding granularity [4,17,18] and standardization, achieving the integration of terminology and classification. As shown in Table 1, although the ICD-10-CCM codes are expanded and refined at the leaf code level, the capture of clinical details is still weaker compared to the ICD-11 codes.

Second, the design concept of the ICD-11 is different. The ICD-10 is an independent classification system, and the standard terminology set for the ICD-10 has been developed independently. Differently, the ICD-11 has good semantic interoperability with other classification systems through harmonized methods. The Systematized Nomenclature of Medicine-Clinical Terms (SNOMED CT) is considered one of the most comprehensive clinical terminologies in the world [19]. The ICD-11 Foundation Component, which includes semantic network concepts and their relationships, is organized around the Common Ontology from a subset of the SNOMED CT [20,21]. The Common Ontology has been harmonized with ICD text definitions, primarily from the SNOMED CT clinical findings hierarchy (findings, disorders, and disease) and secondarily from other hierarchies (situations, events, social context, and so on) [22]. The rich Foundation Component has approximately 80,000 entries and 40,000 synonyms [5]. The ICD-11 linearizations, including the ICD-11 MMS and the International Classification of Diseases for Oncology (ICD-O), are subsets derived from the Foundation Component. The ICD-11 integrates the morphology section of the ICD-O, ICD-O linearization, and tumor node metastasis classification, and the histopathology codes of the ICD-11 are also compatible with the ICD-O [4]. Additionally, the ICD-11 integrates numerous clinical terminologies from some expanded versions of the WHO ICD-10. Compared to other classifications, the ICD-11 has advantages in terms of concept coverage and compatibility.

Third, the ICD-11 has stronger logical links between codes. If multiple codes are needed for disease expression, there is a lack of practical connections between ICD-10 codes. In the ICD-11, precoordination contains the site and morphology, and the postcoordination of morphology and the clinical phenotype is linked to stem codes through an ampersand (&) and slash (/). In addition, the ICD-11 contains rich dimensions of postcoordination. Compared to the ICD-10, which provides information only about topography and morphology, the ICD-11 can also include other dimensions of postcoordination, such as stage, grading, laterality, and the diagnostic method. The ICD-11 coding system is more conducive to unifying national cancer registration in these dimensions. However, due to the limitations of the ICD-10 coding structure, these dimensions of postcoordination were missing when mapped to the ICD-11, with only 0.7 postcoordination per code, and its expression ability cannot be truly reflected.

Moreover, the ICD-11 has more expression flexibility. The ICD-10 coding system is fixed and single. The ICD-11 provides 2 sets of codes: ICD-11 MMS codes and URIs. In ICD-11 MMS codes, a stem code can be used alone or with optional extension codes. A URI is a string of characters that uniquely identifies a particular entity. Therefore, a coding system with good expression flexibility can meet the coding granularity and clinical phenotype mining requirements of hospitals at different levels.

Finally, the ICD-11 has different update mechanisms. The ICD-10 updates relatively slowly. In 1999, the WHO established the Update Advisory Committee as the only authoritative body revising the ICD-10. In 2005, the WHO officially published the second edition of the ICD-10 and continued to issue revisions to the relevant content of the ICD-10 through official channels every year. The ICD-11 provides a web-based coding tool instead of paper environments, which is highly beneficial for timely updates. It has a real-time updated orange version and an annually updated blue version.
Overall, the eleventh revision is more extensive than any other revision since the sixth in 1948. These design-level features make the ICD-11 competitive with other classification systems.

Challenges of Replacement

The improvements described above also pose challenges for replacing the ICD-10 with the ICD-11. Compared to the ICD-10, the ICD-11 has more leaf codes for neoplasms. New extended codes have been added to the ICD-11, and there are also approximately 16,000 extension codes. Hence, the expression ability of the ICD-11 is much greater than that of the ICD-10 and ICD-10-CCM. However, codes with better granularity require clinicians to record diagnoses and treatments in greater detail. It is essential to carry out the necessary interventions to enhance medical record documentation according to ICD-11 before or simultaneously with country-wide implementation [23]. Clinicians will be required to have a good understanding of the ICD-11. Otherwise, the best classification system is just a decoration.

Second, coders are accustomed to using the old classification system and need time to familiarize themselves with the new coding system and tool. Studies have shown that some coders have difficulties coding on the web due to a lack of familiarity with software [24,25]. In this study, 2 skilled coders who received ICD-11 training still had 215 inconsistent codes. Web-based coding requires high professional ability. Studies have also shown that the reliability and accuracy of ICD-11 coding are lower than those of ICD-10 coding [23-25]. In the early stages of ICD-11 promotion, countries involved in the transition need to conduct ICD-11 training and transform coders’ coding concepts. To maintain coding accuracy, developing high-quality ICD-11 training materials, training processes, and local ICD-11 guidelines is effective [23,26].

Moreover, many countries have applied the ICD-10 in different fields, such as mortality and morbidity statistics, diagnosis-related grouping, and cancer registries. There is also a potential challenge with IT systems being able to adopt a new classification that allows for unlimited width in fields. Stakeholders must upgrade their information systems to meet the needs of the ICD-11, which requires a significant amount of time and money to hire medical, IT, and management personnel to complete this transformation. The experience of the United States in replacing ICD versions can provide a better reference. In the United States, because of the significant opposition and reservations expressed by stakeholders, it took 6 years from the adoption of the International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) to complete the transition [27], and the costs associated with the transition were estimated to be between US $475 million and US $1.5 billion dollars, including training, productivity losses, and system changes [28].

Importantly, the ICD-10 and ICD-11 will coexist for a long time before replacement, which contributes to solving some problems. The parallel implementation of the 2 classification systems can ensure government policy continuity. In January 2022, the pilot application program to promote ICD-11 was launched in China, with 59 large general public hospitals from all provinces participating. However, ICD-11 is not fully understood by other hospitals. The impacts of the new classification system implementation included coding accuracy, ICD version mapping, and more [29]. The ICD version mapping can be used for interoperability between coded data sets [30]. Hospitals do not require dual coding and use ICD-11 to report data. If a hospital needs dual coding, a mapping table can significantly reduce the time, at least by half, spent on duplicate coding. A map can also maintain the same accuracy of ICD-11 coding for all hospitals and alert coders if there are any coding errors. Usually, a high-quality map requires substantial manual curation, and some studies use algorithmic mapping approaches, such as sequential mapping [30]. Some scholars have developed hybrid methods, such as automatic mapping and manual review [31], semiautomatic mapping, and manual evaluation [32]. Although automatic coding has many benefits, supervised and manual mapping are still necessary [33]. Undeniably, there may be some issues with the use of mapping tables. A study has shown that 14.5% of International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes used by internists, when mapped to the ICD-10-CM, resulted in potential clinical inaccuracies [34]. In general, a high-quality map would contribute to the smooth transition of ICD-11, and multiple mapping methods can be explored. This study confirms the feasibility of using manual mapping tables for neoplasms between the ICD-10-CCM and ICD-11. To date, the results of some studies on the transition to the ICD-11 have also been optimistic [17,18].

In addition, although the WHO provides an ICD-11 coding tool, countries must continue optimizing this tool. A mapping table can improve the intelligence of the coding tool. If the terms are the same as those in the mapping table, the coding tool can automatically load the coding cluster without selecting postcoordination one by one. The accuracy and efficiency of ICD-11 coding largely depend on how well the coding tool is optimized. It is necessary to continuously optimize the coding tool before fully promoting the ICD-11.

Limitations

This study had some limitations that must be considered. First, this study focused only on the clinical condition of neoplasms. However, the chapter on neoplasms is quite complex and one of the most varied chapters on other conditions, with significant changes. Moreover, patients with cancer may experience different manifestations, including chronic, surgical, and emergency conditions. For example, a patient with long-term chronic hepatitis B causing cirrhosis and liver cancer was admitted to the hospital for emergency surgery due to a tumor rupture. Furthermore, the burden of cancer in China continues to grow with the aging population. According to statistics, the crude cancer mortality rate increased from 108.3 per 100,000 individuals in the 1990-1992 period to 170.1 per 100,000 individuals in 2015 [35]. Significantly, the research method is well thought out. Manual mapping can provide a good understanding of the new ICD-11 features and help individuals familiarize themselves with the characteristics of the new classification systems. As a method, manual mapping can be extended to other conditions. More conditions can be gradually incorporated, especially emergency and surgical conditions, in the future.
Second, this study used only manual mapping, which is time-consuming and laborious. Other methods combined with manual mapping can be continuously explored, especially automatic mapping between the ICD-10-CCM and ICD-11, including algorithmic mapping and machine learning approaches.

**Conclusion**

Neoplasm classification has undergone many improvements in the ICD-11, especially the new coding system, improved expression ability, and good semantic interoperability. The new coding system provides advantages in coding granularity, coding capacity, and expression flexibility. Moreover, 67.8% of ICD-10-CCM neoplasm codes can be fully represented by the ICD-11, and expression ability negatively correlates with grouping changes ($r$=-0.144, $P$<.001). The more significant the changes in a new classification system are, the less information can be expressed when mapped to other databases. The use of URIs and maintenance mechanisms can increase the expression ability for coding in actual cases. In addition, the good semantic interoperability of the ICD-11, integrating numerous clinical terminologies from the SNOMED CT, the ICD-O, the tumor node metastasis classification, and expanded versions of the WHO ICD-10, makes the ICD-11 competitive with other classification systems. The transition to the ICD-11 will inevitably bring numerous challenges for clinicians, coders, policy makers and IT technicians, and many preparations will be necessary.

**Acknowledgments**

The authors would like to thank Dr Xiaoding Liu from the Department of Pathology at Peking Union Medical College Hospital for providing information about pathology and answering our questions. This work was supported by the Fundamental Research Funds for the Central Universities (grant 2022FZZX06-09), and the Zhejiang Kangenbei Hospital-managed soft science research project (grant 2023ZHA-KEB305).

**Conflicts of Interest**

None declared.

**References**


Abbreviations

**ICD-10**: International Classification of Diseases, Tenth Revision
**ICD-10-CCM**: Chinese Clinical Modification of the International Classification of Diseases, Tenth Revision
**ICD-10-CM**: International Classification of Diseases, Tenth Revision, Clinical Modification
**ICD-11 MMS**: International Classification of Diseases, Eleventh Revision, Mortality and Morbidity Statistics
**ICD-11**: International Classification of Diseases, Eleventh Revision
**ICD-9-CM**: International Classification of Diseases, Ninth Revision, Clinical Modification
**ICD-O**: International Classification of Diseases for Oncology
**SNOMED CT**: Systematized Nomenclature of Medicine-Clinical Terms
**URI**: uniform resource identifier
**WHO**: World Health Organization
Behavioral Insights from Vaccine Adoption in Nigeria: Cross-Sectional Survey Findings

Sohail Agha1,2, BA, PhD; Ifeanyi Nsofor3, MSc, MD; Drew Bernard2, BA; Sarah Francis2, BA; Nandan Rao4, MSc

1Behavior Design Lab, Stanford University, Stanford, CA, United States
2Behavioral Insights Lab, Seattle, WA, United States
3Centre for Family Health Initiative, Abuja, Nigeria
4Virtual Labs, Corvallis, OR, United States

Corresponding Author:
Sohail Agha, BA, PhD
Behavior Design Lab
Stanford University
581 Capistrano Way
Stanford, CA, 94305
United States
Phone: 1 1 206 351 9346
Email: sohailagha@gmail.com

Abstract

Background: To generate behavioral insights for the development of effective vaccination interventions, we need approaches that combine rapid and inexpensive survey data collection with instruments based on easy-to-use behavior models. This study demonstrates how an inexpensive digital survey helped identify the drivers of COVID-19 vaccination in Nigeria.

Objective: This study aims to illustrate how behavioral insights can be generated through inexpensive digital surveys.

Methods: We designed and conducted a cross-sectional survey with multistage sampling. Data were collected from Nigerians (aged ≥18 years) from 120 strata based on age, sex, state, and urban or rural location. Respondents were recruited via advertisements on Meta platforms (Facebook and Instagram) using the Virtual Lab open-source tool. We used a Meta Messenger chatbot for data collection; participants were compensated with 400 naira (US $0.87 cents). Data collection took 2 weeks. In total, 957 respondents completed the survey, at an advertising cost of US $1.55 per respondent. An 18-item instrument measuring core motivators, ability barriers, sociodemographic characteristics, and respondents’ vaccination status was pretested before data collection. We ran separate logistic regression models to examine the relationships between vaccine uptake and core motivators, ability barriers, and sociodemographic variables. A final model that predicted vaccine uptake included all 3 sets of variables.

Results: About 56% (n=540) of respondents reported that they had received at least 1 COVID-19 vaccination. Three core motivators were positively associated with vaccine uptake: the belief that the COVID-19 vaccine promised a better life (adjusted odds ratio [aOR] 3.51, 95% CI 2.23-5.52), the belief that the vaccine would allow respondents to do more things they enjoyed (aOR 1.97, 95% CI 1.33-2.93), and respondents’ perception that their friends and family members accepted their decision to get vaccinated (aOR 1.62, 95% CI 1.06-2.48). Two ability barriers were negatively associated with vaccine uptake: cost- or income-related concerns lowered the odds of being vaccinated (aOR 0.35, 95% CI 0.24-0.50) and the lack of availability of vaccines at places respondents routinely visited also lowered their odds of being vaccinated (aOR 0.29, 95% CI 0.21-0.40). After adjusting for other variables, the perceived fear of getting COVID-19 and the hardship associated with the disease were no longer associated with vaccine uptake.

Conclusions: These findings suggest that hope is more important for Nigerians than fear when it comes to vaccine adoption, enjoying life is more important than worrying about getting the disease, and approval from friends and family is more powerful than their disapproval. These findings suggest that emphasizing the benefits of leading a fuller life after being vaccinated is more likely to succeed than increasing Nigerians’ fear of COVID-19. This study identifies a very different set of factors associated with COVID-19 vaccine adoption than previous Nigerian studies.

doi:10.2196/47817
KEYWORDS
behavioral insights; COVID-19; Nigeria; surveys; vaccination

Introduction

Background
Globally, the COVID-19 pandemic has highlighted the importance of behavioral insights for increasing the use of preventive behaviors such as wearing a mask, social distancing, and getting vaccinated. However, recent studies on COVID-19 vaccination emerging from Nigeria and other low- and middle-income countries (LMICs) have primarily focused on identifying gaps in knowledge, attitudes, and beliefs associated with vaccine hesitancy. A number of these studies recommend educating health care workers (HCWs) and members of the general population on vaccine safety and efficacy and assume that equipping people with factually correct information will allay their concerns, increase their perceived risk of acquiring COVID-19, and lead to higher rates of vaccine adoption [1,2].

Data from other Nigerian studies, however, raise questions regarding the strength of the relationship between risk perception and vaccine acceptance. A recent hospital-based study in southern Nigeria found that, while most HCWs perceived themselves at risk of COVID-19, only about half were willing to receive the COVID-19 vaccine [3]. Substantial gaps between risk perception and willingness to accept a COVID-19 vaccine have been observed in the general population in northern Nigeria as well [4]. A study that interviewed over 5000 respondents across all states in Nigeria found that COVID-19 was not perceived as a threat by most respondents [5].

What seems to be an important determinant of vaccine acceptability in Nigeria is trust in the vaccine manufacturing process, health system, government [6], and institutions involved in risk communication on behalf of the government [7]. Yet, despite multiple studies showing a weak relationship between risk perception and willingness to adopt a COVID-19 vaccine in Nigeria, researchers continue to recommend the provision of factually correct information to fill “information gaps” [4].

A recent systematic review of the COVID-19 literature in Nigeria shows that vaccination rates among those at high risk of COVID-19, such as HCWs, were lower than among those at low risk of COVID-19 [8]. Given the weak relationship between risk perception and vaccine uptake in Nigeria, it is not surprising that a recent evaluation found that risk communication efforts in Nigeria were inadequate in sustaining changes in behavior observed at the beginning of the pandemic [9]. Lawal [9] showed that during the first 30 days of the discovery of COVID-19 in Nigeria, and until the national lockdown, public interest in learning about the disease surged. Visits to public places such as grocery stores declined during this period as stringent government policies resulted in reduced mobility of the population. The study by Lawal [9] found that, as the Nigerian population started becoming aware of the disease, there was a slight decline in the number of COVID-19 cases. However, this decline occurred for a relatively short period of time. The number of new cases started increasing again as the initial effects of risk communication interventions dissipated. Lawal [9] concluded that Nigerians listened to messages telling them to take preventive measures such as social distancing or masking for some time but eventually got tired of the messages and stopped responding to them. In part, this was because the recommended public health precautions did not fit well in the context in which they lived their lives [9].

A clear picture of the drivers of COVID-19 vaccine hesitancy does not emerge from the recent public health literature on Nigeria, in part because much of this work is not based on a behavioral framework. The importance of theory-based work to understand the drivers of vaccine acceptance and design appropriate interventions has been emphasized [10]. In the absence of a clearly articulated framework for understanding vaccine-related behavior, it is difficult to interpret the findings of individual studies and arrive at a clear picture of what drives vaccine uptake in Nigeria.

As a result, there is very limited guidance available to support Nigerian practitioners in designing interventions that might accelerate COVID-19 vaccine uptake. For example, although many recent studies emphasize the importance of implementing health promotion interventions or increasing HCWs’ ability to communicate more effectively with members of the general public, most of these studies do not provide any guidance on what the context of this communication should be or what strategy should be used to persuade adults to get vaccinated. Thus, the available research is at a standstill in terms of providing insights that would help in designing more effective behavioral interventions to accelerate vaccine uptake in Nigeria.

An important reason that the literature does not provide a clear direction for the design of behavioral interventions is the lack of use of behavioral frameworks in explaining vaccine acceptance and uptake. Of the more than 20 peer-reviewed publications reviewed for this paper, we found only 1 that used a behavior model to interpret its findings [5]. This is not surprising as researchers have noted the limited use of behavioral frameworks in public health research and practice for over a decade [11-14].

The need for a behavior model that can be used to explain the vaccine adoption process in simple terms that resonate with practitioners is urgent. While a broader discussion of what a practitioner-friendly behavior model should comprise of is merited, the characteristics of such a model have been proposed [15]. A minimum criterion should be that the use of the model leads to deliberate programmatic decisions, a greater emphasis on strengthening activities supported by behavioral research findings, and the elimination of activities that are not evidence based.

The practitioner-friendly model used in this study, the Fogg Behavior Model (FBM), was introduced in the public health literature in 2019 to explain the effects of a social marketing behavior change campaign on the adoption of condoms by married men in Pakistan [13]. More recently, it has been used to identify behavioral drivers associated with the (1) adoption...
of COVID-19 vaccination by a low-income population in Cote d’Ivoire [16], (2) adoption of iron folate by pregnant women in India [15], (3) uptake of COVID-19 vaccination by HCWs in Nigeria [17,18], and (4) use of contraception by adolescent girls and young women in Nigeria [19]. A recent study also demonstrates the use of the FBM in making timely programmatic adjustments to a contraceptive social marketing intervention implemented in Nigeria [20]. To the best of our knowledge, this is the first time that the FBM is being applied to understand the dynamics of COVID-19 vaccine adoption in the general population of Nigeria.

FBM: Motivation and Ability as the Drivers of Behavior

The FBM is a model developed for use by practitioners to understand the drivers of human behavior and assist them in the design of behavior change interventions. Fogg states that behavior happens when motivation, ability, and a prompt happen at the same moment. As shown in Figure 1, the model can be visualized in 2 dimensions. Figure 1 shows motivation along the y-axis and ability along the x-axis. Motivation ranges from high to low for any behavior. Ability also ranges from high to low for any behavior. For simplicity, we describe a behavior as being easy to do or hard to do. For a prompt to work, a person needs to have sufficient motivation and ability. The motivation-ability threshold is reflected by an action line in the FBM. Behavior occurs when a person whose motivation and ability are above the action line is prompted. The prompt does not work if the person does not have sufficient motivation to undertake the behavior and finds the behavior hard to do, that is, they are below the action line [21].

**Methods**

**Questionnaire Design: Core Motivators**

Fogg defines motivation as having 3 components: anticipation, sensation, and belonging. Anticipation reflects the hopes and fears a person associates with a behavior. Sensation reflects the pleasure or pain a person associates with a behavior. Belonging is reflected by the acceptance or rejection of the behavior by people whose opinions a person considers important.

We conducted a review of the literature to identify relevant constructs and appropriate measures of motivation and ability. The survey questionnaire (Multimedia Appendix 1) was designed using instruments developed to test motivation and ability constructs in the FBM. These instruments had been previously tested in the general population in Cote d’Ivoire and with HCWs in Nigeria and had shown acceptable levels of reliability [16,18]. Respondents were allowed to answer how strongly they agreed or disagreed with 6-point Likert scale items measuring core motivators, such as hope, fear, pleasure, pain, acceptance, and rejection, associated with the adoption of the COVID-19 vaccine.

**Ability Barriers**

Fogg defines ability in terms of 5 barriers: time, money, the physical effort required to adopt a behavior, the mental effort required to adopt a behavior, and whether the behavior fits into the person’s routine [21]. Fogg initially considered social
deviance (or social norms) as part of ability but did not include social norms as an ability barrier in later iterations of the model [22], possibly because of the complexity of the relationship between norms and behavior. To capture ability barriers, respondents could answer how strongly they agreed with items on a 5-point Likert scale. These items captured the 5 ability factors related to COVID-19 vaccine uptake: time, money, mental effort, physical effort, and routine. The instrument also contained questions on sociodemographic variables. In total, the instrument comprised 18 questionnaire items.

The instrument was pretested twice in Nigeria with samples of approximately 100 respondents. The first pretest showed that the relationship among variables measuring motivation, ability, and COVID-19 vaccine uptake was in the expected direction, with 1 exception. The variable measuring agreement or disagreement with the question “Many of my friends and family would think poorly of me if they knew I had taken the COVID-19 vaccine” did not demonstrate the expected relationship with vaccine uptake. This was replaced with the statement “Most people I know have obtained the COVID-19 vaccination.” The second pretest showed that the latter response was associated with the outcome in the expected direction. All other items included in the instrument demonstrated the expected relationships with vaccine uptake in both pretests.

Survey Design and Sampling

We implemented a cross-sectional survey with multistage sampling. Nigeria is a large and diverse country with 36 states and the Federal Capital Territory of Abuja. These states and the Federal Capital Territory are grouped into 6 geopolitical zones: northwest, northeast, north central, southwest, southeast, and south-south. Although this survey was not designed to be a representative survey of Nigeria, we aimed to capture the diversity of the country’s population by ensuring that respondents from 1 state within each geopolitical zone were sampled. Thus, the states were selected at the first stage. Sokoto state was selected from the northwest, Bauchi state from the northeast, Niger state from north central, Lagos state from the southwest, Anambra state from the southeast, and Rivers state from the south-south. The state selected within each zone reflects its socioeconomic, religious, and ethnic diversity. In addition, these 6 states had ongoing COVID-19 vaccination campaigns to help ensure that survey findings would be useful for COVID-19 program managers in Nigeria.

The sample was stratified by 6 states, 5 age groups, male and female sex, and urban or rural location. This resulted in 120 strata from which respondents were sampled. We ran a total of 120 different ad sets targeting respondents based on the characteristics mentioned above.

Respondents were recruited via advertisements on the Meta digital ad platforms (Facebook and Instagram) using the Virtual Lab open-source tool [23]. The Virtual Lab tool ran ads targeting respondents in all 120 strata. We used a Meta Messenger chatbot for the survey data collection, compensating respondents who completed the survey with 400 naira (US $0.87 cents) in mobile phone credit. Respondents who clicked on the ads were directed to the messenger bot. Of the 214,335 male and female respondents who reached through the ads, 3660 clicked on the link, 1367 started the survey, and 1011 answered most survey questions.

Respondents could complete the survey in one go or start the survey, stop, and return to complete it later. In total, 957 respondents answered all questions in the survey at an advertising cost of US $1.55 per person. Data from these 957 respondents are used for the analysis.

The outcome of interest for this study was having received at least 1 COVID-19 vaccination. The Government of Nigeria’s data on the number of vaccinations provided in 2022 suggested that 50% of Nigerians had received at least 1 COVID-19 vaccination. A sample size calculation was made using an outcome value of 50% and a design effect of 1.5. We estimated that a sample size of 900 would provide a margin of error of 4 percentage points on the outcome of interest.

Statistical Analysis: Relationships Among Core Motivators, Ability Barriers, and Vaccine Uptake

Univariate analysis was conducted to provide the frequency distributions of core motivators, ability barriers, and sociodemographic characteristics of the sample. Bivariate analysis was conducted to explore the relationships between core motivators, ability factors, sociodemographic characteristics, and having received at least 1 COVID-19 vaccination. Multivariate logistic regression analysis was conducted to identify which core motivators, ability factors, and sample characteristics had a significant relationship with vaccine uptake [24].

We also ran a final multivariate model to determine whether there was any change in the relationship between individual variables and vaccine uptake after taking the 3 sets of variables (core motivators, ability factors, and sociodemographic characteristics) into account. Adjusted odds ratios (aORs) from the analyses are shown in the tables. P values were considered statistically significant at $P<.05$.

Ethical Considerations

The ethical approval for the study was obtained from Nigeria’s National Health Research Ethics Committee (NHREC/01/01/2007). Informed consent was obtained from all respondents to the quantitative survey. Respondents 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. Respondents were informed about their right to refuse to participate in the study or withdraw at any time during the interview.

Results

Core Motivators and COVID-19 Vaccine Uptake

Table 1 shows the frequency distributions of core elements of motivation identified by the FBM, cross-tabulations between core motivators and COVID-19 vaccine uptake and the aOR of COVID-19 vaccination. The first column of Table 1 shows that about three-fourths (n=726, 75.9%) of survey respondents agreed or strongly agreed with the statement that getting vaccinated against COVID-19 allows a person to live a better life, a measure of hope. Nearly half (n=444, 46.4%) of respondents who disagreed with the question “Many of my friends and family would think poorly of me if they knew I had taken the COVID-19 vaccine” did not demonstrate the expected relationship with vaccine uptake. This was replaced with the statement “Most people I know have obtained the COVID-19 vaccination.” The second pretest showed that the latter response was associated with the outcome in the expected direction. All other items included in the instrument demonstrated the expected relationships with vaccine uptake in both pretests.
reported that getting vaccinated protected them and their families from hardship, a measure of fear. Nearly two-thirds (n=610, 63.7%) of respondents agreed or strongly agreed with the statement that getting vaccinated allows a person to do the things they enjoy, a measure of pleasure. About 81.5% (n=780) reported that vaccination reduces the likelihood of getting or spreading COVID-19, a measure of pain. About 79.1% (n=757) of respondents reported that many of their family and friends approve of the COVID-19 vaccination, a variable measuring social acceptance. Consistent with the high social acceptance of the COVID-19 vaccination, rejection of the vaccine was much lower, that is, 37.1% (n=355) of respondents reported that most of their family and friends did not approve of the COVID-19 vaccination.

Table 1. Frequency distributions of core motivators, cross-tabulations, and the adjusted odds of COVID-19 vaccination in Nigeria.

<table>
<thead>
<tr>
<th>Core motivators</th>
<th>Frequency distributions of core motivators (N=957, 100%), n (%)</th>
<th>Nigerian adults who obtained at least 1 vaccination (n=540, 56.4%), n (%)</th>
<th>P value</th>
<th>Adjusted odds of obtaining at least 1 COVID-19 vaccination, aOR (95% CI) a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hope: Vaccination allows a person to live a better life</td>
<td>Agree or strongly agree: 726 (75.9)</td>
<td>483 (66.5)</td>
<td>.001</td>
<td>3.40 (2.27-5.09)</td>
</tr>
<tr>
<td></td>
<td>Disagree or strongly disagree or do not know: 231 (24.1)</td>
<td>57 (24.7)</td>
<td>1.00 (Reference)</td>
<td></td>
</tr>
<tr>
<td>Fear: Getting vaccinated protects people from hardship</td>
<td>Agree or strongly agree: 444 (46.4)</td>
<td>301 (67.8)</td>
<td>.001</td>
<td>1.09 (0.78-1.51)</td>
</tr>
<tr>
<td></td>
<td>Disagree or strongly disagree or do not know: 513 (53.6)</td>
<td>239 (46.6)</td>
<td>1.00 (Reference)</td>
<td></td>
</tr>
<tr>
<td>Pleasure: Getting vaccinated allows people to do more things they enjoy</td>
<td>Agree or strongly agree: 610 (63.7)</td>
<td>411 (67.4)</td>
<td>.001</td>
<td>1.84 (1.30-2.62)</td>
</tr>
<tr>
<td></td>
<td>Disagree or strongly disagree or do not know: 347 (36.3)</td>
<td>129 (37.2)</td>
<td>1.00 (Reference)</td>
<td></td>
</tr>
<tr>
<td>Pain: Getting vaccinated reduces the likelihood of getting or spreading COVID-19</td>
<td>Agree or strongly agree: 780 (81.5)</td>
<td>477 (61.2)</td>
<td>.001</td>
<td>1.15 (0.76-1.73)</td>
</tr>
<tr>
<td></td>
<td>Disagree or strongly disagree or do not know: 177 (18.5)</td>
<td>63 (35.6)</td>
<td>1.00 (Reference)</td>
<td></td>
</tr>
<tr>
<td>Acceptance: Many friends and family approve vaccination</td>
<td>Agree or strongly agree: 757 (79.1)</td>
<td>472 (62.4)</td>
<td>.001</td>
<td>1.76 (1.21-2.58)</td>
</tr>
<tr>
<td></td>
<td>Disagree or strongly disagree or do not know: 200 (20.9)</td>
<td>68 (34)</td>
<td>1.00 (Reference)</td>
<td></td>
</tr>
<tr>
<td>Rejection: Most family and friends do not approve of vaccination</td>
<td>Agree or strongly agree: 355 (37.1)</td>
<td>344 (57.1)</td>
<td>.06</td>
<td>1.06 (0.79-1.43)</td>
</tr>
<tr>
<td></td>
<td>Disagree or strongly disagree or do not know: 602 (62.9)</td>
<td>196 (33.5)</td>
<td>1.00 (Reference)</td>
<td></td>
</tr>
</tbody>
</table>

aPseudo $R^2$=11.85%

The second column of Table 1 shows cross-tabulations between the core motivators of vaccination and vaccine uptake. There were large differences in vaccine uptake by core motivators at the bivariate level. Respondents who associated a COVID-19 vaccination with the hope of a better life had a 42-percentage point higher rate of vaccination (n=483, 66.5% vs n=57, 24.7%; P<.001). Those who feared the hardship that COVID-19 infection would bring had a 21-percentage point higher vaccination rate than other respondents (n=301, 67.8% vs n=239, 46.6%; P<.001). The pleasure that respondents associated with being able to do what they enjoyed doing because of being vaccinated was reflected by a 30-percentage point higher rate of vaccination (n=411, 67.4% vs n=129, 37.2%; P<.001). Respondents’ concern that not being vaccinated would result in getting or spreading COVID-19 was associated with a 25-percentage point higher rate of vaccination (n=477, 61.2% vs n=63, 35.6%; P<.001). Acceptance of the vaccine by friends and family was associated with a 28-percentage point higher vaccination rate (n=472, 62.4% vs n=68, 34%; P<.001). It is interesting that social rejection, or the lack of approval of the vaccination by family members, was not associated with vaccine uptake.

The third column of Table 1 shows the adjusted odds of COVID-19 vaccination. With all core motivators in the model, Nigerians who believed that COVID-19 vaccination was associated with a better life, had 3 times higher odds of getting vaccinated (aOR 3.40, 95% CI 2.27-5.09). Those who believed that getting vaccinated would allow them to do more things that they enjoyed were more likely to be vaccinated (aOR 1.84, 95% CI 1.30-2.62). Friends and family members’ acceptance of their
decision to get vaccinated was associated with a higher vaccination rate (aOR 1.76, 95% CI 1.21-2.58).

### Ability Barriers and COVID-19 Vaccine Uptake

Table 2 shows the frequency distributions of ability factors identified by the FBM, cross-tabulations between ability factors and COVID-19 vaccine uptake, and the adjusted odds of COVID-19 vaccination. The first column of Table 2 shows that about 54% (n=516) of respondents felt that their family or work responsibilities made it difficult for them to get vaccinated. This variable measures the constraint of time. Over a third (n=347, 36.3%) of respondents felt that the cost or loss of income associated with getting vaccinated was a barrier. Nearly 40% (n=437, 46%) of the respondents felt that the decision to get vaccinated was difficult. This variable measures the mental effort required to get vaccinated. About 42% (n=401) of respondents reported that not having the vaccine available in places they routinely visited was a barrier to getting vaccinated. The latter measures the routine associated with adopting a behavior.

<table>
<thead>
<tr>
<th>Ability</th>
<th>Frequency distributions of ability factors (N=957, 100%), n (%)</th>
<th>Nigerian adults who obtained at least one vaccination (n=540, 56.4%), n (%)</th>
<th>P value</th>
<th>Adjusted odds of obtaining at least one COVID-19 vaccinationa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time: Family or work responsibilities make it difficult to find time</td>
<td>Agree or strongly agree 516 (53.9)</td>
<td>277 (53.7)</td>
<td>.06</td>
<td>0.94 (0.69-1.27)</td>
</tr>
<tr>
<td>Disagree or strongly disagree or do not know 441 (46.1)</td>
<td>263 (59.6)</td>
<td>1.00 (Reference)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Money: Costs or loss of income make it difficult</td>
<td>Agree or strongly agree 347 (36.3)</td>
<td>158 (45.5)</td>
<td>&lt;.001</td>
<td>0.45 (0.33-0.62)</td>
</tr>
<tr>
<td>Disagree or strongly disagree or do not know 610 (63.7)</td>
<td>382 (62.6)</td>
<td>1.00 (Reference)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical effort: Physical effort makes it difficult</td>
<td>Agree or strongly agree 382 (39.9)</td>
<td>218 (57.1)</td>
<td>.74</td>
<td>1.46 (1.06-2.01)</td>
</tr>
<tr>
<td>Disagree or strongly disagree or do not know 575 (60.1)</td>
<td>322 (56.0)</td>
<td>1.00 (Reference)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mental effort: Decision to get vaccine is difficult</td>
<td>Agree or strongly agree 437 (45.7)</td>
<td>237 (54.2)</td>
<td>.21</td>
<td>0.94 (0.70-1.27)</td>
</tr>
<tr>
<td>Disagree or strongly disagree or do not know 520 (54.3)</td>
<td>303 (58.3)</td>
<td>1.00 (Reference)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Routine: Vaccine not available where I routinely visit</td>
<td>Agree or strongly agree 401 (41.9)</td>
<td>141 (35.2)</td>
<td>&lt;.001</td>
<td>0.21 (0.16-0.28)</td>
</tr>
<tr>
<td>Disagree or strongly disagree or do not know 556 (58.1)</td>
<td>399 (71.8)</td>
<td>1.00 (Reference)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

aPseudo $R^2=12\%$.

The second column of Table 2 shows cross-tabulations between ability factors and vaccine uptake. Respondents who agreed or strongly agreed with the statement that costs or loss of income were a barrier reported a 17-percentage point lower vaccination rate compared to others (n=158, 45.5% vs n=382, 62.6%; P<.001). The lack of availability of the COVID-19 vaccine in places that they routinely visited was associated with a 37-percentage point lower rate of vaccination (n=141, 35.2% vs n=399, 71.8%; P<.001).

The third column of Table 2 shows the adjusted odds of a COVID-19 vaccination. With all ability factors in the model, Nigerians who believed that the cost or the loss of income made it difficult to obtain a COVID-19 vaccination were less likely to get vaccinated (aOR 0.45, 95% CI 0.33-0.62). The lack of availability of vaccines at places respondents routinely visited was associated with a lower likelihood of vaccination (aOR 0.21, 95% CI 0.16-0.28). Contrary to our expectations, respondents who felt that physical effort made it difficult to get vaccinated were more likely to be vaccinated (aOR 1.46, 95% CI 1.06-2.01).

### Sociodemographic Factors and COVID-19 Vaccine Uptake

Table 3 shows the frequency distributions of sociodemographic characteristics of respondents in the sample, cross-tabulations between these characteristics and vaccine uptake, and the adjusted odds of COVID-19 vaccination. The first column of Table 3 shows that, as expected from a digital survey, the sample had relatively young participants: 56% (n=531) of respondents were aged 18-29 years and 16% (n=151) were aged 40 years and older. Male participants represented a higher proportion of the sample (n=592, 61.9%). About 39% (n=371) of respondents had a primary or secondary school certificate, one quarter

https://www.i-jmr.org/2024/1/e47817
(page number not for citation purposes)
(n=231, 24.1%) had an ordinary national diploma (OND) or a higher national diploma (HND), and one-third of respondents (n=317, 33.4%) had a bachelor’s or higher degree. A majority of respondents were from urban areas: 57% (n=544) were from cities, 34% (n=323) from towns, and 9% (n=90) from rural areas.

**Table 3.** Frequency distributions of sociodemographic variables, cross-tabulations, and the adjusted odds of COVID-19 vaccination in Nigeria.

<table>
<thead>
<tr>
<th>Demographic</th>
<th>Sample characteristics (N=957, 100%), n (%)</th>
<th>Nigerian adults who obtained at least 1 vaccination (n=540, 56.4%), n (%)</th>
<th>P value</th>
<th>Adjusted odds of obtaining at least 1 COVID-19 vaccination, aOR (95% CI)a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-29</td>
<td>531 (55.5)</td>
<td>300 (56.5)</td>
<td>.71</td>
<td>1.31 (0.89-1.93)</td>
</tr>
<tr>
<td>30-39</td>
<td>275 (28.7)</td>
<td>159 (57.8)</td>
<td></td>
<td>1.28 (0.85-1.93)</td>
</tr>
<tr>
<td>≥40</td>
<td>151 (15.8)</td>
<td>81 (53.8)</td>
<td></td>
<td>1.00 (Reference)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td>.06</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>592 (61.9)</td>
<td>348 (58.8)</td>
<td></td>
<td>1.00 (Reference)</td>
</tr>
<tr>
<td>Female</td>
<td>365 (38.1)</td>
<td>192 (52.6)</td>
<td></td>
<td>0.73 (0.56-0.95)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Primary or secondary school certificate</td>
<td>371 (38.8)</td>
<td>188 (50.7)</td>
<td></td>
<td>1.00 (Reference)</td>
</tr>
<tr>
<td>Ordinary national diploma (OND)</td>
<td>111 (11.6)</td>
<td>72 (64.9)</td>
<td>1.90 (1.22-2.98)</td>
<td></td>
</tr>
<tr>
<td>Higher national diploma (HND)</td>
<td>120 (12.5)</td>
<td>87 (72.5)</td>
<td>2.62 (1.64-4.19)</td>
<td></td>
</tr>
<tr>
<td>Bachelors or higher</td>
<td>317 (33.4)</td>
<td>169 (53.3)</td>
<td>1.12 (0.81-1.55)</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>38 (4)</td>
<td>24 (63.2)</td>
<td></td>
<td>1.79 (0.89-3.60)</td>
</tr>
<tr>
<td>Location</td>
<td></td>
<td></td>
<td>.02</td>
<td></td>
</tr>
<tr>
<td>City</td>
<td>544 (56.8)</td>
<td>320 (58.8)</td>
<td></td>
<td>1.00 (Reference)</td>
</tr>
<tr>
<td>Town</td>
<td>323 (33.8)</td>
<td>181 (56.0)</td>
<td></td>
<td>0.88 (0.68-1.17)</td>
</tr>
<tr>
<td>Rural</td>
<td>90 (9.4)</td>
<td>39 (43.3)</td>
<td></td>
<td>0.56 (0.35-0.89)</td>
</tr>
</tbody>
</table>

Pseudo $R^2=2.79\%$.

The second column of Table 3 shows cross-tabulations between sociodemographic characteristics and vaccine uptake. There was no statistically significant difference in the COVID-19 vaccination rate by age or sex. Education was associated with vaccine uptake: respondents with an OND (n=72, 64.9% vs n=188, 50.7%; $P<.001$) or an HND (n=87, 72.5% vs n=188 50.7%; $P<.001$) were more likely to be vaccinated than respondents with a primary or secondary school certificate. Urban residence was also associated with higher vaccine uptake: respondents from rural areas were significantly less likely to have obtained the COVID-19 vaccination (n=39, 43.3% vs n=320, 58.8%; $P=.02$).

The third column of Table 3 shows the adjusted odds of COVID-19 vaccination. With all sociodemographic characteristics in the model, female participants were less likely to get vaccinated (aOR 0.73, 95% CI 0.56-0.95). Having an OND (aOR 1.90, 95% CI 1.22-2.98) or a HND (aOR 2.62, 95% CI 1.64-4.19) increased a respondent’s likelihood of being vaccinated. Nigerians living in rural areas were less likely to be vaccinated (aOR 0.56, 95% CI 0.35-0.89).

**Full Model: Core Motivators, Ability Barriers, and Sociodemographic Characteristics**

Table 4 shows the adjusted odds of COVID-19 vaccine uptake in Nigeria. The 3 core motivators identified earlier remained significant after adjusting for ability factors and sociodemographic characteristics. Nigerians who believed that the COVID-19 vaccination was associated with the promise of a better life were more likely to be vaccinated (aOR 3.51, 95% CI 2.23-5.52). Nigerians who felt that the vaccination would allow them to do more things they enjoyed were more likely to be vaccinated (aOR 1.97, 95% CI 1.33-2.93). Respondents’ friends and family members’ acceptance of their decision to get vaccinated was associated with a higher likelihood of their being vaccinated (aOR 1.62, 95% CI 1.06-2.48).
The relationships between ability factors and vaccine uptake remained important after adjusting for sociodemographic characteristics and core motivators. Nigerians with cost- or income-related concerns were less likely to obtain a COVID-19 vaccination (aOR 0.35, 95% CI 0.24-0.50). The lack of availability of vaccines at places they routinely visited made them less likely to get vaccinated (aOR 0.29, 95% CI 0.21-0.40).

After adjusting for motivation and ability, female participants were no longer less likely to obtain a COVID-19 vaccination. Respondents with an OND (aOR 2.27, 95% CI 1.34-3.84) or HND (aOR 3.57, 95% CI 2.04-6.24) were more likely to be vaccinated than those with primary or secondary school certificates. Rural residents were less likely to be vaccinated than residents living in cities (aOR 0.54, 95% CI 0.31-0.92).

### Discussion

#### Principal Findings

The findings of this study show that 56.4% (n=540) of Nigerian adults who responded to the digital survey had obtained at least 1 COVID-19 vaccination by October 2022. Several core motivators were associated with vaccine uptake, after adjusting

---

**Table 4.** Adjusted odds (aOR) of COVID-19 vaccination in Nigeria.

<table>
<thead>
<tr>
<th>Core motivators</th>
<th>Adjusted odds of obtaining at least 1 COVID-19 vaccination, aOR (95% CI)³</th>
</tr>
</thead>
<tbody>
<tr>
<td>Getting vaccinated allows a person to live a better life (hope)</td>
<td>3.51 (2.23-5.52)</td>
</tr>
<tr>
<td>Getting vaccinated protects people from hardship (fear)</td>
<td>1.03 (0.71-1.49)</td>
</tr>
<tr>
<td>Allows people to do more things they enjoy (pleasure)</td>
<td>1.97 (1.33-2.93)</td>
</tr>
<tr>
<td>Reduces the likelihood of getting or spreading COVID-19 (pain)</td>
<td>1.17 (0.74-1.84)</td>
</tr>
<tr>
<td>Many friends and family approve vaccination (acceptance)</td>
<td>1.62 (1.06-2.48)</td>
</tr>
<tr>
<td>Most family and friends do not approve vaccination (rejection)</td>
<td>1.18 (0.83-1.67)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ability factors</th>
<th>Adjusted odds of obtaining at least 1 COVID-19 vaccination, aOR (95% CI)³</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family or work responsibilities make it difficult (time)</td>
<td>0.85 (0.61-1.19)</td>
</tr>
<tr>
<td>Costs or loss of income make it difficult (money)</td>
<td>0.35 (0.24-0.50)</td>
</tr>
<tr>
<td>Physical effort makes it difficult (physical effort)</td>
<td>1.45 (1.02-2.07)</td>
</tr>
<tr>
<td>The decision to get the vaccine is difficult (mental effort)</td>
<td>1.08 (0.77-1.51)</td>
</tr>
<tr>
<td>Vaccine not available where I routinely visit (routine)</td>
<td>0.29 (0.21-0.40)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sociodemographic factors</th>
<th>Adjusted odds of obtaining at least 1 COVID-19 vaccination, aOR (95% CI)³</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>18-29</td>
<td>1.09 (0.68-1.74)</td>
</tr>
<tr>
<td>30-39</td>
<td>1.14 (0.70-1.86)</td>
</tr>
<tr>
<td>≥40</td>
<td>1.00 (Reference)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1.00 (Reference)</td>
</tr>
<tr>
<td>Female</td>
<td>0.80 (0.58-1.10)</td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
</tr>
<tr>
<td>Primary or secondary school certificate</td>
<td>1.00 (Reference)</td>
</tr>
<tr>
<td>Ordinary national diploma (OND)</td>
<td>2.27 (1.34-3.84)</td>
</tr>
<tr>
<td>Higher national diploma (HND)</td>
<td>3.57 (2.04-6.24)</td>
</tr>
<tr>
<td>Bachelors or higher</td>
<td>1.26 (0.86-1.85)</td>
</tr>
<tr>
<td>Other</td>
<td>1.88 (0.82-4.33)</td>
</tr>
<tr>
<td><strong>Location</strong></td>
<td></td>
</tr>
<tr>
<td>City</td>
<td>1.00 (Reference)</td>
</tr>
<tr>
<td>Town</td>
<td>0.90 (0.64-1.26)</td>
</tr>
<tr>
<td>Rural</td>
<td>0.54 (0.31-0.92)</td>
</tr>
</tbody>
</table>

³Pseudo $R^2=23.51\%$.
for ability factors and sociodemographic variables. The belief that COVID-19 vaccination allows a person to live a better life and that it allows them to do more things that they enjoy increases the likelihood of being vaccinated. The belief that many friends and family members approve of COVID-19 vaccination is also associated with a greater likelihood of being vaccinated. Several ability barriers were also correlated with vaccine uptake, after adjusting for other variables. Respondents who felt that costs or loss of income associated with getting vaccinated made it difficult to get vaccinated were less likely to get vaccinated. The lack of availability of the COVID-19 vaccine at places respondents routinely visited was also negatively associated with vaccine uptake. The study also found a higher likelihood of Nigerians with OND or HND being vaccinated compared with those with primary or secondary school certificates and a lower likelihood of being vaccinated among rural respondents.

**Strengths and Limitations**

While this study provides useful insights for program design, its limitations should be acknowledged. The first limitation of this study is that no causal inferences can be made from it because of its cross-sectional design. This design limitation may explain an unexpected study finding: after adjusting for other variables, the physical effort required to get vaccinated was associated with a higher rate of vaccination. This finding may reflect reverse causality; those who are vaccinated may be more aware of the physical effort required to obtain a COVID-19 vaccination. Further investigation is needed to determine whether the positive relationship between the perceived physical effort required to get vaccinated and receiving a vaccination holds only for those who have been vaccinated or for the full sample.

The second limitation of this study is that variables that were not measured may be responsible for the observed relationships. For example, while the relationship between the belief that vaccination allows a person to enjoy life more and vaccine uptake is powerful, there is a possibility that unmeasured factors are driving this relationship. Thus, developing messages around how vaccination can help a person lead a fuller life and testing them through relatively inexpensive digital campaigns would be important prior to implementing an at-scale campaign that focuses on this message.

The third limitation of the study is that it is not representative of all Nigerians in the 6 states in which it was conducted. This is reflected in the higher educational status of the survey sample: about 33% (n=317) of respondents had a bachelor’s or higher education. The participants were also relatively young; about 56% (n=531) of respondents were between 18 and 29 years. Moreover, male participants comprised a higher proportion of the sample than female participants. These findings are not uncommon for digital surveys conducted in LMICs.

A strength of this study is its cost efficiency and the timeliness with which the survey was conducted compared to face-to-face household surveys. A major barrier to the use of behavioral insights by practitioners in LMICs is the cost of data collection. Behavioral research is not well-funded in LMICs. Inexpensive digital surveys could substantially increase the ability of practitioners in LMICs to use behavioral insights to develop interventions that increase vaccine uptake.

**Future Directions**

We do not know the extent to which the findings from surveys conducted by recruiting respondents through web-based advertising are comparable to the findings from population-based household surveys. Although some studies show broadly similar patterns between digital and population-based surveys [25], more research is needed to identify what types of systematic differences may exist between these 2 survey modalities. It is important, for example, to learn whether inferences from digital surveys apply to the behavior of individuals who are not on digital platforms.

Given our inability to generalize these findings beyond Nigerians who are on Facebook and Instagram, how can the findings of this survey be used? First, the findings may be used to design interventions on digital platforms as well as to evaluate the effectiveness of those interventions. Digital behavior change interventions may be evaluated by experimental studies on digital platforms that compare vaccine uptake between intervention and control groups. A growing proportion of the Nigerian population is now on Facebook and Instagram: between 31 million and 36 million Nigerians 13 and older use Facebook and Instagram each month. High exposure to messages that associate the COVID-19 vaccination with a better, more fulfilling life is achievable through advertising on digital platforms and at a fraction of the cost of advertising on traditional mass media channels such as television.

Future interventions could build upon the findings of this study by conducting qualitative research to determine which motivation or ability factors are relevant in locations where interventions are planned. A survey conducted in the most densely populated, low-income commune in Yugpognon, Cote d’Ivoire, using the FBM found broadly comparable findings: motivation and ability were powerful drivers of vaccine adoption, although the specific elements of motivation and ability that were relevant in Yugpognon were, not surprisingly, different [16].

Our findings raise several questions that should be answered through additional research. Answers to these questions may help in the design of more effective COVID-19 service delivery interventions. It would be useful, for example, to learn whether some of the places routinely visited by Nigerians are amenable to serving as COVID-19 vaccine delivery sites. Are such potential vaccine delivery sites likely to vary by age, sex, by urban or rural residence and are they suitable for cost-efficient provision of COVID-19 vaccinations?

Our sample consisted primarily of Nigerians living in cities and towns, with a minority of respondents living in rural areas. A larger proportion of rural respondents may be obtained from digital surveys that oversample rural areas. This may be done by capping the number of respondents from urban areas and including more rural areas. This may be done by targeting places that are amenable to digital interventions. It would be useful, for example, to learn whether digital interventions that increase vaccine uptake are feasible and effective in different settings.
Comparison to Prior Work

Overall, the findings of our study provide a very different perspective on vaccine adoption in Nigeria than what is available in the peer-reviewed literature. Recent Nigerian studies on COVID-19 vaccine uptake have primarily focused on identifying gaps in knowledge, attitudes, and beliefs associated with vaccine hesitancy [1-3]. These studies place emphasis on the perceived risk of disease as a driver of vaccine uptake, despite mixed evidence on the role of risk perception on vaccine uptake in Nigeria [5,6,9]. Several of these recent studies propose that equipping people with factually correct information will allay their concerns, increase their perceived risk of acquiring COVID-19, and lead to higher rates of vaccine adoption. By contrast, a small but rapidly growing body of work is putting vaccine adoption in a behavioral context [13-19,26]. These studies take motivation and ability for behavior change into account in explaining the range of barriers that influence immunization decisions and suggest how programs should help individuals overcome them.

Conclusions

These findings help us consider a very different approach to intervention design—one that builds upon what people want for their future, what gives them pleasure, and how they are influenced by the approval of their friends and family members. Our findings suggest that hope is more important for Nigerians than fear when it comes to vaccine adoption, social approval is more powerful than social disapproval, and enjoying life is more important than worrying about getting the disease.

These findings suggest that an approach that is based on increasing the perception of hope and pleasure associated with vaccine adoption as well as increasing network members’ social approval is likely to increase COVID-19 vaccine adoption in Nigeria. Our analysis also suggests that financial considerations play an important role in the uptake of COVID-19 vaccination in Nigeria. The costs associated with reaching a vaccination site or the loss of income associated with being away from work are important determinants of vaccine adoption. Nigerians with limited flexibility at work may find it challenging to visit a vaccination site during the hours that it is open. Consistent with this finding, making COVID-19 vaccines available at places that Nigerians visit routinely may have a large impact on vaccine uptake.

The use of a behavior model to understand drivers of COVID-19 vaccine uptake in Nigeria has helped provide a different perspective on vaccine-related decision-making in Nigeria than what is currently available in the published literature. The FBM, a model of human behavior rather than a model of health behavior per se, considers a broad range of factors influencing motivation, including an individual’s hopes and fears, the sensation of pleasure or pain that they get from a particular behavior, and the social influences on them associated with their identity. The model also measures ability constraints including bandwidth-related constraints such as time or cognitive constraints, financial constraints, physical effort–related constraints, and habits or routine-related constraints. By comparison to behavior models that focus on perceived risk of and susceptibility to disease, the FBM situates behavior within the broader context of a person’s life.

Acknowledgments

The authors gratefully acknowledge the study participants for the information they provided. Funding for this study was provided by The Bill & Melinda Gates Foundation. The funders had no input on the interpretation or publication of the study results.

Data Availability

The data sets generated and analyzed during this study are available from the corresponding author on reasonable request.

Authors' Contributions

IN developed and submitted the institutional review board protocol to Nigeria’s National Health Research Ethics Committee. SA and DB conceptualized the study. SA conducted the statistical analysis and wrote the first draft of the report. IN provided important context on Nigeria that enabled the interpretation of the study findings. DB developed the survey instrument and interpreted the study findings. DB and SF were responsible for acquiring funding for the study. SF and NR designed the digital data collection instruments and implemented the data collection and quality control procedures. All authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey questionnaire.

References


Abbreviations

- **aOR**: adjusted odds ratio
- **FBM**: Fogg Behavior Model
- **HCW**: health care worker
- **HND**: higher national diploma
- **LMICs**: low- and middle-income countries
- **OND**: ordinary national diploma

©Sohail Agha, Ifeanyi Nsofor, Drew Bernard, Sarah Francis, Nandan Rao. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 26.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.
The Effect of Body Temperature Changes on the Course of Treatment in Patients With Pneumonia and Sepsis: Results of an Observational Study

Domen Guzelj\textsuperscript{1*}; Anže Grubelnik\textsuperscript{1*}; Nina Greif\textsuperscript{1*}; Petra Povalej Bržan\textsuperscript{2*}, PhD; Jure Fluher\textsuperscript{3*}, MD; Žiga Kalamar\textsuperscript{3*}, MD; Andrej Markota\textsuperscript{3*}, MD, PhD

\textsuperscript{1}Faculty of Medicine, University of Maribor, Maribor, Slovenia
\textsuperscript{2}Faculty of Electrical Engineering and Computer Science, University of Maribor, Maribor, Slovenia
\textsuperscript{3}Medical Intensive Care Unit, University Medical Centre Maribor, Maribor, Slovenia

*all authors contributed equally

Corresponding Author:
Andrej Markota, MD, PhD
Medical Intensive Care Unit
University Medical Centre Maribor
Ljubljanska ulica 5
Maribor, 2000
Slovenia
Phone: 386 23212008
Email: andrej.markota@ukc-mb.si

Abstract

Background: Traditionally, patients who are critically ill with infection and fever have been treated with antipyretics or even physically cooled. Presumed benefits of the reduction of body temperature are mostly based on decreased metabolic demands. However, it has been shown that decreasing body temperature in patients who are critically ill is not associated with improvement in treatment outcomes. Additionally, there is some data to support the use of temperature modulation (therapeutic hyperthermia) as an adjuvant treatment strategy in patients with infection.

Objective: This study aims to determine the effect of body temperature on the course of intensive care unit (ICU) treatment of patients who are mechanically ventilated with pneumonia, sepsis, and positive tracheal aspirates on admission.

Methods: We performed a single-center retrospective study. Core body temperature was measured in all patients. We analyzed associations between average temperatures in the first 48 hours after admission to ICU and ICU treatment parameters. Additionally, patients were divided into three groups: patients with negative tracheal aspirates 1 week after ICU admission (P-N group), patients with a different pathogen in tracheal aspirates 1 week after ICU admission (P-HAP group), and patients with a persisting pathogen in tracheal aspirates 1 week after ICU admission (P-P group). Differences in body temperature and interventions aimed at temperature modulation were determined.

Results: We observed a significantly higher average temperature in the first 48 hours after admission to ICU in patients who survived to hospital discharge compared to nonsurvivors (mean 37.2 °C, SD 1 °C vs mean 36.9 °C, SD 1.6 °C; \( P=0.04 \)). We observed no associations between average temperatures in the first 48 hours after ICU admission and days of mechanical ventilation in the first 7 days of treatment (\( \rho=-0.090; P=0.50 \)), the average maximum daily requirement for noradrenaline in the first 7 days of treatment (\( \rho=0.029; P=0.80 \)), average maximum FiO\textsubscript{2} in the first 7 days of ICU treatment (\( \rho=0.040; P=0.70 \)), and requirement for renal replacement therapy in the first 7 days of ICU treatment (mean 37.3 °C, SD 1.4 °C vs mean 37.0 °C, SD 1.3 °C; \( P=0.23 \)). In an additional analysis, we observed a significantly greater use of paracetamol in the P-N group (mean 1.0, SD 1.1 g vs mean 0.4, SD 0.7 g vs mean 0.4, SD 0.8 g; \( P=0.009 \)), a trend toward greater use of active cooling in the first 24 hours after ICU admission in the P-N group (n=11, 44% vs n=14, 33.3% vs n=16, 32%; \( P=0.57 \)), and no other significant differences in parameters of ICU treatment between patient groups.

Conclusions: We observed better survival in patients who developed higher body temperatures in the first 48 hours after admission to the ICU; however, we observed no changes in other treatment parameters. Similarly, we observed greater use of...
paracetamol in patients with negative tracheal aspirates 1 week after ICU admission. Our results support the strategy of temperature tolerance in patients who are intubated with pneumonia and sepsis.


KEYWORDS
fever; targeted temperature management; pneumonia; sepsis; intensive care unit

Introduction

Increased body temperature has been recognized as a sign of illness for more than 2000 years and antipyretics have been used for at least 100 years with the aim of lowering body temperature in patients who are febrile [1]. However, a body of evidence exists ranging from studies in the preantibiotic era to a recently published pilot trial, where increased body temperature is investigated as a treatment option in patients with infection [2-4]. A fever-range increase of body temperature is a highly preserved response that is probably beneficial in patients who develop fever as a part of a normal immune response to infection [5]. Recently, we have observed that patients with lower body temperatures are at a higher risk of acquiring the presence of multidrug-resistant pathogens [6]. The aim of this study was to investigate the association between body temperature and the course of intensive care unit (ICU) treatment in patients with pneumonia and sepsis and to evaluate the effect of temperature on the persistence of positive control tracheal aspirates.

Methods

Study Design and Settings

We performed retrospective observational data collection from January 1, 2018, to December 31, 2021. The study was performed in a medical ICU in a tertiary center and supported by an institutional research grant (grant IRP-2022/01-01).

Ethical Considerations

The study was approved by University Medical Centre Maribor Ethical Committee, and informed consent was waived because of the observational and retrospective nature of the study (No. UKC-MB-KME-35/22). Patients’ data have been anonymized and deidentified. No compensation was provided to the participants.

Study Population

We included adult (aged >18 years) patients with pneumonia, septic shock, positive tracheal aspirates within 24 hours after admission (“admission” tracheal aspirates), and tracheal aspirates withdrawn 8-14 days after admission to ICU (“week 2” tracheal aspirates). We excluded patients who were treated with targeted temperature management for accidental hypo- or hyperthermia, patients after cardiac arrest, and patients who are neurocritical (eg, patients with meningitis, encephalitis, ischemic or hemorrhagic stroke, or subarachnoid hemorrhage). Temperature management in our patient cohort was as per the treating physician.

Measurements

We collected basic demographic data and data related to ICU treatment, namely outcome of ICU treatment, ICU length of stay, core body temperature during ICU stay, use of renal replacement therapy, use of acetaminophen, maximum concentration of noradrenaline, maximum fraction of inspired oxygen, maximum level of positive end-expiratory pressure and maximum minute ventilation, and microbiological results of tracheal aspirates. Source data was paper based for temperature and therapeutic charts and electronic for other data. Core body temperature was used for study data. As per department policy, temperature was measured via thermal probe urinary catheters (Rüsch Sensor Urinary Catheters, Teleflex Medical, Athlone, Ireland), and temperature measurements were continuously displayed on ICU monitors (Philips IntelliVue MX800 Patient Monitoring System, Koninklijke Philips N.V., Amsterdam, Netherlands). If insertion of a urinary catheter is not possible, then an esophageal temperature probe is inserted, but temperature measurement in all of the included patients was performed via a urinary catheter. Temperature results from 2-hourly notations were used for statistical analysis. We compared the association between body temperature and course of treatment parameters between three groups of patients: patients with sterile week 2 tracheal aspirates (P-P group), patients with a different pathogen in week 2 tracheal aspirates (P-HAP group), and patients with a persistent pathogen presence in week 2 tracheal aspirates (P-P group).

Data Analysis

Statistical analyses were carried out using R (version 4.1.1: R Foundation for Statistical Computing). Nominal variables are presented with frequencies (percentages) and numerical variables with means (SDs) or medians (IQRs) when the normality assumption is violated. For the comparison of nominal dichotomous variables, the Fisher exact test was used. Continuous variables were first assessed for normality using the D’Agostino omnibus test. A comparison of continuous variables across groups was carried out using the Kruskal-Wallis test. Dunn post hoc test with Bonferroni correction was used to adjust for multiple comparisons. The association between average temperature in the first 48 hours and the duration of mechanical ventilation, the average maximum daily requirement for noradrenaline and the average maximum FiO2 in the first 7 days was evaluated using Spearman correlation. Generalized linear models were used to additionally estimate the ICU and hospital survival in association with the abovementioned risk factors. A statistically significant observation was considered at $P < 0.05$. 

https://www.i-jmr.org/2024/1/e52590
Results

Baseline Characteristics

In all, 117 patients were included in the study analysis; 84 (71.8%) were male, and the mean age was 63.7 (SD 13.5) years. All patients were invasively mechanically ventilated in the ICU on day 1. Mean APACHE II and SOFA scores on admission were 21.7 (SD 6.6) and 10.1 (SD 2.6), respectively. A total of 77 (65.8%) were discharged alive from the ICU, and 46 (39.3%) patients were discharged alive from the hospital. P-N, P-HAP, and P-P groups consisted of 25, 42, and 50 patients, respectively. We observed no significant differences between the P-N, P-HAP, and P-P groups in all treatment parameters apart from the use of paracetamol on day 1, which was significantly greater in the P-N group (mean 1.0, SD 1.1 g vs mean 0.4, SD 0.7 g vs mean 0.4, SD 0.8 g; \( P = .009 \)). General demographic data and parameters describing the course of the treatment in the ICU are described in Table 1. The study patient population and inclusion flowchart are presented in Figure 1.

Table 1. Demographic data and parameters describing the course of treatment in the intensive care unit (ICU).

<table>
<thead>
<tr>
<th>Label and variable</th>
<th>Group</th>
<th>Total</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients, n (%)</td>
<td>P-N(^a)</td>
<td>P-P(^b)</td>
<td>P-HAP(^c)</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>5 (15.2)</td>
<td>18 (54.5)</td>
<td>10 (30.3)</td>
</tr>
<tr>
<td>Male</td>
<td>20 (23.8)</td>
<td>32 (38.1)</td>
<td>32 (38.1)</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>62.0 (52.0-74.0)</td>
<td>64.0 (57.2-71.8)</td>
<td>67.0 (62.0-76.0)</td>
</tr>
<tr>
<td>Min-max</td>
<td>23.0-84.0</td>
<td>26.0-84.0</td>
<td>39.0-86.0</td>
</tr>
<tr>
<td>BMI (kg/m(^2))</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>28.7 (7.0)</td>
<td>27.6 (4.3)</td>
<td>29.7 (6.4)</td>
</tr>
<tr>
<td>Min-max</td>
<td>17.2-47.7</td>
<td>18.8-39.2</td>
<td>16.3-44.8</td>
</tr>
<tr>
<td>Length of hospitalization (days)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>22.0 (17.0-29.0)</td>
<td>19.5 (15.0-26.8)</td>
<td>17.5 (13.0-26.8)</td>
</tr>
<tr>
<td>Min-max</td>
<td>8.0-53.0</td>
<td>10.0-43.0</td>
<td>8.0-69.0</td>
</tr>
<tr>
<td>ICU survival, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Died</td>
<td>6 (15.0)</td>
<td>16 (40.0)</td>
<td>18 (45.0)</td>
</tr>
<tr>
<td>Survived</td>
<td>19 (26.1)</td>
<td>34 (44.2)</td>
<td>24 (31.2)</td>
</tr>
<tr>
<td>Hospital survival, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Died</td>
<td>13 (18.3)</td>
<td>31 (43.7)</td>
<td>27 (38.0)</td>
</tr>
<tr>
<td>Survived</td>
<td>12 (26.1)</td>
<td>19 (41.3)</td>
<td>15 (32.6)</td>
</tr>
<tr>
<td>SOFA(^d) at the time of admission</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (IQR)</td>
<td>9.5 (7.0-12.0)</td>
<td>10.0 (9.0-12.0)</td>
<td>10.0 (9.0-12.0)</td>
</tr>
<tr>
<td>Min-max</td>
<td>5.0-16.0</td>
<td>5.0-15.0</td>
<td>4.0-15.0</td>
</tr>
<tr>
<td>APACHE(^e) II at the time of admission</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>20.4 (6.3)</td>
<td>22.0 (6.4)</td>
<td>22.1 (7.2)</td>
</tr>
<tr>
<td>Min-max</td>
<td>6.0-32.0</td>
<td>7.0-36.0</td>
<td>5.0-40.0</td>
</tr>
<tr>
<td>Change of antibiotic in 7 days after admission (no), n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>7 (18.4)</td>
<td>17 (44.7)</td>
<td>14 (36.8)</td>
</tr>
</tbody>
</table>

\(^a\)P-N: patients with sterile week 2 tracheal aspirates.
\(^b\)P-HAP: patients with a different pathogen in week 2 tracheal aspirates.
\(^c\)P-P group: patients with a persistent pathogen presence in week 2 tracheal aspirates.
\(^d\)SOFA: Sequential Organ Failure Assessment.
\(^e\)APACHE: Acute Physiology and Chronic Health Evaluation.
**Main Results**

Body temperature changes in the first 28 days of treatment are presented in Figure 2. We observed a significantly higher average temperature in the first 48 hours after admission to ICU in patients who survived to hospital discharge compared to nonsurvivors (mean 37.2 °C, SD 1 °C vs mean 36.9 °C, SD 1.6 °C; \( P=.04 \)) and nonsignificant differences in average temperatures in the first 48 hours after admission to ICU between survivors and nonsurvivors to ICU discharge (mean 37.0 °C, SD 1.3 °C vs mean 37.0 °C, SD 1.2 °C; \( P=.60 \)). We observed no associations between average temperatures in the first 48 hours after ICU admission and days of mechanical ventilation in the first 7 days of treatment (\( \rho=-0.090; P=.30 \)), the average maximum daily requirement for noradrenaline in the first 7 days of treatment (\( \rho=-0.029; P=.80 \)), average maximum FiO\(_2\) in the first 7 days of ICU treatment (\( \rho=0.040; P=.70 \)), and requirement for renal replacement therapy in the first 7 days of ICU treatment (mean 37.3 °C, SD 1.4 °C vs mean 37.0 °C, SD 1.3 °C; \( P=.23 \)).
P-N, P-HAP, and P-P groups consisted of 25, 42, and 50 patients each, respectively. No significant differences in average body temperature in the first 48 hours were observed between the P-P, P-HAP, and P-N groups. Temperature variations increased after day 16 (Figure 2), however, and approximately two-thirds of patients were discharged from the ICU or died before day 16. Both hospital and ICU survival were higher in the P-N group compared to the P-HAP and P-P groups; however, the differences were not statistically significant (Figure 3). We observed statistically significant greater use of paracetamol in the first 24 hours after ICU admission in the P-N group compared to the P-HAP and P-P groups (mean 1.0, SD 1.1 g vs mean 0.4, SD 0.7 g vs mean 0.4, SD 0.8 g; $P=0.009$). Active cooling was used more frequently on day 1 in the P-N group, but no significant differences were observed (Figure 4).

**Figure 2.** Average body temperature by groups in the first 28 days of treatment in the intensive care unit. P-HAP: patients with a different pathogen in week 2 tracheal aspirates; P-N: patients with sterile week 2 tracheal aspirates; P-P: patients with a persistent pathogen presence in week 2 tracheal aspirates.

**Figure 3.** Intensive care unit (ICU) and hospital survival by groups. P-HAP: patients with a different pathogen in week 2 tracheal aspirates; P-N: patients with sterile week 2 tracheal aspirates; P-P: patients with a persistent pathogen presence in week 2 tracheal aspirates.
Figure 4. Active cooling on day one of hospitalization in the intensive care unit divided by groups. P-HAP: patients with a different pathogen in week 2 tracheal aspirates; P-N: patients with sterile week 2 tracheal aspirates; P-P: patients with a persistent pathogen presence in week 2 tracheal aspirates.

Microbiological Data
In all, 148 different causal pathogens were isolated. In admission tracheal aspirates, the majority (n=124, 66%) of pathogens were gram-negative bacteria, followed by gram-positive bacteria (n=56, 29.8%) and fungal pathogens (4.3%). In week 2 tracheal aspirates, the majority (n=128, 86.5%) of pathogens were gram-negative bacteria, followed by gram-positive bacteria (n=14, 9.5%) and fungal pathogens (n=6, 4.1%). We observed no differences between the P-N, P-HAP, and P-P groups in causative pathogens in admission tracheal aspirates (Table 2). There were no differences between the P-N, P-HAP, and P-P groups in the rate of change of antimicrobial therapy within the first week after ICU admission, and we observed no changes in the duration of initial (combined empiric and antibiogram-guided antimicrobial therapy) and cumulative ICU-stay antibiotic therapy (mean 7.6, SD 2.7 days vs mean 8.4, SD 3.1 days vs mean 8.2, SD 2.9 days; P=.55; mean 15.2, SD 5.6 days vs mean 14.8, SD 6.8 days vs mean 15.9, SD 6.2 days; P=.70; respectively).
Table 2. Isolated pathogens from the first and second samples divided by groups.

<table>
<thead>
<tr>
<th>Genus</th>
<th>Group, n (%)</th>
<th>P-P\textsuperscript{a}</th>
<th>P-HAP\textsuperscript{b}</th>
<th>P-N\textsuperscript{c}</th>
<th>Total, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>8 (4.3)</td>
<td>15 (8.0)</td>
<td>6 (3.2)</td>
<td>29 (15.4)</td>
</tr>
<tr>
<td></td>
<td>Gram-positive bacteria</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>5 (2.7)</td>
<td>11 (5.9)</td>
<td>5 (2.7)</td>
<td>21 (11.2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3 (1.6)</td>
<td>2 (1.1)</td>
<td>1 (0.5)</td>
<td>6 (3.2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>16 (8.5)</td>
<td>28 (14.9)</td>
<td>12 (6.4)</td>
<td>56 (29.8)</td>
</tr>
<tr>
<td>Sample 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Staphylococcus</td>
<td>21 (11.2)</td>
<td>5 (2.7)</td>
<td>11 (5.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Streptococcus</td>
<td>6 (3.2)</td>
<td>1 (0.5)</td>
<td>5 (2.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Corynebacterium</td>
<td>56 (29.8)</td>
<td>12 (6.4)</td>
<td>28 (14.9)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>4 (2.7)</td>
<td>1 (0.5)</td>
<td>3 (2.0)</td>
<td>4 (2.7)</td>
</tr>
<tr>
<td>Sample 2</td>
<td></td>
<td>1 (0.7)</td>
<td>1 (0.7)</td>
<td>1 (0.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Staphylococcus</td>
<td>2 (1.4)</td>
<td>N/A\textsuperscript{d}</td>
<td>7 (4.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Streptococcus</td>
<td>1 (0.7)</td>
<td>N/A</td>
<td>2 (1.4)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Enterococcus</td>
<td>0 (0)</td>
<td>N/A</td>
<td>1 (0.7)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>7 (4.7)</td>
<td>7 (4.7)</td>
<td>N/A</td>
<td>14 (9.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gram-negative bacteria</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Klebsiella</td>
<td>17 (9.0)</td>
<td>6 (3.2)</td>
<td>8 (4.3)</td>
<td>31 (16.5)</td>
</tr>
<tr>
<td></td>
<td>Haemophilus</td>
<td>4 (2.1)</td>
<td>7 (3.7)</td>
<td>6 (3.2)</td>
<td>17 (9.0)</td>
</tr>
<tr>
<td></td>
<td>Escherichia</td>
<td>9 (4.8)</td>
<td>4 (2.1)</td>
<td>1 (0.5)</td>
<td>14 (7.4)</td>
</tr>
<tr>
<td></td>
<td>Enterobacter</td>
<td>9 (4.8)</td>
<td>3 (1.6)</td>
<td>2 (1.1)</td>
<td>14 (7.4)</td>
</tr>
<tr>
<td></td>
<td>Pseudomonas</td>
<td>10 (5.3)</td>
<td>0 (0)</td>
<td>1 (0.5)</td>
<td>11 (5.9)</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>23 (12.2)</td>
<td>7 (3.7)</td>
<td>7 (3.7)</td>
<td>37 (19.7)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>72 (38.3)</td>
<td>27 (14.4)</td>
<td>25 (13.3)</td>
<td>124 (66)</td>
</tr>
<tr>
<td>Sample 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Klebsiella</td>
<td>16 (10.8)</td>
<td>12 (8.1)</td>
<td>N/A</td>
<td>28 (18.9)</td>
</tr>
<tr>
<td></td>
<td>Pseudomonas</td>
<td>16 (10.8)</td>
<td>8 (5.4)</td>
<td>N/A</td>
<td>24 (16.2)</td>
</tr>
<tr>
<td></td>
<td>Escherichia</td>
<td>11 (7.4)</td>
<td>4 (2.7)</td>
<td>N/A</td>
<td>15 (10.1)</td>
</tr>
<tr>
<td></td>
<td>Enterobacter</td>
<td>8 (5.4)</td>
<td>6 (4.1)</td>
<td>N/A</td>
<td>14 (9.5)</td>
</tr>
<tr>
<td></td>
<td>Acinetobacter</td>
<td>10 (6.8)</td>
<td>3 (2.0)</td>
<td>N/A</td>
<td>13 (8.8)</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>18 (12.2)</td>
<td>16 (10.8)</td>
<td>N/A</td>
<td>34 (23.0)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>79 (53.4)</td>
<td>49 (33.1)</td>
<td>N/A</td>
<td>128 (86.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fungi</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Candida</td>
<td>1 (0.5)</td>
<td>6 (3.2)</td>
<td>1 (0.5)</td>
<td>8 (4.3)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>89 (47.3)</td>
<td>61 (32.4)</td>
<td>38 (20.2)</td>
<td>188 (100)</td>
</tr>
<tr>
<td>Sample 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Candida</td>
<td>5 (3.4)</td>
<td>1 (0.7)</td>
<td>N/A</td>
<td>6 (4.1)</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>91 (61.5)</td>
<td>57 (38.5)</td>
<td>N/A</td>
<td>148 (100)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}P-P group: patients with a persistent pathogen presence in week 2 tracheal aspirates.

\textsuperscript{b}P-HAP: patients with a different pathogen in week 2 tracheal aspirates.

\textsuperscript{c}P-N: patients with sterile week 2 tracheal aspirates.

\textsuperscript{d}N/A: not applicable.
Discussion

Principal Findings

We observed higher initial body temperatures in patients who survived to hospital discharge and a trend toward higher initial temperatures in patients who survived to ICU discharge. We did not observe any differences in body temperature between the P-N, P-HAP, and P-P groups of patients, but we did observe a significantly greater use of paracetamol in the P-N group of patients and a trend toward greater use of active cooling in the P-N group of patients. Additionally, we observed a trend toward better survival to ICU and hospital discharge in the P-N group of patients.

Traditionally, antipyretic therapy (mostly paracetamol), and in some cases active cooling, have been used to lower body temperature in patients who are febrile with infection. The presumed patient benefit comes from decreased metabolic demands associated with lower body temperature [7,8].

The survival of our patients is comparable to studies performed by Gursel and Demirtas [9], Depuydt et al [10], and Qiao et al [11] who reported survival to ICU discharge of patients who were critically ill with pneumonia in the range between 46% and 75%, compared to 65.8% survival to ICU discharge in our patients. They included patients with similar severity of illness, with SOFA scores in the range between 4 and 6 points and APACHE II scores around 20 points, compared to the SOFA score 10.1 (SD 2.6) and APACHE II score 21.7 (SD 6.6) in our study.

Our findings are in line with a number of other studies where better survival was observed in patients with infection and physiological-grade fever (ie, body temperature between around 37 °C and 39 °C). Both Kushimoto et al [12] in a prospective observational study and Shimazui et al [13] in a retrospective observational study observed higher mortality in patients with sepsis and lower body temperature on admission. Lee et al [14] also observed lower mortality in patients with physiological-grade fever in their prospective study. Rumbus et al [15] performed a meta-analysis of 42 studies reporting body temperature and mortality in patients with sepsis. They discovered a correlation between higher body temperature and lower mortality; higher body temperature predicted better survival, and hypothermia predicted lower survival. Similar to our findings the number of patients with body temperature over 39.5 °C was very low. Thomas-Rüddel et al [16] performed a secondary analysis of a large data set of patients with sepsis and discovered that initial body temperatures were distributed in two peaks: a smaller peak at around 35.5 °C and a larger peak approximately twice as large at around 38 °C. Again, the highest survival rate was observed in patients with hyperthermia and the lowest in patients with hypothermia. They also observed that ambient temperatures were significantly associated with body temperatures; lower ambient temperatures were associated with hypothermia, and higher outside temperatures were associated with hyperthermia. The potential benefits of active warming of patients with infection were further highlighted by Drewry et al [4], who performed a pilot randomized controlled study evaluating the use of therapeutic hyperthermia in patients with sepsis and observed a significantly better survival in the hyperthermia group; however, there were no differences in the primary results.

We can speculate that higher doses of paracetamol and greater use of active cooling probably decreased body temperatures in the P-N group of patients; however, this was not associated with any improvement during ICU treatment. We observed no differences in the use of noradrenalin, parameters of mechanical ventilation, requirement for renal replacement therapy, or organ dysfunction scores, and we observed a trend toward greater ICU and hospital survival in the P-N group. Our results suggest that there is probably no clinical benefit associated with the treatment of physiological-grade fever. Similarly, Zhang et al [17] and Ye et al [18] performed retrospective studies on patients with fever and sepsis, and they observed no beneficial effects of antipyretic therapy and possible harm associated with the use of external cooling. In a prospective randomized controlled trial, Young et al [19] observed approximately 0.5 °C lower body temperature in patients who received paracetamol (4 g daily) compared to patients who received a placebo. They also observed that patients who received more paracetamol experienced a longer ICU stay if they were nonsurvivors, and shorter ICU stay if they were survivors, which was explained by the probable effect of lower body temperature on clinicians’ perception of the patients’ prognosis. Lower body temperature in the paracetamol group was not associated with any improvement in the ICU course of treatment parameters [19,20].

To our knowledge, there is no other data to compare our results regarding the clearance or persistence of pathogens in tracheal aspirates. In 42.7% (n=50) of our patients (ie, the P-P group of patients), the same pathogen persisted in tracheal aspirates 1 week after initial samples were obtained. There were no differences in causative pathogens between the P-N, P-HAP, and P-P groups of patients, and we observed no significant changes in baseline data between the groups, including the rate of change of empiric antibiotic therapy. There was a trend toward higher temperature in the first 48 hours in the P-N group, compared to the P-HAP and P-P groups, possibly indicating that tolerance of hyperthermia could be beneficial for patients who are intubated with pneumonia and sepsis. There were no significant differences in the course of ICU treatment parameters despite greater use of paracetamol and active cooling, indicating that pharmacological or active cooling has no benefit in patients where fever is part of an appropriate response to infection.

There are a number of limitations to our study. We performed a single-center retrospective observational study with all inherent biases associated with this study design. Of the initial 488 patients with pneumonia, 249 were not included because there were no tracheal aspirates 1 week after ICU admission (because of extubation, death, or no clinical need for obtaining samples if the patients were still intubated). Additionally, target temperatures were defined by the treating physicians. However, there are no guidelines regarding target temperatures for this patient population, and we observed no differences in baseline data between the different groups.
Conclusions
To conclude, we observed better survival in patients who developed higher body temperatures in the first 48 hours after admission to the ICU; however, we observed no changes in other treatment parameters. Additionally, we observed greater use of interventions aimed at cooling the patients (use of paracetamol and a trend toward greater use of active cooling) in patients with negative tracheal aspirates 1 week after ICU admission. Additionally, in this group of patients, we observed a trend toward better survival. Our results speak against the use of interventions aimed at the reduction of body temperature and support the strategy of temperature tolerance in patients who are mechanically ventilated with pneumonia and sepsis.

Acknowledgments
The study was supported by an institutional research grant (grant IRP-2022/01-01).

Authors’ Contributions
AM, JF, and ŽK conceptualized the study. AM and PPB developed the methodology. PPB used the software. AM and PPB performed the validation. PPB performed the formal analysis. DG, AG, and NG conducted the investigation and acquired the resources. AM curated the data curation. DG, AG, and NG prepared the original draft. AM, PPB, JF, ŽK, DG, AG, and NG reviewed and edited the paper. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest
None declared.

References


**Abbreviations**

APACHE: acute physiology and chronic health evaluation  
ICU: intensive care unit  
P-HAP: patients with a different pathogen in week 2 tracheal aspirates  
P-N: patients with sterile week 2 tracheal aspirates  
P-P: patients with a persistent pathogen presence in week 2 tracheal aspirates  
SOFA: sequential organ failure assessment

©Domen Guzelj, Anže Grubelnik, Nina Greif, Petra Povalej Bržan, Jure Fluher, Žiga Kalamar, Andrej Markota. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 01.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.
Evaluation of the Accuracy, Credibility, and Readability of Statin-Related Websites: Cross-Sectional Study

Eunice Ling¹, BAppSc, MD; Domenico de Pieri¹*, MD; Evenne Loh¹*, MD; Karen M Scott¹, BEd, MA, PhD; Stephen C H Li², MBBS, MSci, MPH, MBA, MAACB; Heather J Medbury¹,³, BSc, PhD

¹Sydney Medical School, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia
²Institute of Clinical Pathology and Medical Research, Westmead Hospital, Westmead, Australia
³Vascular Biology Research Centre, Surgery, Westmead Hospital, Westmead, Australia
*these authors contributed equally

Corresponding Author:
Heather J Medbury, BSc, PhD
Vascular Biology Research Centre
Surgery
Westmead Hospital
REN Building
Westmead, 2145
Australia
Phone: 61 8890 3668
Email: heather.medbury@sydney.edu.au

Abstract

Background: Cardiovascular disease (CVD) represents the greatest burden of mortality worldwide, and statins are the most commonly prescribed drug in its management. A wealth of information pertaining to statins and their side effects is on the internet; however, to date, no assessment of the accuracy, credibility, and readability of this information has been undertaken.

Objective: This study aimed to evaluate the quality (accuracy, credibility, and readability) of websites likely to be visited by the general public undertaking a Google search of the side effects and use of statin medications.

Methods: Following a Google web search, we reviewed the top 20 consumer-focused websites with statin information. Website accuracy, credibility, and readability were assessed based on website category (commercial, not-for-profit, and media), website rank, and the presence or absence of the Health on the Net Code of Conduct (HONcode) seal. Accuracy and credibility were assessed following the development of checklists (with 20 and 13 items, respectively). Readability was assessed using the Simple Measure of Gobbledegook scores.

Results: Overall, the accuracy score was low (mean 14.35 out of 20). While side effects were comprehensively covered by 18 websites, there was little information about statin use in primary and secondary prevention. None of the websites met all criteria on the credibility checklist (mean 7.8 out of 13). The median Simple Measure of Gobbledegook score was 9.65 (IQR 8.825-10.85), with none of the websites meeting the recommended reading grade of 6, even the media websites. A website bearing the HONcode seal did not mean that the website was more comprehensive or readable.

Conclusions: The quality of statin-related websites tended to be poor. Although the information contained was accurate, it was not comprehensive and was presented at a reading level that was too difficult for an average reader to fully comprehend. As such, consumers risk being uninformed about this pharmacotherapy.

(KEYWORDS
statins; consumer health information; readability; credibility; accuracy; digital health, health information seeking; cardiovascular; mortality; management; pharmacotherapy; risk; medication

**Introduction**

**Background**

Cardiovascular diseases (CVDs) are the primary cause of death globally, with an estimated 17.9 million people dying of CVDs in 2021. This represents 31% of all global deaths. Of these deaths, 85% are due to heart attack and stroke, whose most common etiology is atherosclerosis [1]—the development of fatty plaque within artery walls. A key pharmacological treatment for atherosclerosis is statin therapy. It has a role in the primary and secondary prevention of vascular events, with a lowering of low-density lipoprotein cholesterol leading by 2 mmol/L, typically giving a 10% absolute benefit (the reduction in the probability of an event’s occurrence within a population receiving treatment) for those diagnosed with vascular disease and a 5% absolute benefit for those with risk factors yet without having experienced a vascular event [2]. This creates issues when we consider that patients may be biochemically abnormal (with hypercholesterolemia) but asymptomatic. Such patients may doubt the use of the prescribed statin therapy as they determine the cost-benefit balance between tangible adverse effects and theoretical benefits. This may prompt the consultation of alternative sources of knowledge to aid decision-making. In this era of shared decision-making, where patients participate in the medical decisions that affect their health [3], it is essential that the information they access is high quality and easily understood.

**Use of the Internet for Health Information Seeking**

In this milieu, the internet has risen as a key source of health-related information, with 79% of adults seeking web-based health information in America and 79% to 86% in China, the Philippines, Hong Kong, Indonesia, and Vietnam [4,5]. Comparatively, seeking web-based health information is as popular as playing games or downloading music from the web [6]. Notably, the COVID-19 pandemic has presented unprecedented challenges, catapulting society further into a future dependence on telehealth and internet-assisted health care [7,8]. As such, traditional in-clinic and leaflet modes of health information delivery are being supplanted, and in some cases supplanted, by internet searches. With this dramatic change in the terrain upon which patients and their families are attaining information, it is crucial to determine the quality of web-based health information put forth to them.

Health literacy is defined as “the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions” [9]. It requires a complex group of skills such as reading, listening, analyzing, and decision-making, as well as the ability to apply the aforementioned skills to health situations [10]. Those with poor health literacy are vulnerable to undertaking unnecessary tests and treatments or, conversely, refusing beneficial tests and treatments. In part, they may be misled into assessing the quality of web-based health information based on its search result ranking, image quality, celebrity endorsement, and website authorship rather than relying on the criteria of established quality guidelines [11,12].

Health literacy–related knowledge and skills are particularly deficient among vulnerable populations, who are also more likely to experience CVD [12,13]. Unfortunately, those with poor health literacy are susceptible to the influence of mass media and emotionally persuasive texts. This may explain the response to the television program “Catalyst” in Australia [14], whose criticism of statins resulted in 11% of patients (in a survey by the Australian National Heart Foundation of 1094 patients) who watched the program ceasing to take their cholesterol-lowering medication and significant and sustained changes in statin usage, with 2.6% fewer statins (equivalent to 14,005 dispensing) each week [15].

**Internet Standards**

There is considerable heterogeneity in the quality of web-based health information [16]. The quality of information can be examined in 3 domains: accuracy, credibility, and readability [17]. Each is defined as follows: accuracy is the intent to be evidence based and safe by adequately offering a complete, unbiased picture and its relevance [17]; credibility is the attribution of source and authorship and the disclosure of conflicts of interest for the presented information [18]; and readability is the ease of understanding due to the style of writing, describing the reading comprehension level a layperson requires to understand a text [19,20]. Complex wording reduces engagement with, and application of, content [21]. This leaves patients vulnerable to becoming ill-informed and at risk of adverse health outcomes [22]. The quality of web-based health information is, therefore, a pressing issue, which the Health on the Net Foundation aims to address by providing a Health on the Net Code of Conduct (HONcode) seal, an internet-based certification of medical and health websites that adheres to a set of publishing principles regarding the source and purpose of medical information. However, few consumers are aware of Health on the Net or the HONCode.

While there have been assessments of the quality of websites on many key areas of health care, such as diabetes [23], obesity [24], and hypertension [25], and on surgical interventions [26,27], there is a dearth of assessment about the quality of web-based information pertaining to medications. Such assessments are necessary to inform clinicians of the quality of content likely to be accessed by their patients, who are particularly interested in the likelihood and nature of adverse side effects. We aimed to assess the quality of consumer health information on websites about statins through consideration of accuracy, credibility, and readability.

**Methods**

**Study Design**

**Overview**

Through this cross-sectional study, we analyzed the accuracy, credibility, and readability of websites that were most commonly presented to patients searching for the keywords “statin” and “statin side effects.” We determined each website’s search engine ranking, category (commercial, not-for-profit, or media), and the presence or absence of the HONcode seal. Furthermore,
we determined the relationship between the accuracy, credibility, and readability of the websites found.

In selecting websites on statins to analyze, we aimed to emulate a typical consumer’s search for web-based health information. A web search was conducted using Google, in keeping with evidence that it accounts for more than half of all web traffic [28-30] and is an increasingly preferred search engine by the general public [31]: 91% of American adults using the internet use a search engine, and of those, 83% use Google more often than other search engines [4,32]. To conduct the search, location filters, user information, search history, cached data, and cookies were disabled, and sponsored results were excluded to avoid inadvertent search bias. The search terms used were “statin” and “statin side effects,” following the advice of our lipidologist coauthor (SL) that the generic term “statin,” rather than specific medication names, was commonly used in discussion with patients in clinical practice and that “statin side effects” were a key concern of patients.

The first 20 ranked websites in the Google search results page were analyzed (after removing any duplication from search results of the 2 search terms). We did not identify further websites in the search results, given that, in general, websites returned on the first Google search results page generate 92% of all traffic from an average search [16,33]. This drops by 95% for the second page and by 78% and 58% for subsequent pages [31]. Thus, we did not aim to identify all websites on statins but rather to emulate an authentic consumer search.

**Website Search Rank**

The effect of the association between search result ranking and accuracy, credibility, and readability was considered. Given that engagement is highest with the first 5 websites in search results, garnering 67% of all clicks from an average search [16,33]. This drops by 95% for the second page and by 78% and 58% for subsequent pages [31]. Thus, we did not aim to identify all websites on statins but rather to emulate an authentic consumer search.

**Website Categorization**

In the interest of determining whether the nature of the authorship of the websites had a bearing on their accuracy, credibility, and readability, each of the 20 chosen websites was categorized into 3 types: commercial (defined as a website that generates revenue or cash and is not affiliated with the government), not-for-profit (a website that garners support for a cause rather than revenue, including government and charities), and media (a website that reports new findings or stories, with the primary purpose of the website being news reporting).

**Presence of HONcode Seal**

We assessed whether the quality of the websites was associated with the presence or absence of the HONcode seal.

**Measures**

**Overview**

In assessing the accuracy, credibility, and readability of the websites, we considered existing tools and developed study-specific tools where necessary.

**Accuracy**

For accuracy, we developed a statin-specific tool that took into account medical guidelines. The checklist (Table 1) was designed after referring to other studies purporting to assess the quality of web-based health information [17]. A key difference here was that we were looking at a specific treatment. Three features were considered: (1) the intent to be evidence based; (2) safety, in that a website should adequately offer users a complete, unbiased picture of statin treatment; and (3) relevance, in that it is reasonable to expect a website to address the criteria in the checklist [17]. The checklist was intrinsically linked to a website’s comprehensiveness, consistent with other studies that have evaluated completeness as an integral part of accuracy [16].

To meet these features, guidelines from the American College of Cardiology and the Australian Heart Foundation [37] were synthesized into short statements, which formed the accuracy criteria. These statements formed a checklist that each website was required to address to be considered “accurate.” The development of the statements was further informed and determined by a review of the treatment of cholesterol in light of its evidence base [38], as well as criteria from the treatment section of the validated DISCERN tool [39]. This section (items 9-16 of the DISCERN tool) addresses issues of risk, benefit, and how the website guides decision-making surrounding treatment options [39]. Combining these sources ensured that a higher score would be awarded to websites providing the most evidence-based information. In total, 20 equally weighted criteria were devised, and a score of 20 was arbitrarily defined as a minimum acceptable standard.

Each accuracy item was scored as: “present and complete” (2), “present but incomplete” (1), “absent” (0), or “inaccurate” (*). A maximum score of 40 could be awarded to each website assessed. “Incompleteness” was defined as a nonexact or indirect mention of a topic outlined in a criterion rather than an explicit statement. Two reviewers (DdP and E loh) completed the assessment.

Any hyperlinks that navigated to information within a website were followed and the data were included in the final assessment; links leading to external websites were not followed. Embedded videos were analyzed. Once each reviewer concluded their analysis, the results were compared. Discrepancies were resolved through discussion until reaching a consensus.
Table 1. Website accuracy checklist.

<table>
<thead>
<tr>
<th>Accuracy criteria</th>
<th>Descriptor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Mentions that cholesterol is a modifiable risk factor for cardiovascular disease [35]</td>
</tr>
<tr>
<td>2</td>
<td>Mentions that consultation with a doctor is essential before and while taking statins and when ceasing them [36]</td>
</tr>
<tr>
<td>3</td>
<td>Lists conditions for which statins are used [36]</td>
</tr>
<tr>
<td>4</td>
<td>Defines the target population for statin therapy [37]</td>
</tr>
<tr>
<td>5</td>
<td>Mentions the importance of adherence to statin therapy [36]</td>
</tr>
<tr>
<td>6</td>
<td>Addresses the subtleties of primary prevention [38]</td>
</tr>
<tr>
<td>7</td>
<td>Mentions that statins are about reducing complications of high cholesterol rather than achieving a specific (low-density lipoprotein) cholesterol [38]</td>
</tr>
<tr>
<td>8</td>
<td>Describes or at least lists the benefits of statin therapy [39]</td>
</tr>
<tr>
<td>9</td>
<td>Describes or at least lists the side effects or risks of statin therapy [39]</td>
</tr>
<tr>
<td>10</td>
<td>Describes how treatment affects the overall quality of life [39]</td>
</tr>
<tr>
<td>11</td>
<td>Mentions low to moderate dose statin therapy is recommended in primary prevention [37]</td>
</tr>
<tr>
<td>12</td>
<td>Specifically addresses rhabdomyolysis [36]</td>
</tr>
<tr>
<td>13</td>
<td>Describes the approximate financial burden to the patient [36]</td>
</tr>
<tr>
<td>14</td>
<td>Describes the duration of treatment before an effect is measurable [39]</td>
</tr>
<tr>
<td>15</td>
<td>Describes how statins work or at least what they do [39]</td>
</tr>
<tr>
<td>16</td>
<td>Describes what may happen without treatment [39]</td>
</tr>
<tr>
<td>17</td>
<td>Explores the possibility of using alternative therapies to statins [39]</td>
</tr>
<tr>
<td>18</td>
<td>Mentions that statins must not be used during pregnancy [37]</td>
</tr>
<tr>
<td>19</td>
<td>Describes drug interactions or at least lists them [39]</td>
</tr>
<tr>
<td>20</td>
<td>Mentions that statins do not replace a healthy lifestyle [36]</td>
</tr>
</tbody>
</table>

Credibility

In developing the criteria to be included in the assessment of credibility (Table 2), DISCERN was chosen as a reference, as well as other studies that used DISCERN or another available tool for website assessment. However, as the 5-point Likert scale used in DISCERN can be subject to response style bias [40], a present (1) or absent (0) scale was adopted as it has been shown to improve the objectivity of data collection [41-43].
Table 2. Website credibility checklist.

<table>
<thead>
<tr>
<th>Item</th>
<th>Criteria</th>
<th>Score criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Referencing or citations obtained from peer-reviewed journals</td>
<td>1 point if the articles for which the references are obtained are published in peer-reviewed journals [16,44]</td>
</tr>
<tr>
<td>2</td>
<td>Website updated within last 24 months</td>
<td>The latest update should be within the past 24 months [45]</td>
</tr>
<tr>
<td>3</td>
<td>Avoids anecdotal evidence for making claims</td>
<td>Does not use anecdotal evidence as a basis for claims; quoting a case study without using claims is acceptable [46]</td>
</tr>
<tr>
<td>4</td>
<td>Mailing address present</td>
<td>Physical contact address of the website clearly stated [45]</td>
</tr>
<tr>
<td>5</td>
<td>Contact information available</td>
<td>Contact information including name, position, telephone number, address, and email [47]</td>
</tr>
<tr>
<td>6</td>
<td>Sponsorship stated</td>
<td>Any sponsorship should be clearly stated</td>
</tr>
<tr>
<td>7</td>
<td>Organizational privacy policy stated</td>
<td>Organization privacy policy should be clearly stated [47]</td>
</tr>
<tr>
<td>8</td>
<td>Declaration of the author’s qualification</td>
<td>Author’s qualification should be health care related [16,44]</td>
</tr>
<tr>
<td>9</td>
<td>Paid access tab present</td>
<td>If paid access is available, the difference in the information obtained from paid vs unpaid access should be clearly stated [48]</td>
</tr>
<tr>
<td>10</td>
<td>Disclosure of funding or conflicts of interest</td>
<td>Conflicts of interest and funding disclosure should be clearly stated [44]</td>
</tr>
<tr>
<td>11</td>
<td>The presence of an HONcode seal or third-party certification</td>
<td>Presence of a HONcode(^a) seal or any other third-party certification [16]</td>
</tr>
<tr>
<td>12</td>
<td>Advertisement neutral</td>
<td>Advertisements should steer clear from the website information (eg, no pop-ups related to the website content) [16]</td>
</tr>
<tr>
<td>13</td>
<td>Disclaimer regarding web-based health information</td>
<td>A disclaimer should be clearly stated that web-based health information does not replace a practitioner’s advice [44]</td>
</tr>
</tbody>
</table>

\(^a\)HONcode: Health on the Net Code of Conduct.

Each website was appraised according to this list. A score was allocated for each website’s front page, with internal links explored only if relevant. Data (credibility scores) were undertaken as independent assessments by 2 assessors (E Loh and DdP). The results were compared, and if discrepancies arose, discussions were held to clarify the score, with external input from advisors (HM, SL, and KS) obtained where appropriate.

Readability

For readability, various tools are available, including the Flesch Kincaid Reading Ease, Flesch Kincaid Grade Level, Simple Measure of Gobbledygook (SMOG), and Average Grade Level. We used SMOG as it is considered the gold standard for assessing the readability of health care material and has a high correlation with the other scoring systems [24,49]. Importantly, the outcome measure is easy to understand as, for example, an SMOG readability grade of 6 represents a text comprehensible to all individuals with sixth-grade reading skills and above [50-53]. This grade level was set as the basis of readability, given that the available literature sets this as the standard for “superior” readability. To use this tool, texts from the 20 selected websites were copied and saved as separate Microsoft Word (Microsoft Corp) and plain text documents for analysis, deleting text unrelated to the health information topic (eg, author information or disclaimers) to prevent this from confounding the scoring. A single web-based readability calculator [54] was used to generate the scores.

Data Analysis

The website category and ranking findings were compared by ANOVA, and differences between websites with and without the HONcode seal were analyzed with 2-tailed \(t\) tests. In addition, the relationship between credibility and readability with accuracy was assessed by Pearson correlation.

Ethical Considerations

As the research was not conducted on human subjects, no ethics review was required.

Results

Selected Websites

The top 20 websites returned by the search are listed in Multimedia Appendix 1 [55-74]. Of the 20 websites chosen from the search, 45\% (n=9) were categorized as commercial, 45\% (n=9) not-for-profit, and 10\% (n=2) media (Table 3). There was an even distribution of commercial and not-for-profit websites across the 4 quartiles, with both media websites found in the fourth quartile. Eight of the websites bore the HONcode seal.
### Table 3. Top 20 statin websites’ category, HONcode\textsuperscript{a} presence or absence, accuracy, credibility, and readability.

<table>
<thead>
<tr>
<th>Website rank</th>
<th>Category</th>
<th>HONcode Seal</th>
<th>Accuracy</th>
<th>Credibility</th>
<th>Readability (SMOG\textsuperscript{b} score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Commercial</td>
<td>Yes</td>
<td>20</td>
<td>6</td>
<td>10.7</td>
</tr>
<tr>
<td>2</td>
<td>Not-for-profit</td>
<td>Yes</td>
<td>9</td>
<td>11</td>
<td>11.3</td>
</tr>
<tr>
<td>3</td>
<td>Not-for-profit</td>
<td>No</td>
<td>23</td>
<td>9</td>
<td>12.5</td>
</tr>
<tr>
<td>4</td>
<td>Commercial</td>
<td>Yes</td>
<td>16</td>
<td>6</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td>Commercial</td>
<td>No</td>
<td>10</td>
<td>4</td>
<td>7.7</td>
</tr>
<tr>
<td>6</td>
<td>Not-for-profit</td>
<td>No</td>
<td>15</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>7</td>
<td>Commercial</td>
<td>No</td>
<td>14</td>
<td>10</td>
<td>9.6</td>
</tr>
<tr>
<td>8</td>
<td>Not-for-profit</td>
<td>No</td>
<td>13</td>
<td>5</td>
<td>10.8</td>
</tr>
<tr>
<td>9</td>
<td>Not-for-profit</td>
<td>No</td>
<td>16</td>
<td>11</td>
<td>10.2</td>
</tr>
<tr>
<td>10</td>
<td>Commercial</td>
<td>Yes</td>
<td>11</td>
<td>10</td>
<td>8.6</td>
</tr>
<tr>
<td>11</td>
<td>Commercial</td>
<td>Yes</td>
<td>16</td>
<td>9</td>
<td>9.4</td>
</tr>
<tr>
<td>12</td>
<td>Commercial</td>
<td>No</td>
<td>13</td>
<td>6</td>
<td>11</td>
</tr>
<tr>
<td>13</td>
<td>Not-for-profit</td>
<td>No</td>
<td>17</td>
<td>9</td>
<td>8.9</td>
</tr>
<tr>
<td>14</td>
<td>Commercial</td>
<td>No</td>
<td>8</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>15</td>
<td>Not-for-profit</td>
<td>No</td>
<td>8</td>
<td>7</td>
<td>10.3</td>
</tr>
<tr>
<td>16</td>
<td>Media</td>
<td>No</td>
<td>8</td>
<td>4</td>
<td>7.8</td>
</tr>
<tr>
<td>17</td>
<td>Not-for-profit</td>
<td>Yes</td>
<td>16</td>
<td>11</td>
<td>6.7</td>
</tr>
<tr>
<td>18</td>
<td>Commercial</td>
<td>Yes</td>
<td>19</td>
<td>8</td>
<td>9.7</td>
</tr>
<tr>
<td>19</td>
<td>Commercial</td>
<td>Yes</td>
<td>21</td>
<td>7</td>
<td>9.2</td>
</tr>
<tr>
<td>20</td>
<td>Media</td>
<td>No</td>
<td>14</td>
<td>2</td>
<td>8.1</td>
</tr>
</tbody>
</table>

\textsuperscript{a}HONcode: Health on the Net Code of Conduct.
\textsuperscript{b}SMOG: Simple Measure of Gobbledygook.

### Accuracy
The mean website accuracy score was 14.35 (SD 4.43). In terms of accuracy, the 3 highest-scoring websites were Wikipedia (score of 23), Drugs.com (score of 21), and Medicine.net (score of 20). These were the only websites to achieve a score of 20 or above. No website contradicted any checklist criterion. The top 3 performing checklist criteria were related to side effects and statin mechanism of action (criteria 9, 12, and 15), with a score of “present and complete” for each of these criteria achieved by 18, 14, and 12 websites, respectively. Although side effects were covered to some degree in all websites, criteria about drug safety (criteria 18 and 19) were complete in only 8 and 7 websites, respectively. Other poorly performing criteria reflected the lack of detail about primary prevention (criteria 11 and 6), with a score of “absent” assigned to 19 and 17 of the websites, respectively.

### Credibility
None of the sampled websites met all credibility criteria for a perfect score of 13. The mean score overall was 7.45, with a range of 2-11. Importantly, 12 websites referenced peer-reviewed journal articles as a source of information, and 15 avoided anecdotal evidence for making claims. Media and some commercial websites reported personal opinions. While only 6 websites provided an organization’s contact details, the others provided an email address or feedback form for contact purposes. Sponsorship was explicitly stated in 12 websites, with reference to either government or private organizations. All websites declared their organization’s privacy policy, including websites with lower overall credibility scores. Only 8 websites declared author qualifications, which were primarily health related. None of the websites required paid access. Funding sources were fully disclosed in 11 websites, with the remaining 9 not reporting their source of funding or conflicts of interest. Twelve websites either had no advertisements or non–health care advertisements; the 8 websites that did not meet this criterion were commercial or media websites. Only 5 websites did not include a disclaimer that web-based health information does not replace a practitioner’s advice: all of these websites were commercial or media websites.

### Readability
Overall, for SMOG readability, the median was 9.65 (IQR 8.825-10.85) and the average was 9.875 (SD 1.75), that is, above the ninth-grade level. None of the websites met the recommended grade level of 6; even the media websites required an eighth-grade level of comprehension.

---

https://www.i-jmr.org/2024/1/e42849

Interact J Med Res 2024 | vol. 13 | e42849 | p.77

(page number not for citation purposes)
Correlation Between Accuracy, Credibility, and Readability

No significant correlation was evident between the correlation between credibility and accuracy ($P=.23$) (Figure 1).

**Figure 1.** Relationship between website credibility (black triangles) and correlation readability (blue squares) with accuracy ($P=.23$).

**Website Search Rank**

Websites that featured prominently in search results were not necessarily the most accurate, with no significant difference between the quartiles ($P=.64$). Indeed, of the 20 websites reviewed, the second highest scoring website for accuracy was ranked in the 19th position on search results. Similarly, there was no difference in credibility ($P=.63$) or readability ($P=.06$) between the quartiles.

**Website Categorization**

Comparing commercial with not-for-profit websites, 2-tailed $t$ tests revealed there was no significant difference in terms of accuracy ($P=.275$), credibility ($P=.83$), or readability ($P=.452$). As there were only 2 media websites, a comparison with them was not made. Notably, they had the lowest scores for credibility, but both scored among the most readable.

**Presence of HONcode Seal**

Of the 20 websites, 8 were HONcode certified, with 6 of these categorized as commercial websites and 2 as not-for-profit. The mean accuracy scores for websites with and without the HONcode seal were 16 (SD 4.2) and 13.25 (SD 4.4), respectively, but this was not significantly different ($P=.18$). The presence or absence of the HONcode seal did not preclude a website from scoring at either end of the accuracy scale. Although the 8 websites with the HONcode seal scored higher in credibility (mean 8.5, SD 2) than websites that were not HONcode certified (mean 6.75, SD 2.7), this was not significantly different ($P=.139$). There was no significant difference in readability scores ($P=.92$) when comparing websites with HONcode seal status or lack thereof (mean 9.83, SD 1.9 and mean 9.91, SD 1.7, respectively).

**Discussion**

This study found that overall, the quality of websites with statin-related information tended to be poor. The website content was not sufficiently comprehensive, and the reading level was too difficult for the average reader to fully comprehend. The credibility of the websites varied, although overall websites bearing the HONcode seal had higher credibility than those without.

Here, we formally assessed the quality of websites addressing statins and their side effects. The finding that the quality of information is of variable caliber is consistent with studies investigating web-based health information on other topics [4,17,75]. Although the criteria used by Google’s ranking algorithm is confidential, Google’s guidelines state that it uses a series of algorithms that account for the words of the query, relevance and usability of web pages, the expertise of sources, ease of use on mobile device interfaces, as well as location and settings to determine the results displayed [76,77]. However, this study demonstrates that the most prominent websites in the Google search ranking are not necessarily of high quality.

The lack of correlation between accuracy and credibility or readability is a concern if patients are using the information to understand their condition and take action related to it. Patients with poor health literacy may use inaccurate and untrustworthy information in deciding whether to see a health professional following the onset of symptoms or whether to undertake tests and treatments that may be unnecessary or recommended by health professionals [12]. Vulnerable populations are at higher risk of having poor health literacy and experiencing CVD.
[12,13], making them especially vulnerable to inaccurate, untrustworthy, and unreadable websites on statins.

While most websites analyzed as part of this study scored low in accuracy, this tended to be attributed to a lack of completeness of information rather than a lack of factual information. While the checklist developed here may be stringent, it would be reasonable to expect that websites dedicated to statins would be comprehensive. The lack of comprehensiveness in the information provided on the websites could result in consumers overlooking important details unless they browse through multiple websites. Furthermore, visits to multiple websites may not generate clarity but confusion. This is due to the increased likelihood of encountering inconsistent information, particularly as websites have different agendas based on the website type. That said, the commercial websites scored, on average, just as well as the not-for-profit websites, indicating that they can be a valuable source of information for consumers. It also indicates that government and other not-for-profit websites will be required to at least match the accuracy of commercial websites if they are to remain relevant in Google’s search algorithm, as having information-rich content is a factor that contributes to higher search rankings [78].

When browsing the internet, one would expect government and other not-for-profit websites to provide credible information. However, some of these websites returned relatively low credibility scores and overall were not significantly more credible than commercial websites. Over half of the 20 websites analyzed provided evidence-based information and avoided anecdotal evidence, increasing their credibility rating [16,44]; however, it was concerning that 5 websites provided information based on anecdotal evidence. As expected, media websites received low credibility scores as news articles about statins contained personal views and anecdotes. Other indicators of credibility were lacking by a large proportion of websites, in particular author qualifications and details about sponsorship, funding, and conflicts of interest [16,44]. Furthermore, many of the commercial and media websites included advertisements, including health-related advertisements [16], and 5 of them did not include the disclaimer that web-based health information does not replace a practitioner’s advice [44]. Thus, even patients with a degree of health literacy would find it difficult to accurately appraise the credibility of many of these websites on statins.

Given that the general public is unlikely to be fully equipped to gauge the credibility of web-based health information presented [12], clinicians could advise that patients identify the presence of the HONcode seal as this merits some confidence in the information presented [16]. However, the code does not necessarily imply that websites are comprehensive. Additionally, website developer application for the HONcode seal is voluntary, so high-quality websites may not bear the HONcode seal. The finding that the readability of the websites with the HONcode seal was not at a suitable level means that such websites may not represent digestible patient health information. Furthermore, a practical issue is that the HONcode seal is at the bottom of the web page and is thus not necessarily evident at first glance.

Many patient demographic groups have been found to read at a level more than 3 years below their completed educational years [79]. Thus, the study results may not be indicative of the severity of the problem posed by websites with high readability scores in terms of the general public’s understanding of web-based information [79]. Those with limited literacy skills tend to have poorer health status due to a lack of knowledge and understanding of health care issues and a diminished ability to participate in shared decision-making in the clinical context [80]. They also tend to have poorer compliance with treatment recommendations and subsequent disease progression, as well as a higher risk for seeking emergency care and more frequent and longer inpatient admissions [27].

Additionally, other factors besides readability play into the way a text is received, including logical and sequential presentation of information. Additionally, alternative media such as images and graphs provide a well-documented “picture superiority” effect that boosts understanding of and engagement with a text [81], although some of these may also require interpretation by consumers.

A limitation of this study of these websites is that the internet is dynamic, with websites updated at any time. The search used in undertaking the study is constrained temporally in its noniterative nature, as well as its method, which used only the major search engine Google. While a metasearch capturing results from multiple search engines would provide a more comprehensive view of the information about statins on the internet, it is unlikely to represent the behavior of the public [82]. Additionally, only 8 of the websites in the study were updated in some way after completion of this study, and the information on some websites is dated as more than 10 years old.

Overall, this study has demonstrated that within the surfeit of information available on the internet regarding statin therapy, the quality of websites is of mixed caliber. The content of information is generally accurate but incomplete, while credibility is variable. Readability is generally of a level too difficult for the general public to comprehend. This suggests a need for guidance to website developers of health care websites in order to capitalize on the vast potential of the internet to equip patients with the empowerment of improved health information and health literacy. It also highlights that clinicians will need to be educated themselves about what is on the internet and what constitutes accuracy, credibility, and readability in order to impart this knowledge to their patients. During the COVID-19 pandemic, the methods through which patients seek information about their health have shifted toward increasingly internet-based means, making the quality of information on the internet of particular significance in the current climate and for the foreseeable future.
Acknowledgments
The findings were originally developed for the award of a Doctor of Medicine. Publication charges were funded by New South Wales (NSW) Health Pathology.

Authors’ Contributions
E Ling attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. The project was planned by HM with input from all authors. Checklists were developed by E Loh and DdP with the assistance of SL and KS. Data were collected by E Ling, E Loh, and DdP. Data analysis was done by E Ling, E Loh, and DdP with assistance from HM. Data interpretation and manuscript review were done by all authors.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Top 20 websites analyzed in the study.

References
76. How search algorithms work mountain view. Google. URL: https://www.google.com/search/howsearchworks/algorithms/ [accessed 2023-12-29]

Abbreviations

CVD: cardiovascular disease
HONcode: Health on the Net Code of Conduct
SMOG: Simple Measure of Gobbledygook

©Eunice Ling, Domenico de Pieri, Evenne Loh, Karen M Scott, Stephen C H Li, Heather J Medbury. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 14.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.
A Preliminary Checklist (METRICS) to Standardize the Design and Reporting of Studies on Generative Artificial Intelligence–Based Models in Health Care Education and Practice: Development Study Involving a Literature Review

Malik Sallam¹,²,³, MD, PhD; Muna Barakat⁴, PhD; Mohammed Sallam⁵, MSc, PharmD

¹Department of Pathology, Microbiology and Forensic Medicine, School of Medicine, The University of Jordan, Amman, Jordan
²Department of Clinical Laboratories and Forensic Medicine, Jordan University Hospital, Amman, Jordan
³Department of Translational Medicine, Faculty of Medicine, Lund University, Malmo, Sweden
⁴Department of Clinical Pharmacy and Therapeutics, Faculty of Pharmacy, Applied Science Private University, Amman, Jordan
⁵Department of Pharmacy, Mediclinic Parkview Hospital, Mediclinic Middle East, Dubai, United Arab Emirates

Corresponding Author:
Malik Sallam, MD, PhD
Department of Pathology, Microbiology and Forensic Medicine
School of Medicine
The University of Jordan
Queen Rania Al-Abdullah Street-Aljubeiha
Amman, 11942
Jordan
Phone: 962 0791845186
Fax: 962 65353388
Email: malik.sallam@ju.edu.jo

Abstract

Background: Adherence to evidence-based practice is indispensable in health care. Recently, the utility of generative artificial intelligence (AI) models in health care has been evaluated extensively. However, the lack of consensus guidelines on the design and reporting of findings of these studies poses a challenge for the interpretation and synthesis of evidence.

Objective: This study aimed to develop a preliminary checklist to standardize the reporting of generative AI-based studies in health care education and practice.

Methods: A literature review was conducted in Scopus, PubMed, and Google Scholar. Published records with “ChatGPT,” “Bing,” or “Bard” in the title were retrieved. Careful examination of the methodologies employed in the included records was conducted to identify the common pertinent themes and the possible gaps in reporting. A panel discussion was held to establish a unified and thorough checklist for the reporting of AI studies in health care. The finalized checklist was used to evaluate the included records by 2 independent raters. Cohen $\kappa$ was used as the method to evaluate the interrater reliability.

Results: The final data set that formed the basis for pertinent theme identification and analysis comprised a total of 34 records. The finalized checklist included 9 pertinent themes collectively referred to as METRICS (Model, Evaluation, Timing, Range/Randomization, Individual factors, Count, and Specificity of prompts and language). Their details are as follows: (1) Model used and its exact settings; (2) Evaluation approach for the generated content; (3) Timing of testing the model; (4) Transparency of the data source; (5) Range of tested topics; (6) Randomization of selecting the queries; (7) Individual factors in selecting the queries and interrater reliability; (8) Count of queries executed to test the model; and (9) Specificity of the prompts and language used. The overall mean METRICS score was 3.0 (SD 0.58). The tested METRICS score was acceptable, with the range of Cohen $\kappa$ of 0.558 to 0.962 ($P<.001$ for the 9 tested items). With classification per item, the highest average METRICS score was recorded for the “Model” item, followed by the “Specificity” item, while the lowest scores were recorded for the “Randomization” item (classified as suboptimal) and “Individual factors” item (classified as satisfactory).

Conclusions: The METRICS checklist can facilitate the design of studies guiding researchers toward best practices in reporting results. The findings highlight the need for standardized reporting algorithms for generative AI-based studies in health care, considering the variability observed in methodologies and reporting. The proposed METRICS checklist could be a preliminary
helpful base to establish a universally accepted approach to standardize the design and reporting of generative AI-based studies in health care, which is a swiftly evolving research topic.


**KEYWORDS**
guidelines; evaluation; meaningful analytics; large language models; decision support

## Introduction

The integration of generative artificial intelligence (AI) models into health care education and practice holds promising perspectives with numerous possibilities for continuous improvement [1-5]. Examples of generative AI-based conversational models characterized by ease of use and perceived usefulness include ChatGPT by OpenAI, Bing by Microsoft, and Bard by Google [6-8].

The vast potential of generative AI-based models in health care can be illustrated as follows. First, generative AI-based models can facilitate streamlining of the clinical workflow, with subsequent improvement in efficiency manifested in reduced time for care delivery and reduced costs [1,9-11]. Second, generative AI-based models can enhance personalized medicine with a huge potential to achieve refined prediction of disease risks and outcomes [1,12,13]. Third, generative AI-based models can be implemented to improve health literacy among lay individuals through the provision of easily accessible and understandable health information [1,14,15].

Despite the aforementioned advantages of generative AI-based models in health care, several valid concerns were raised, which should be considered carefully owing to their serious consequences [1,4,16]. For example, the lack of clarity on how generative AI-based models are trained raises ethical concerns [17,18]. Additionally, these models have an inherent bias in the generated content based on the modality of training used for their development and updates [17,18]. Importantly, the generation of inaccurate or misleading content, which might appear scientifically plausible to nonexperts (referred to as “hallucinations”), could have profound negative impacts in health care settings [1,19-21]. Furthermore, the integration of generative AI-based models in health care could raise complex medicolegal and accountability questions, compounded by the issues of data privacy and cybersecurity risks [1,4,22,23].

Similarly, the use of generative AI-based models can cause a paradigm shift in information acquisition, particularly in health care education [1,24-26]. However, careful consideration of the best policies and practices to incorporate AI-based models in health care education is needed [27]. This issue involves the urgent need to address the issues of inaccuracies, possible academic dishonesty, decline in critical thinking development, and deterioration of practical training skills [1].

Recently, a remarkable number of studies investigated the applicability and disadvantages of prominent generative AI-based conversational models, such as ChatGPT, Microsoft Bing, and Google Bard, in various health care and educational settings [1,2,4,28-34]. However, synthesizing evidence from such studies can be challenging owing to several reasons. Variations in methodologies implemented in various studies as well as in the reporting standards is a major limitation. This issue could hinder the efforts aiming to compare and contrast the results of generative AI-based studies in health care, contributing to the complexity in this domain. This variability arises from several factors, including different settings of the tested models, prompt variability, varying approaches used to evaluate the generated content of generative AI-based models, varying range of tested topics, and possible bias in selecting the tested subjects. Additionally, variability can be related to the different number and varying expertise of individual raters of content quality, as well as the variable number of queries executed, among other factors [35-37].

Therefore, it is important to initiate and develop an approach that can aid in establishing standardized reporting practices for studies aiming to evaluate the content of generative AI-based models, particularly in health care. This standardization can be crucial to facilitate the design of generative AI-based studies in health care, ensuring rigor and achieving precise comparison and credible synthesis of findings across different studies. Thus, we aimed to propose a preliminary framework (checklist) to establish proper guidelines for the design and reporting of findings of generative AI-based studies that address health care–related topics.

## Methods

### Study Design

The study was based on a literature review to highlight the key methodological aspects in studies that investigated 3 generative AI-based models (ChatGPT, Bing, and Bard) in health care education and practice. The literature review was conducted to identify relevant literature indexed in databases up to November 11, 2023 [38]. The databases used for this literature search were Scopus, PubMed/MEDLINE, and Google Scholar.

### Ethical Considerations

This study did not involve human subjects, and thus, the requirement of ethical permission was waived.

### Literature Search to Identify Relevant Records

The Scopus string query was as follows: (TITLE-ABS-KEY (“artificial intelligence” OR “AI”) AND TITLE-ABS-KEY (“healthcare” OR “health care”) AND TITLE-ABS-KEY (“education” OR “practice”)) AND PUBYEAR > 2022 AND DOCTYPE (ar OR re) AND (LIMIT-TO (PUBSTAGE , “final”)) AND (LIMIT-TO (SRCTYPE , “j”)) AND (LIMIT-TO (LANGUAGE , “English”)). The Scopus search yielded a total of 843 documents.
The PubMed advanced search tool was used as follows: ("artificial intelligence"[Title/Abstract] OR “AI”[Title/Abstract]) AND ("healthcare"[Title/Abstract] OR “health care”[Title/Abstract]) AND (“education”[Title/Abstract] OR “practice”[Title/Abstract]) AND (“2022/12/01”[Date - Publication] : “2023/11/11”[Date - Publication]). The PubMed search yielded a total of 564 records.

In Google Scholar and using the Publish or Perish software (version 8), in the title words and in the years 2022-2023, the search was as follows: “artificial intelligence” OR “AI” AND “healthcare” OR “health care” AND “education” OR “practice,” with a maximum of 999 records retrieved [39].

Criteria for Record Inclusion

The records from the 3 databases were merged using EndNote 20.2.1 software. This was followed by removal of duplicate records and removal of preprints by using the following function: ANY FIELD preprint OR ANY FIELD xiv OR ANY FIELD SSRN OR ANY FIELD Researchgate OR ANY FIELD researchsquare OR ANY FIELD. The retrieved records were eligible for the final screening step given the following inclusion criteria: (1) Original article; (2) English record; (3) Published (peer reviewed); and (4) Assessment in health care practice or health care education. Finally, the imported references were subjected to the search function in EndNote as follows: Title contains ChatGPT OR Title contains Bing OR Title contains Bard. The selection of the included records was performed by the first author (Malik Sallam).

Development of the Initial Checklist Items

Initial development of the proposed checklist began with the assessment of the methodology and results sections of the included records, a majority of which were regarded as cross-sectional descriptive studies. Then, we referred to 2 commonly used reporting and quality guidelines to proactively explore pertinent themes for the proposed checklist based on the nature of the included records: (1) STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) Statement: guidelines for reporting observational studies (checklist: cross-sectional studies) and (2) CASP (Critical Appraisal Skills Programme) Checklist (CASP qualitative studies checklist) [40,41]. This facilitated the allocation of ethical considerations, including transparency, methodological rigor, and issues related to bias, in the proposed checklist. Then, the 3 authors conducted an independent content review process to identify all the possible essential themes and best practices in generative AI-based health care studies among the included records. Finally, the authors had a collaborative discussion to refine the selected themes and classify these themes into specific components relevant to the study objectives. Special attention was given to aspects of method and result reporting that were perceived to impact the quality and reproducibility of the records, as identified by the 3 authors.

Establishing the Final Checklist Criteria

Careful examination of the included records resulted in the compilation of 3 independent lists of “pertinent themes,” which are herein defined as being critical or recurring in the reporting of results of generative AI-based studies. A thorough discussion among the authors followed to reach a consensus on the pertinent themes. Recurring themes were defined as those found in the methods of at least three separate records. Critical aspects were defined as those that would impact the conclusions of the included records as agreed by the 3 authors.

The final pertinent themes were selected based on their author-perceived significance in the quality and reproducibility of the findings. A final list of 9 themes was agreed upon by the authors as follows: (1) the “Model” of the generative AI-based tool or tools used in the included record and the explicit mention of the exact settings employed for each tool; (2) the “Evaluation” approach to assess the quality of the content generated by the generative AI-based model in terms of objectivity to reach unbiased findings and subjectivity; (3) the exact “Timing” of generative AI-based model testing and its duration; (4) the “Transparency” of data sources used to generate queries for the generative AI-based model testing, including the permission to use copyrighted content; (5) the “Range” of topics tested (single topic, multiple related topics, or various unrelated topics, as well as the breadth of intertopic and intratopic queries tested); (6) the degree of “Randomization” of topics selected to be tested to consider the potential bias; (7) the “Individual” subjective role in evaluating the content and the possibility of interrater reliability concordance or discordance; (8) the number (“Count”) of queries executed on each generative AI model entailing the sample size of queries tested; and (9) the “Specificity” of prompts used on each generative AI-based model, including the exact phrasing of each prompt and the presence of feedback and learning loops, and the “Specificity” of the language or languages used in testing, besides any other cultural issues.

Thus, the final checklist was termed METRICS (Model, Evaluation, Timing, Range/Randomization, Individual factors, Count, and Specificity of prompts and language).

Scoring the METRICS Items and Classification of the METRICS Score

Testing of the included records was performed by 2 independent raters (Malik Sallam and Mohammed Sallam) independently, with each METRICS item scored using a 5-point Likert scale as follows: 5=excellent, 4=very good, 3=good, 2=satisfactory, and 1=suboptimal. For the items that were deemed “not applicable” (eg, individual factors for studies that employed objective methods for evaluation), no score was given. The scores for the 2 raters were then averaged. The average METRICS score was calculated as the sum of average scores for each applicable item divided by 10 minus the number of items deemed not applicable.

The subjective assessment of the 2 raters was performed based on predefined criteria as a general guide. For example, if the exact dates of model queries were mentioned, the “Timing” item was scored as excellent. The count of queries was agreed to be categorized as excellent if it was more than 500, while a single case or no mention of the count was considered suboptimal. For the prompt attributes, scores were assigned based on the availability of exact prompts, explicit mention of the language used, and details of prompting. Thus, prompts and language specificity were appraised positively if the study clearly and explicitly made the exact prompts available and if
there was an explicit mention of the language employed in the prompts. The evaluation method was agreed to be rated higher for objective assessments with full details and lower for subjective assessments. The explicit mention of the method of interrater reliability testing was agreed to be scored higher for the “Individual” item. Transparency was assessed based on the comprehensiveness of the data source, and the presence of full database disclosure and permission to use the data was agreed to be given an excellent score. Randomization was agreed to be scored the lowest for the absence of details and the highest for explicit detailed descriptions.

Finally, we decided to highlight the records that scored the highest for each METRICS item. The decision to take this approach was based on an attempt to refrain from providing examples for the other quality categories to avoid premature conclusions regarding the quality of the included studies owing to the preliminary pilot nature of the METRICS tool.

Statistical and Data Analysis
The average METRICS scores were classified into distinct categories of equal weights as follows: excellent (score 4.21-5.00), very good (3.41-4.20), good (2.61-3.40), satisfactory (1.81-2.60), and suboptimal (1.00-1.80).

The Cohen $\kappa$ measure was used to assess the interrater reliability by 2 independent raters. The Cohen $\kappa$ measure was categorized as follows: <0.20, poor agreement; 0.21-0.40, fair agreement; 0.41-0.60, good agreement; 0.61-0.80, very good agreement; and 0.81-1.00, excellent agreement.

For statistical analysis, we used IBM SPSS Statistics for Windows, Version 26.0 (IBM Corp). A $P$ value <.05 was considered significant.

Results
Description of the Included Studies
A total of 34 studies were included in the final analysis that aimed to establish the METRICS criteria (Figure 1). The most common source of records was Cureus, with 9 of the 34 records (27%), followed by BMJ Neurology Open, with 2 of the 34 records (6%). The remaining 23 records were published in 23 different journals.

Evaluation of the Included Records Based on MED-METRICS Items
The METRICS checklist was divided into the following 3 parts: “Model” attributes, “Evaluation” approach, and features of “Data” (MED-METRICS).

The complete details of the model attributes of the included studies are presented in Table 1.

ChatGPT was tested in all the records included (34/34, 100%), followed by Google Bard (5/34, 15%) and Bing Chat (5/34, 15%). The exact dates of generative AI-based model queries were explicitly mentioned in 13 of the 34 records (38%). The count of cases or questions that were tested in the studies ranged from a single case to 2576 questions. The majority of studies (23/34, 68%) tested the AI models based on queries in the English language.

The complete details of the evaluation approach of the content generated by the AI-based models in the included studies are presented in Table 2.

Objective evaluation of the content generated by the generative AI-based model was noted in 15 of the 34 records (44%).

The complete details of the features of data used to generate queries for testing on the generative AI-based models, including the range of topics and randomization, are presented in Table 3.

Explicit mention of the randomization process was only noted in 4 of the 34 included studies (9%). Of the 34 records, 6 (18%) involved broad multidisciplinary medical exam questions (18%). Moreover, 2 studies (6%) explicitly mentioned the permission to use the data for the studies.
Table 1. Details of the model attributes of the included studies.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Model</th>
<th>Timing</th>
<th>Count</th>
<th>Specificity of the prompts and language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al-Ashwal et al</td>
<td>ChatGPT-3.5, ChatGPT-4, Bing, and Bard with unclear settings</td>
<td>One month, May 2023</td>
<td>255 drug-drug pairs</td>
<td>The exact prompts were provided in English.</td>
</tr>
<tr>
<td>Alfertshofer et al</td>
<td>ChatGPT with unclear settings</td>
<td>Not provided</td>
<td>1800 questions</td>
<td>The exact prompt was used for each question. The questions were taken from United States, United Kingdom, Italian, French, Spanish, and Indian exams. A new session was used for each question.</td>
</tr>
<tr>
<td>Ali et al</td>
<td>ChatGPT Feb 9 free version with unclear settings</td>
<td>Not provided</td>
<td>50 items</td>
<td>Information was provided fully in the supplementary file of the article in English.</td>
</tr>
<tr>
<td>Aljindan et al</td>
<td>ChatGPT-4 with unclear settings</td>
<td>Not provided</td>
<td>220 questions</td>
<td>Initial prompting that involved role playing as a medical professional. The language was English.</td>
</tr>
<tr>
<td>Altmimi et al</td>
<td>ChatGPT-3.5</td>
<td>Single day not otherwise specified</td>
<td>9 questions</td>
<td>The exact prompts were provided in English.</td>
</tr>
<tr>
<td>Baglivo et al</td>
<td>Bing, ChatGPT, Chatsonic, Bard, and YouChat with full details of the mode and large language model, including plugins</td>
<td>Exact dates were provided for each model (April 12, 13, and 14, 2023, and July 13, 2023)</td>
<td>15 questions</td>
<td>Italian was used.</td>
</tr>
<tr>
<td>Biswas et al</td>
<td>ChatGPT-3.5 with unclear settings</td>
<td>Exact date provided (March 16, 2023)</td>
<td>11 questions</td>
<td>The exact prompts were provided in English, with a new session for each question.</td>
</tr>
<tr>
<td>Chen et al</td>
<td>ChatGPT-4 with unclear settings</td>
<td>Not provided</td>
<td>560 questions</td>
<td>The exact prompts were provided in English.</td>
</tr>
<tr>
<td>Deiana et al</td>
<td>ChatGPT-3.5 and ChatGPT-4 with unclear settings</td>
<td>Not provided</td>
<td>11 questions</td>
<td>The exact prompts were not explicitly provided. English was used. A new session was used for each question. Up to three iterations were allowed for incorrect responses.</td>
</tr>
<tr>
<td>Fuchs et al</td>
<td>ChatGPT-3 and ChatGPT-4 with unclear settings</td>
<td>Exact dates provided (February 19, 2023, and March 25, 2023)</td>
<td>60 questions</td>
<td>Dental medicine questions were translated from German to English, while the other questions were already present in English. Exact prompts were provided in English with prompting in 2 groups for the same questions: one group was primed with instructions, while the other was not primed. A total of 20 trials were conducted per group, and chat history was cleared after each trial.</td>
</tr>
<tr>
<td>Ghosh &amp; Bir</td>
<td>ChatGPT (version March 14, 2023) with unclear settings</td>
<td>March 14 and 16, 2023</td>
<td>200 questions</td>
<td>The exact prompts and language were not explicitly provided. The first response was taken as final, and the option of “regenerate response” was not used.</td>
</tr>
<tr>
<td>Giannos</td>
<td>ChatGPT-3 and ChatGPT-4 with unclear settings</td>
<td>Not provided</td>
<td>69 questions</td>
<td>Not explicitly provided.</td>
</tr>
<tr>
<td>Gobira et al</td>
<td>ChatGPT-4 with unclear settings</td>
<td>Not provided</td>
<td>125 questions</td>
<td>Portuguese was used.</td>
</tr>
<tr>
<td>Grewal et al</td>
<td>ChatGPT-4 with unclear settings</td>
<td>The first week of May 2023</td>
<td>Not clear</td>
<td>The exact prompts were provided in English. One follow-up prompt was used for enhancement of some prompts.</td>
</tr>
<tr>
<td>Guerra et al</td>
<td>ChatGPT-4 with unclear settings</td>
<td>Not provided</td>
<td>591 questions</td>
<td>The exact prompts were provided, while the language was not explicitly provided.</td>
</tr>
<tr>
<td>Hamed et al</td>
<td>ChatGPT-4 with unclear settings</td>
<td>Not provided</td>
<td>Not clear</td>
<td>The exact prompts and language were not explicitly provided. Different prompts were tried to identify the most suitable.</td>
</tr>
<tr>
<td>Hoch et al</td>
<td>ChatGPT (May 3rd version) with unclear settings</td>
<td>May 5 and 7, 2023</td>
<td>2576 questions</td>
<td>The exact prompts were provided, while the language was not explicitly provided.</td>
</tr>
<tr>
<td>Authors</td>
<td>Model</td>
<td>Timing</td>
<td>Count</td>
<td>Specificity of the prompts and language</td>
</tr>
<tr>
<td>-------------------------</td>
<td>--------------------------------------------</td>
<td>---------------------------------------------</td>
<td>------------------------</td>
<td>---------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Juhi et al [59]</td>
<td>ChatGPT with unclear settings</td>
<td>February 20, 2023, to March 5, 2023</td>
<td>40 drug-drug pairs</td>
<td>The exact prompts were provided in English.</td>
</tr>
<tr>
<td>Kuang et al [60]</td>
<td>ChatGPT with unclear settings</td>
<td>Not provided</td>
<td>Not clear</td>
<td>The exact prompts were not explicitly provided. English was used.</td>
</tr>
<tr>
<td>Kumari et al [61]</td>
<td>ChatGPT-3.5, Bard, and Bing with unclear settings</td>
<td>July 30, 2023</td>
<td>50 questions</td>
<td>The exact prompts were not explicitly provided. English was used.</td>
</tr>
<tr>
<td>Kung et al [62]</td>
<td>ChatGPT-3.5 and ChatGPT-4 with unclear settings</td>
<td>July 2023</td>
<td>215 questions</td>
<td>Not clear</td>
</tr>
<tr>
<td>Lai et al [63]</td>
<td>ChatGPT-4 (May 24; Version 3.5) with unclear settings</td>
<td>Not provided</td>
<td>200 questions</td>
<td>The exact prompts and language were not explicitly provided. Three attempts to answer the complete set of questions over 3 weeks (once per week), with a new session for each question.</td>
</tr>
<tr>
<td>Lyu et al [64]</td>
<td>ChatGPT with unclear settings</td>
<td>Mid-February 2023</td>
<td>Not clear</td>
<td>The exact prompts were provided in English.</td>
</tr>
<tr>
<td>Moise et al [65]</td>
<td>ChatGPT-3.5 with unclear settings</td>
<td>Not provided</td>
<td>23 questions</td>
<td>The exact prompts were provided in English, with a new session for each question.</td>
</tr>
<tr>
<td>Oca et al [66]</td>
<td>ChatGPT, Bing, and Bard with unclear settings</td>
<td>April 11, 2023</td>
<td>20 queries for each model</td>
<td>The exact prompts were provided in English.</td>
</tr>
<tr>
<td>Oztermeli &amp; Oztermeli [67]</td>
<td>ChatGPT-3.5 with unclear settings</td>
<td>Not provided</td>
<td>1177 questions</td>
<td>The exact prompts were not explicitly provided. Turkish was used, with a new session for each question.</td>
</tr>
<tr>
<td>Pugliese et al [68]</td>
<td>ChatGPT with unclear settings</td>
<td>March 25, 2023</td>
<td>15 questions</td>
<td>The exact prompts were provided in English, with a new session for each question.</td>
</tr>
<tr>
<td>Sallam et al [69]</td>
<td>ChatGPT (default model) with unclear settings</td>
<td>February 25, 2023</td>
<td>Not provided</td>
<td>The exact prompts were provided in English.</td>
</tr>
<tr>
<td>Seth et al [70]</td>
<td>ChatGPT-3.5, Bard, and Bing AI</td>
<td>Not provided</td>
<td>6 questions</td>
<td>The exact prompts were provided in English.</td>
</tr>
<tr>
<td>Suthar et al [71]</td>
<td>ChatGPT-4 with unclear settings</td>
<td>Not provided</td>
<td>140 cases</td>
<td>The exact prompts were not explicitly provided. English was used.</td>
</tr>
<tr>
<td>Walker et al [72]</td>
<td>ChatGPT-4 with unclear settings</td>
<td>Not provided</td>
<td>5 cases</td>
<td>The exact prompts were not explicitly provided. English was used, with a new session for each question.</td>
</tr>
<tr>
<td>Wang et al [73]</td>
<td>ChatGPT-3.5 and ChatGPT-4 with unclear settings</td>
<td>February 14, 2023, for ChatGPT-3.5 and May 14-16, 2023, for ChatGPT-4</td>
<td>300 questions</td>
<td>The exact prompts were provided. Chinese and English were used. The prompts were enhanced though role play.</td>
</tr>
<tr>
<td>Wang et al [74]</td>
<td>ChatGPT-3.5 with unclear settings</td>
<td>March 5-10, 2023</td>
<td>Not clear</td>
<td>Chinese (Mandarin) and English were used. Examples of prompts were provided.</td>
</tr>
<tr>
<td>Zhou et al [75]</td>
<td>ChatGPT-3.5 with unclear settings</td>
<td>April 24-25, 2023</td>
<td>Single case and multiple poll questions</td>
<td>The exact prompts were provided in English.</td>
</tr>
<tr>
<td>Authors</td>
<td>Evaluation of performance</td>
<td>Individual role and interrater reliability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>----------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Al-Ashwal et al [42]</td>
<td>Objective via 2 different clinical reference tools</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alftertschofer et al [43]</td>
<td>Objective based on the key answers, with the questions screened independently by 4 investigators</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ali et al [44]</td>
<td>Objective for multiple-choice questions and true or false questions, and subjective for short-answer and essay questions</td>
<td>Assessment by 2 assessors independently with intraclass correlation coefficient for agreement</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aljindan et al [45]</td>
<td>Objective based on key answers and historical performance metrics</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Altunimie et al [46]</td>
<td>Subjective</td>
<td>Not clear; Assessment for accuracy, informativeness, and accessibility by clinical toxicologists and emergency medicine physicians</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baglivo et al [47]</td>
<td>Objective based on key answers and comparison with 5th year medical students’ performance</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biswas et al [48]</td>
<td>Subjective by a 5-member team of optometry teaching and expert staff with over 100 years of clinical and academic experience between them; Independent evaluation on a 5-point Likert scale ranging from very poor to very good</td>
<td>The median scores across raters for each response were studied; The score represented rater consensus, while the score variance represented disagreements between the raters</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chen et al [49]</td>
<td>Objective based on key answers</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deiana et al [50]</td>
<td>Subjective based on qualitative assessment of correctness, clarity, and exhaustiveness; Each response rated using a 4-point Likert scale scored from strongly disagree to strongly agree</td>
<td>Independent assessment by 2 raters with experience in vaccination and health communication topics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fuchs et al [51]</td>
<td>Objective based on key answers</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ghosh &amp; Bir [52]</td>
<td>Objective based on key answers; Subjectivity by raters’ assessment</td>
<td>Scoring by 2 assessors on a scale of 0 to 5, with 0 being incorrect and 5 being fully correct, based on a preselected answer key</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Giannos [53]</td>
<td>Objective based on key answers</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gobira et al [54]</td>
<td>Objective based on key answers, with an element of subjectivity through classifying the responses as adequate, inadequate, or indeterminate</td>
<td>Two raters independently scored the accuracy; After individual evaluations, the raters performed a third assessment to reach a consensus on the questions with differing results</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grewal et al [55]</td>
<td>Not clear</td>
<td>Not clear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guerra et al [56]</td>
<td>Subjective through comparison with the results of a previous study on the average performance of users, and a cohort of medical students and neurosurgery residents</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hamed et al [57]</td>
<td>Subjective</td>
<td>Not clear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoch et al [58]</td>
<td>Objective based on key answers</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Juhi et al [59]</td>
<td>Subjective and the use of Stockley’s Drug Interactions Pocket Companion 2015 as a reference key</td>
<td>Two raters reached a consensus for categorizing the output</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kuang et al [60]</td>
<td>Subjective</td>
<td>Not clear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kumari et al [61]</td>
<td>Subjective; Content validity checked by 2 experts of curriculum design</td>
<td>Three independent raters scored content based on their correctness, with an accuracy score ranging from 1 to 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kung et al [62]</td>
<td>Objective based on key answers</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lai et al [63]</td>
<td>Objective based on key answers</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lyu et al [64]</td>
<td>Subjective</td>
<td>Two experienced radiologists (with 21 and 8 years of experience) evaluated the quality of the ChatGPT responses</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moise et al [65]</td>
<td>Subjective through comparison with the latest American Academy of Otolaryngology—Head and Neck Surgery Foundation Clinical Practice Guideline: Tymanostomy Tubes in Children (Update)</td>
<td>Two independent raters evaluated the output; The interrater reliability was assessed using the Cohen $\kappa$ test; To confirm consensus, responses were reviewed by the senior author</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oca et al [66]</td>
<td>Not clear</td>
<td>Not clear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Authors</td>
<td>Evaluation of performance</td>
<td>Individual role and interrater reliability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>-------------------------------------------------------------------------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oztermeli &amp; Oztermeli [67]</td>
<td>Objective based on key answers</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pugliese et al [68]</td>
<td>Subjective using the Likert scale for accuracy, completeness, and comprehensiveness</td>
<td>Multirater: 10 key opinion leaders in nonalcoholic fatty liver disease and 1 nonphysician with expertise in patient advocacy in liver disease independently rating the AI² content</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sallam et al [69]</td>
<td>Subjective based on correctness, clarity, and conciseness</td>
<td>Fleiss multirater κ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seth et al [70]</td>
<td>Subjective through comparison with the current health care guidelines for rhinoplasty, and evaluation by a panel of plastic surgeons through a Likert scale to assess the readability and complexity of the text and the education level required for understanding, and the modified DISCERN² score</td>
<td>Not clear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Suthar et al [71]</td>
<td>Subjective by 3 fellowship-trained neuroradiologists, using a 5-point Likert scale, with 1 indicating “highly improbable” and 5 indicating “highly probable”</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walker et al [72]</td>
<td>Modified EQIP² Tool with comparison with UK National Institute for Health and Care Excellence guidelines for gall-stone disease, pancreatitis, liver cirrhosis, or portal hypertensive, and the European Association for Study of the Liver guidelines</td>
<td>All answers were assessed by 2 authors independently, and in case of a contradictory result, resolution was achieved by consensus; The process was repeated 3 times per EQIP item; Wrong or out of context answers, known as “AI hallucinations,” were recorded</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wang et al [73]</td>
<td>Subjective</td>
<td>Unclear</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wang et al [74]</td>
<td>Objective based on key answers</td>
<td>Not applicable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zhou et al [75]</td>
<td>Subjective</td>
<td>Unclear</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

²AI: artificial intelligence.
³DISCERN is an instrument for judging the quality of written consumer health information on treatment choices.
⁴EQIP: Ensuring Quality Information for Patients.
Table 3. Classification of the included studies based on the features of data used to generate queries for testing on the generative artificial intelligence–based models.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Transparency</th>
<th>Range</th>
<th>Randomization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al-Ashwal et al [42]</td>
<td>Full description using 2 tools for assessment: Micromedex, a subscription-based drug-drug interaction screening tool, and Drugs.com, a free database</td>
<td>Narrow; Drug-drug interaction prediction</td>
<td>Nonrandom; Purposeful selection of the drugs by a clinical pharmacist; 5 drugs paired with the top 51 prescribed drugs</td>
</tr>
<tr>
<td>Alftershofer et al [43]</td>
<td>Full description using the question bank AMBOSS, with official permission for the use of the AMBOSS USMLE step 2CK practice question bank for research purposes granted by AMBOSS</td>
<td>Broad</td>
<td>Randomly extracted</td>
</tr>
<tr>
<td>Ali et al [44]</td>
<td>Developed by the researchers and reviewed by a panel of experienced academics for accuracy, clarity of language, relevance, and agreement on correct answers; Evaluation of face validity, accuracy, and suitability for undergraduate dental students</td>
<td>Narrow intersubject (dentistry); Broad intra-subject in restorative dentistry, periodontics, fixed prosthodontics, removable prosthodontics, endodontics, pedodontics, orthodontics, preventive dentistry, oral surgery, and oral medicine</td>
<td>Not clear</td>
</tr>
<tr>
<td>Aljindan et al [45]</td>
<td>Saudi Medical Licensing Exam questions extracted from the subscription CanadaQBank website</td>
<td>Broad in medicine, with 30% of the questions from medicine, 25% from obstetrics and gynecology, 25% from pediatrics, and the remaining 20% from surgery</td>
<td>Randomized through 4 researchers to ensure comprehensive coverage of questions and eliminate potential bias in question selection</td>
</tr>
<tr>
<td>Altamimi et al [46]</td>
<td>Snakebite management information guidelines derived from the World Health Organization, Centers for Disease Control and Prevention, and the clinical literature</td>
<td>Narrow</td>
<td>Not clear</td>
</tr>
<tr>
<td>Baglivo et al [47]</td>
<td>The Italian National Medical Residency test</td>
<td>Narrow; Vaccination-related questions from the Italian National Medical Residency Test</td>
<td>Not clear</td>
</tr>
<tr>
<td>Biswas et al [48]</td>
<td>Constructed based on the frequently asked questions on the myopia webpage of the Association of British Dispensing Opticians and the College of Optometrists</td>
<td>Narrow involving 9 categories: 1 each for disease summary, cause, symptom, onset, prevention, complication, natural history of untreated myopia, and prognosis, and 3 involving treatments</td>
<td>Not clear</td>
</tr>
<tr>
<td>Chen et al [49]</td>
<td>BoardVitals, which is an online question bank accredited by the Accreditation Council for Continuing Medical Education</td>
<td>Neurology-based; Broad intrasubject: basic neuroscience; behavioral, cognitive, psychiatry; cerebrovascular; child neurology; congenital; cranial nerves; critical care; demyelinating disorders; epilepsy and seizures; ethics; genetic; headache; imaging or diagnostic studies; movement disorders; neuro-ophthalmology; neuro-otology; neuroinfectious disease; neurologic complications of systemic disease; neuromuscular; neurotoxicology; nutrition, metabolic; oncology; pain; pharmacology; pregnancy; sleep; and trauma</td>
<td>Not clear</td>
</tr>
<tr>
<td>Deiana et al [50]</td>
<td>Questions concerning vaccine myths and misconceptions by the World Health Organization</td>
<td>Narrow on vaccine myths and misconceptions</td>
<td>Not clear</td>
</tr>
<tr>
<td>Fuchs et al [51]</td>
<td>Digital platform self-assessment questions tailored for dental and medical students at the University of Bern’s Institute for Medical Education</td>
<td>Broad with multiple-choice questions designed for dental students preparing for the Swiss Federal Licensing Examination in Dental Medicine, and allergists and immunologists preparing for the European Examination in Allergy and Clinical Immunology</td>
<td>Not clear</td>
</tr>
<tr>
<td>Ghosh &amp; Bir [52]</td>
<td>Department question bank, which is a compilation of first and second semester questions from various medical universities across India</td>
<td>Biochemistry</td>
<td>Random without details of randomization</td>
</tr>
<tr>
<td>Giannos [53]</td>
<td>Specialty Certificate Examination Neurology Web Questions bank</td>
<td>Neurology and neuroscience</td>
<td>Not clear</td>
</tr>
<tr>
<td>Authors</td>
<td>Transparency</td>
<td>Range</td>
<td>Randomization</td>
</tr>
<tr>
<td>-------------------------</td>
<td>--------------</td>
<td>-----------------------------------------------------------------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Gobira et al [54]</td>
<td>National Brazilian Examination for Revalidation of Medical Diplomas issued by Foreign Higher Education Institutions (Revalida)</td>
<td>Preventive medicine, gynecology and obstetrics, surgery, internal medicine, and pediatrics</td>
<td>Not clear</td>
</tr>
<tr>
<td>Grewal et al [55]</td>
<td>Not clear</td>
<td>Radiology</td>
<td>Not clear</td>
</tr>
<tr>
<td>Guevara et al [56]</td>
<td>Questions released by the Congress of Neurological Surgeons in the self-assessment neurosurgery exam</td>
<td>Neurosurgery across 7 subspecialties: tumor, cerebrovascular, trauma, spine, functional, pediatrics, and pain or nerve</td>
<td>Not clear</td>
</tr>
<tr>
<td>Hamed et al [57]</td>
<td>Guidelines adapted from Diabetes Canada Clinical Practice Guidelines Expert Committee, the Royal Australian College of General Practitioners, Australian Diabetes Society position statement, and the Joint British Diabetess Societies</td>
<td>Management of diabetic ketoacidosis</td>
<td>Not clear</td>
</tr>
<tr>
<td>Hoch et al [58]</td>
<td>Question database of an online learning platform funded by the German Society of Otorhino-Laryngology, Head and Neck Surgery</td>
<td>Otolaryngology with a range of 15 distinct otolaryngology subspecialties, including allergology, audiology, ear, nose and throat, tumors, face and neck, inner ear and skull base, larynx, middle ear, oral cavity and pharynx, nose and sinuses, phoniatrics, salivary glands, sleep medicine, vestibular system, and legal aspects</td>
<td>Not clear</td>
</tr>
<tr>
<td>Juhi et al [59]</td>
<td>A list of drug-drug interactions from previously published research</td>
<td>Narrow on drug-drug interaction</td>
<td>Not clear</td>
</tr>
<tr>
<td>Kumari et al [61]</td>
<td>Designed by experts in hematology-related cases</td>
<td>Hematology with the following intrasubject aspects: case solving, laboratory calculations, disease interpretations, and other relevant aspects of hematology</td>
<td>Not clear</td>
</tr>
<tr>
<td>Kung et al [62]</td>
<td>Orthopedic In-Training Examination</td>
<td>Orthopedics</td>
<td>Not clear</td>
</tr>
<tr>
<td>Lai et al [63]</td>
<td>The United Kingdom Medical Licensing Assessment, which is a newly derived national undergraduate medical exit examination</td>
<td>Broad in medicine with the following aspects: acute and emergency, cancer, cardiovascular, child health, clinical hematology, ear, nose and throat, endocrine and metabolic, gastrointestinal including liver, general practice and primary health care, genetics and genomics, infection, medical ethics and law, medicine of older adults, mental health, musculoskeletal, neuroscience obstetrics and gynecology, ophthalmology, palliative and end of life care, perioperative medicine and anesthesia, renal and urology, respiratory, sexual health, social and population health, and surgery</td>
<td>Not clear</td>
</tr>
<tr>
<td>Lyu et al [64]</td>
<td>Chest computed tomography and brain magnetic resonance imaging screening reports collected from the Atrium Health Wake Forest Baptist clinical database</td>
<td>Radiology</td>
<td>Not clear</td>
</tr>
<tr>
<td>Oca et al [66]</td>
<td>Not clear</td>
<td>Narrow involving only queries on accurate recommendation of close ophthalmologists</td>
<td>Not clear</td>
</tr>
<tr>
<td>Authors</td>
<td>Transparency</td>
<td>Range</td>
<td>Randomization</td>
</tr>
<tr>
<td>---------</td>
<td>--------------</td>
<td>-------</td>
<td>---------------</td>
</tr>
<tr>
<td>Oztermeli &amp; Oztermeli [67]</td>
<td>Turkish medical specialty exam, prepared by the Student Selection and Placement Center</td>
<td>Broad: basic sciences including anatomy, physiology-histology-embryology, biochemistry, microbiology, pathology, and pharmacology; clinical including internal medicine, pediatrics, general surgery, obstetrics and gynecology, neurology, neurosurgery, psychiatry, public health, dermatology, radiology, nuclear medicine, otorhinolaryngology, ophthalmology, orthopedics, physical medicine and rehabilitation, urology, pediatric surgery, cardiovascular surgery, thoracic surgery, plastic surgery, anesthesiology and reanimation, and emergency medicine</td>
<td>Not clear</td>
</tr>
<tr>
<td>Pugliese et al [68]</td>
<td>Expert selection of 15 questions commonly asked by patients with nonalcoholic fatty liver disease</td>
<td>Narrow involving nonalcoholic fatty liver disease aspects</td>
<td>Not clear</td>
</tr>
<tr>
<td>Sallam et al [69]</td>
<td>Panel discussion of experts in health care education</td>
<td>Broad on health care education, medical, dental, pharmacy, and public health</td>
<td>Not clear</td>
</tr>
<tr>
<td>Seth et al [70]</td>
<td>Devised by 3 fellows of the Royal Australasian College of Surgeons, with experience in performing rhinoplasty and expertise in facial reconstructive surgery</td>
<td>Narrow involving technical aspects of rhinoplasty</td>
<td>Not clear</td>
</tr>
<tr>
<td>Suthar et al [71]</td>
<td>Quizzes from the Case of the Month feature of the American Journal of Neuroradiology</td>
<td>Narrow involving radiology</td>
<td>Not clear</td>
</tr>
<tr>
<td>Walker et al [72]</td>
<td>Devised based on the Global Burden of Disease tool</td>
<td>Narrow involving benign and malignant hepatopancreatobiliary-related conditions</td>
<td>Not clear</td>
</tr>
<tr>
<td>Wang et al [73]</td>
<td>Medical Exam Help. A total of 10 inpatient and 10 outpatient medical records to form a collection of Chinese medical records after desensitization</td>
<td>Clinical medicine, basic medicine, medical humanities, and relevant laws</td>
<td>Not clear</td>
</tr>
<tr>
<td>Wang et al [74]</td>
<td>The Taiwanese Senior Professional and Technical Examinations for Pharmacists downloaded from the Ministry of Examination website</td>
<td>Broad involving pharmacology and pharmaceutical chemistry, pharmaceutical analysis and pharmacognosy (including Chinese medicine), pharmaceutics and biopharmaceutics, dispensing pharmacy and clinical pharmacy, therapeutics, pharmacy administration, and pharmacy law</td>
<td>Not clear</td>
</tr>
<tr>
<td>Zhou et al [75]</td>
<td>A single clinical case from OrthoBullets, a global clinical collaboration platform for orthopedic surgeons; Written permission to use their clinical case report</td>
<td>Very narrow involving a single orthopedic case</td>
<td>Not clear</td>
</tr>
</tbody>
</table>

**Examples of Optimal Reporting of Each Criterion Within the METRICS Checklist**

The records with the highest scores for each METRICS item, as determined by the average subjective interrater assessments, are shown in Table 4.
Table 4. Included records that had the highest METRICS score per item.

<table>
<thead>
<tr>
<th>Item</th>
<th>Issues considered in each item</th>
<th>Excellent or very good reporting examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1 Model</td>
<td>What is the model of the generative AI tool used for generating content, and what are the exact settings for each tool?</td>
<td>Baglivo et al [47]: Bing, ChatGPT, Chatsonic, Bard, and YouChat, with full details of the mode and large language model, including plugins</td>
</tr>
<tr>
<td>#2 Evaluation</td>
<td>What is the exact approach used to evaluate the content generated by the generative AI-based model and is it an objective or subjective evaluation?</td>
<td>Alfertshofer et al [43]: Objective based on key answers, with the questions screened independently by 4 investigators; Ali et al [44]: Objective for multiple-choice questions and true or false questions, and subjective for short-answer and essay questions; Aljindan et al [45]: Objective based on key answers and historical performance metrics; and Baglivo et al [47]: Objective based on key answers and comparison with 5th year medical students’ performance</td>
</tr>
<tr>
<td>#3a Timing</td>
<td>When is the generative AI model tested exactly and what are the duration and timing of testing?</td>
<td>Baglivo et al [47]; Biswas et al [48]; Fuchs et al [51]; Ghosh &amp; Bir [52]; Hoch et al [58]; Juhi et al [59]; Kumari et al [61]; Kung et al [62]; Oca et al [66]; Pugliese et al [68]; Sallam et al [69]; Wang et al [73]; and Zhou et al [75]</td>
</tr>
<tr>
<td>#3b Transparency</td>
<td>How transparent are the data sources used to generate queries for the generative AI-based model?</td>
<td>Alfertshofer et al [43]</td>
</tr>
<tr>
<td>#4a Range</td>
<td>What is the range of topics tested and are they intersubject or intrasubject with variability in different subjects?</td>
<td>Ali et al [44]; Chen et al [49]; Hoch et al [58]; and Wang et al [73]</td>
</tr>
<tr>
<td>#4b Randomization</td>
<td>Was the process of selecting the topics to be tested on the generative AI-based model randomized?</td>
<td>Alfertshofer et al [43] and Aljindan et al [45]</td>
</tr>
<tr>
<td>#5 Individual</td>
<td>Is there any individual subjective involvement in generative AI content evaluation? If so, did the authors describe the details in full?</td>
<td>Ali et al [44] and Moise et al [65]</td>
</tr>
<tr>
<td>#6 Count</td>
<td>What is the count of queries executed (sample size)?</td>
<td>Alfertshofer et al [43]; Chen et al [49]; Guerra et al [56]; Hoch et al [58]; and Oztermeli &amp; Oztermeli [67]</td>
</tr>
<tr>
<td>#7 Specificity of the prompt or language</td>
<td>How specific are the exact prompts used? Were those exact prompts provided fully? Did the authors consider the feedback and learning loops? How specific are the language and cultural issues considered in the generative AI model?</td>
<td>Alfertshofer et al [43]; Biswas et al [48]; Fuchs et al [51]; Grewal et al [55]; Wang et al [73]; Moise et al [65]; and Pugliese et al [68]</td>
</tr>
</tbody>
</table>

*AI: artificial intelligence.

**Interrater Assessment of the Included Records Based on METRICS Scores**

The overall mean METRICS score was 3.0 (SD 0.58). For each item, the $\kappa$ interrater reliability ranged from 0.558 to 0.962 ($P<.001$ for the 9 tested items), indicating good to excellent agreement (Table 5). With classification per item, the highest average METRICS score was recorded for the “Model” item, followed by the “Specificity” item, while the lowest scores were recorded for the “Randomization” item (classified as suboptimal) and “Individual factors’ item (classified as satisfactory) (Table 5).
The principal finding of this study was the establishment of 9 studies in health care. Needed efforts aiming to standardize the reporting of AI-based model content testing can be illustrated as follows.

First, the variability in generative AI model types used to conduct queries and the variability in settings pose significant challenges for cross-study comparisons. The significant impact of generative AI models on the resultant output is related to the distinct architectures and capabilities of various generative AI models, with expected variability in the performance and quality of the content generated by generative AI. Consequently, it is important to consider these variations when evaluating research using different generative AI models. These issues can be illustrated clearly in the records included in this study that conducted contemporary analysis of at least two models. For example, Al-Ashwal et al. showed that Bing had the highest accuracy and specificity for predicting drug-drug interaction, outperforming Bard, ChatGPT-3.5, and ChatGPT-4. Moreover, Baglivo et al. showed not only intermodel variability but also intramodel variability in performance in the domain of public health. Additionally, in the context of providing information on rhinoplasty, Seth et al. showed that this intermodel variability in performance was the most comprehensible with the content of Bard, followed by ChatGPT and Bing.

Second, the continuous updating of generative AI models introduces significant temporal variability, which would influence the comparability of studies conducted at different times. The updates of generative AI models enhance their capabilities and performance. Consequently, this temporal variability can lead to inconsistencies in synthesizing evidence, as the same model may demonstrate different outputs over time. Therefore, when analyzing or comparing studies involving AI models, it is crucial to consider the specific version and state of the model at the time of each study to accurately interpret and compare results. In this context, it is important to conduct

Table 5. Interrater reliability per METRICS item.

| METRICSa item                              | Score | Quality  | Cohen κ | Asymptotic standard error | Approximate T | P value  \\n|--------------------------------------------|-------|----------|---------|---------------------------|---------------|----------
| Model                                      | 3.72 (0.58) | 2.5-5.0 | Very good | 0.820 | 0.090 | 6.044 | <.001 |
| Timing                                     | 2.90 (1.93) | 1.0-5.0 | Good | 0.853 | 0.076 | 6.565 | <.001 |
| Count                                      | 3.04 (1.32) | 1.0-5.0 | Good | 0.962 | 0.037 | 10.675 | <.001 |
| Specificity of prompts and language        | 3.44 (1.25) | 1.0-5.0 | Very good | 0.765 | 0.086 | 8.083 | <.001 |
| Evaluation                                 | 3.31 (1.16) | 1.0-5.0 | Good | 0.885 | 0.063 | 9.668 | <.001 |
| Individual factors                         | 2.50 (1.42) | 1.0-5.0 | Satisfactory | 0.865 | 0.087 | 6.860 | <.001 |
| Transparency                               | 3.24 (1.01) | 1.0-5.0 | Good | 0.558 | 0.112 | 5.375 | <.001 |
| Range                                      | 3.24 (1.07) | 2.0-5.0 | Good | 0.836 | 0.076 | 8.102 | <.001 |
| Randomization                              | 1.31 (0.87) | 1.0-4.0 | Suboptimal | 0.728 | 0.135 | 5.987 | <.001 |
| Overall                                    | 3.01 (0.58) | 1.5-4.1 | Good | 0.381 | 0.086 | 10.093 | <.001 |

bThe mean scores represent the results of evaluating the included studies averaged for the 2 rater scores.

Discussion

Principal Findings

The interpretation and synthesis of credible scientific evidence based on studies evaluating commonly used generative AI-based conversational models (eg, ChatGPT, Bing, and Bard) can be challenging. This is related to the discernible variability in the methods used for the evaluation of such models, as well as the varying styles of reporting. Such variability is fathomable considering the emerging nature of this research field with less than a year of reporting at the time of writing. Therefore, a standardized framework to guide the design of such studies and to delineate the best reporting practices can be beneficial, since rigorous methodology and clear reporting of findings are key attributes of science to reach reliable conclusions with real-world implications.

In this study, a preliminary checklist referred to as “METRICS” was formulated, which can help researchers aspiring to test the performance of generative AI-based models in various aspects of health care education and practice. It is crucial to explicitly state that the proposed METRICS checklist in this study cannot be claimed to be comprehensive or flawless. Nevertheless, this checklist could provide a solid base for future studies and much needed efforts aiming to standardize the reporting of AI-based studies in health care.

The principal finding of this study was the establishment of 9 key themes that are recommended to be considered in the design, testing, and reporting of generative AI-based models in research, particularly in the health care domain. These features included the model of the generative AI tool, evaluation approach, timing of testing and transparency, range of topics tested and randomization of queries, individual factors in the design and assessment, count of queries, and specific prompts and languages used. The relevance of these themes in the design and reporting of generative AI-based model content testing can be illustrated as follows.
future longitudinal studies to discern the exact effect of changes in performance of commonly used generative AI models over time.

Third, the count of queries in the evaluation of generative AI models was identified among the pertinent themes of assessment. This appears understandable since studies employing a larger number of queries can provide a more comprehensive evaluation of the tested model. An extensive number of queries can reveal minor weaknesses, despite the difficulty to establish what constitutes an “extensive” number of queries of the minimum number of queries needed to reveal the real performance of a generative AI model in a particular topic. In this study, the number of queries varied from a single case to more than 2500 questions, showing the need for standardization and establishing a clear guide on the number of queries deemed suitable [58, 75].

Fourth, a key theme identified in this study was the nature and language of the prompts used to conduct the studies. This involved the imperative of explicitly stating the used prompts and the language in which they were framed. The exact prompting approach and the presence of cultural and linguistic biases appear to be critical factors that can influence the quality of content generated by generative AI-based models [83]. Slight differences in wording or context in the prompt used to generate the generative AI content can lead to recognizable differences in the content generated [36, 84, 85]. Additionally, feedback mechanisms and learning loops that allow generative AI-based models to learn from interactions can change the model performance for the same query, which might not be consistently accounted for in all studies. These minor variations in prompts across different studies can also complicate the synthesis of evidence, highlighting the need for standardizing such an aspect. Additionally, as highlighted above, generative AI-based models may exhibit biases based on their training data, affecting performance across various cultural or linguistic contexts [86-88]. Thus, studies conducted in different regions or involving various languages might yield varying results. In this study, we found that a majority of the included records tested generative AI-based models using the English language, highlighting the need for more studies on other languages to reveal the possible variability in performance based on language. Comparative studies involving multiple languages can reveal such inconsistencies, for example, the study by Wang et al [74].

In the aforementioned study assessing ChatGPT performance in the Taiwanese pharmacist licensing exam, the performance in the English test was better than that in the Chinese test across all tested subjects [74]. Another comprehensive study by Alfertshofer et al [43] that assessed the performance of ChatGPT on 6 different national medical licensing exams highlighted the variability in performance per country and language. Based on the previous points, more studies that encompass diverse language and cultural perspectives are essential to assess and address possible cultural and language biases in generative AI-based models. Additionally, the design of generative AI-based models trained on a more diverse set of languages and cultural contexts is important to ensure that the training data sets are representative of different linguistic groups, which is of paramount importance in health care. Furthermore, cross-cultural validation studies appear valuable for understanding the performance of generative AI-based models in various language and cultural settings. These approaches could enhance the broad applicability of generative AI-based models in health care to ensure the fair distribution of the benefits of generative AI technologies.

Fifth, an important theme highlighted in this study was the approach of evaluating the content generated by generative AI-based models. Variable methods of assessment can introduce a discernible methodological variability. Specifically, the use of objective assessment ensures consistency in assessment. On the other hand, subjective assessment, even by experts, can vary despite the professional judgment and deep understanding provided by such an expert opinion [89].

Similarly, the number and expertise of evaluators or raters involved in constructing and evaluating the generative AI-based studies were identified as a pertinent theme in this study [90, 91]. Variations in rater numbers across studies can lead to inconsistencies in synthesized evidence [68, 69, 92]. Additionally, the method used to establish agreement (eg, kappa statistics and consensus meetings) might differ in various studies, affecting the comparability of results.

Finally, data-pertinent issues were identified as key themes in this study. These involved the need for full transparency regarding the sources of data used to create the queries (eg, question banks, credible national and international guidelines, clinical reports, etc) [93, 94]. Additionally, ethical considerations, such as consent to use copyrighted material and consent or anonymization of the clinical data, should be carefully stated in the evaluation studies of AI-based models. An important aspect that appeared suboptimal in the majority of included records was randomization to reduce or eliminate potential bias in query selection. Thus, this important issue should be addressed in future studies to allow unbiased evaluation of the content of generative AI-based models. Another important aspect is the need to carefully select the topics to be tested, which can belong to a broad domain (eg, medical examination) or a narrow domain (eg, a particular specialty) [95-97]. A comprehensive description of topics is essential to reveal subtle differences in generative AI performance across various domains. Biased query coverage per topic may result in unreliable conclusions regarding generative AI model performance.

The value of the METRICS checklist in guiding researchers toward more rigorous design and transparent reporting of health care studies assessing the performance of generative AI-based models can be highlighted as follows. Two studies exemplified the practical utility of the METRICS checklist (presented in its preprint form) across different research scenarios (health care education and health care practice) [98]. The first study conducted a detailed assessment of ChatGPT-3.5 performance in medical microbiology multiple choice questions compared with dental students [99]. By applying the METRICS checklist retrospectively, the study quality was delineated, including the identification of potential limitations, such as the absence of randomization, thus offering a more critical evaluation of the research design [99].
The second study investigating the performance of both ChatGPT-3.5 and ChatGPT-4 in the context of diagnostic microbiology case scenarios was conceived based on the METRICS checklist [100]. The prospective incorporation of the METRICS checklist was particularly instrumental in refining the study design and the reporting of results [100].

Thus, the aforementioned studies emphasize the effectiveness of the METRICS checklist as a tool to standardize research practice in a rapidly growing research field. The real-world application of the METRICS checklist has been valuable in identifying potential research limitations and in enhancing the overall structure and clarity of the reporting of results. These studies also demonstrate the value of the METRICS checklist for guiding researchers toward more rigorous design and transparent reporting of generative AI-based studies in health care [99,100].

Limitations
It is crucial to explicitly mention the need for careful interpretation of the findings based on the following limitations. First, the initial search process involved the broad term “artificial intelligence” and was conducted by a single author, which may have inadvertently resulted in missing relevant references. The record selection process was further limited by the screening of record titles for only 3 generative AI models (ChatGPT, Bing, and Bard) using the EndNote search function. Additionally, the reliance on including published English records, indexed in Scopus, PubMed, or Google Scholar, could raise concerns about potential selection bias and the exclusion of relevant studies. However, it is important to consider this limitation in light of the context of our study, which represents a preliminary report that needs to be validated by future comprehensive and exhaustive studies. Second, it is important to acknowledge that a few pertinent themes could have been overlooked despite our attempt to achieve a thorough analysis, given the limited number of authors. Therefore, future follow-up studies can benefit from inclusion of authors with diverse backgrounds, including different health care disciplines, computer scientists, researchers in the field of human-AI interaction, and AI developers. Additionally, the subjective nature of the pertinent theme selection can be considered as another important caveat in this study. This shortcoming extended to involve the raters’ subjective assessments in assigning different METRICS scores. Moreover, the equal weight given to each item of the checklist in the METRICS score might not be a suitable approach, given the possibility of varying importance of each component. Thus, future comprehensive studies should focus on the relative importance of each METRICS component and its possible impact on the reporting of generative AI model performance. Third, the focus on a few specific generative AI-based conversational models (i.e., ChatGPT, Bing, and Bard) can potentially overlook the nuanced aspects of other generative AI models. Nevertheless, our approach was justified by the popularity and widespread use of these particular generative AI-based models. However, it is important for future studies to expand the scope to include models, such as Llama or Claude, which could provide a more comprehensive evaluation of the utility of the METRICS checklist. Lastly, we fully and unequivocally acknowledge that the METRICS checklist is preliminary and needs further verification to ensure its valid applicability.

Future Perspectives
The METRICS checklist proposed in this study could be a helpful step toward establishing useful guidelines to design and report the findings of generative AI-based studies. The integration of generative AI models in health care education and practice necessitates a collaborative approach involving health care professionals, researchers, and AI developers. Synthesis of evidence with critical appraisal of the quality of each element in the METRICS checklist is recommended for continuous enhancement of AI output, which would result in successful implementation of AI models in health care while avoiding possible concerns. Regular multidisciplinary efforts and iterative revisions are recommended to ensure that the METRICS checklist properly reflects its original intended purpose of improving the quality of study design and result reporting in this swiftly evolving research field. Future studies would benefit from expanding the scope of literature review and data inclusion, with the incorporation of a wider range of databases, languages, and AI models. This is crucial for reaching the ultimate aim of standardization of the design and reporting of generative AI-based studies.

Conclusions
The newly devised “METRICS” checklist may represent a key initial step to motivate the standardization of reporting of generative AI-based studies in health care education and practice. Additionally, the establishment of this algorithm can motivate collaborative efforts to develop universally accepted reporting guidelines for the design and reporting of results of generative AI-based studies. In turn, this can enhance the comparability and reliability of evidence synthesis from these studies. The METRICS checklist, as presented by the findings of this study, can help to elucidate the strengths and limitations of the content generated by generative AI-based models, guiding their future development and application. The standardization offered by the METRICS checklist can be important to ensure the reporting of reliable and replicable results. Subsequently, this can result in the exploitation of the promising potential of generative AI-based models in health care while avoiding its possible concerns. The METRICS checklist could mark the significant progress in the evolving research field. Nevertheless, there is room for refinement through revisions and updates to verify its validity.

Data Availability
The data used in this study are available upon request from the corresponding author (Malik Sallam).
Authors' Contributions

Malik Sallam contributed to conceptualization; Malik Sallam, MB, and Mohammed Sallam contributed to methodology; Malik Sallam, MB, and Mohammed Sallam contributed to formal analysis; Malik Sallam, MB, and Mohammed Sallam performed the investigation; Malik Sallam, MB, and Mohammed Sallam contributed to data curation; Malik Sallam contributed to writing—original draft preparation; Malik Sallam, MB, and Mohammed Sallam contributed to writing—review and editing; Malik Sallam contributed to visualization; Malik Sallam contributed to supervision; and Malik Sallam contributed to project administration. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

None declared.

References


35. Eysenbach G. The Role of ChatGPT, Generative Language Models, and Artificial Intelligence in Medical Education: A Conversation With ChatGPT and a Call for Papers. JIMIR Med Educ 2023 Mar 06;9:e46885 [FREE Full text] [doi: 10.2196/46885] [Medline: 36863937]


41. CASP Qualitative Studies Checklist. Critical Appraisal Skills Programme. URL: https://casp-uk.net/casp-tools-checklists/ [accessed 2023-11-10]


92. Ferreira AL, Chu B, Grant-Kels JM, Ogunleye T, Lipoff JB. Evaluation of ChatGPT Dermatology Responses to Common Patient Queries. JMIR Dermatol 2023 Nov 17;6:e49280 [FREE Full text] [doi: 10.2196/49280] [Medline: 37976093]


Abbreviations

AI: artificial intelligence
CASP: Critical Appraisal Skills Programme
METRICS: Model, Evaluation, Timing, Range/Randomization, Individual factors, Count, and Specificity of prompts and language

https://www.i-jmr.org/2024/1/e54704
Interact J Med Res 2024 | vol. 13 | e54704 | p.103
(page number not for citation purposes)
Please cite as:
Sallam M, Barakat M, Sallam M
A Preliminary Checklist (METRICS) to Standardize the Design and Reporting of Studies on Generative Artificial Intelligence–Based Models in Health Care Education and Practice: Development Study Involving a Literature Review
URL: https://www.i-jmr.org/2024/1/e54704
doi:10.2196/54704
PMID:38276872

©Malik Sallam, Muna Barakat, Mohammed Sallam. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 15.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.
Factors Associated With Worsened Mental Health of Health Care Workers in Canada During the COVID-19 Pandemic: Cross-Sectional Survey Study

Ali AbdulHussein1, BEng, MEng, MBA; Zahid Ahmad Butt2, PhD; Stanko Dimitrov1, PhD; Brian Cozzarin1, PhD

1Department of Management Science, Faculty of Engineering, University of Waterloo, Waterloo, ON, Canada
2School of Public Health Sciences, Faculty of Health, University of Waterloo, Waterloo, ON, Canada

Corresponding Author:
Ali AbdulHussein, BEng, MEng, MBA
Department of Management Science
Faculty of Engineering
University of Waterloo
200 University Ave W
Waterloo, ON, N2L 3G1
Canada
Phone: 1 (519) 888 4567
Fax: 1 (519) 888 4567
Email: a9abdulh@uwwaterloo.ca

Abstract

Background: Health care workers (HCWs) in Canada have endured difficult conditions during the COVID-19 pandemic. Many worked long hours while attending to patients in a contagious environment. This introduced an additional burden that may have contributed to worsened mental health conditions.

Objective: In this study, we examine the factors associated with worsened mental health conditions of HCWs as compared to before the start of the pandemic.

Methods: We use data from a survey of HCWs by Statistics Canada. A regression model is used to estimate the odds ratios (ORs) of worsened mental health after the start of the pandemic. The estimated odds ratio (OR) is associated with different independent variables that include demographics (age, sex, immigration status, and geographic area), occupational factors (work status, occupational group, and exposure category), and different access levels to personal protective equipment (PPE).

Results: Of 18,139 eligible participants surveyed, 13,990 (77.1%) provided valid responses. We found that HCWs younger than 35 years old were more likely (OR 1.14, 95% CI 1.03-1.27; P=.01) to exhibit worsened mental health as compared to the reference group (35-44 years old). As for sex, male HCWs were less likely (OR 0.76, 95% CI 0.67-0.86; P<.001) to exhibit worsened mental health as compared to female HCWs. Immigrant HCWs were also less likely (OR 0.57, 95% CI 0.51-0.64; P<.001) to exhibit worsened mental health as compared to nonimmigrant HCWs. Further, HCWs working in Alberta had the highest likelihood of exhibiting worsened mental health as compared to HCWs working elsewhere (Atlantic provinces, Quebec, Manitoba, Saskatchewan, Ontario, British Columbia, and Northern Territories). Frontline workers were more likely (OR 1.26, 95% CI 1.16-1.38; P<.001) to exhibit worsened mental health than nonfrontline HCWs. Part-time HCWs were less likely (OR 0.85, 95% CI 0.76-0.93; P<.001) to exhibit worsened mental health than full-time HCWs. HCWs who reported encountering COVID-19 cases were more likely (OR 1.55, 95% CI 1.41-1.70; P<.001) to exhibit worsened mental health than HCWs who reported no contact with the disease. As for PPE, HCWs who never had access to respirators, eye protection, and face shields are more likely to exhibit worsened mental health by 1.31 (95% CI 1.07-1.62; P<.001), 1.51 (95% CI 1.17-1.96; P<.001), and 1.41 (95% CI 1.05-1.92; P=.02) than those who always had access to the same PPE, respectively.

Conclusions: Different HCW groups experienced the pandemic differently based on their demographic and occupational backgrounds as well as access to PPE. Such findings are important to stakeholders involved in the planning of personalized support programs and aid mental health mitigation in future crises. Certain groups require more attention.

KEYWORDS
health care workers; COVID-19; mental health; demographic factors; occupational factors; access to PPE; pandemic; health care system; psychological trauma; psychological; trauma; respirators; eye protection; face shields; support

Introduction
On March 11, 2020, the World Health Organization declared COVID-19 a global pandemic. The pandemic resulted in devastating health impacts on populations and a crisis within the health care system [1]. This health care system was tasked to handle the unprecedented inflow of patients. Functions within the system that were impacted include emergency departments, intensive care units, physician services, and long-term care units [2]. Health care workers (HCWs) of different occupational groups battled the pandemic. Overwhelmed hospitals in Canada canceled less urgent surgeries by up to 80% by June 2020 [3]. These patterns shifted the workload on HCWs and had an impact on the overall health care system. These conditions not only demanded more hospital capacity but also put an overwhelming strain on HCWs [4]. In addition to the operational pressure, HCWs also suffered from the lack of personal protection equipment (PPE), especially at the beginning of the pandemic [5]. HCWs were at the frontline in battling this pandemic. This battle has put pressure on their mental health conditions [6-8].

In this paper, we assess the various factors associated with worsened HCWs’ mental health conditions as compared to before the start of the pandemic. Although experiencing mental health conditions may be a daily occurrence for some HCWs, the duration and severity during the pandemic were different. HCWs were also at a higher risk of infection, adding to the risk of further mental health conditions [9]. In the past, HCWs have experienced mental health problems during other outbreaks, including the Middle East respiratory syndrome and the severe acute respiratory syndrome [10]. Such conditions have been studied in the literature [11]. In this study, we use a recent data set by Statistics Canada from a national cross-sectional survey that was conducted in the fall of 2020 to assess the impact of COVID-19 on HCWs. Unlike work in the literature, we comprehensively assess the impact of demographic and occupational factors as well as the availability of PPE on HCWs’ mental health conditions [9,12,13].

In the global literature, authors assessed the mental health conditions of HCWs during such outbreaks globally [14-16]. The impact of demographic, social, and occupational factors was reportedly linked to various mental health conditions. Researchers also assessed the prevalence of stress, anxiety, and other psychological well-being indicators of HCWs in Oman during the pandemic [12]. The focus was on young female HCWs who encountered confirmed or suspected COVID-19 cases during their work. Another study in Turkey examined the relationship between the perceived risk of infection and the mental health conditions of HCWs during the COVID-19 pandemic [13]. In addition, a study in the Chinese province of Hubei was conducted early in the pandemic to assess the psychological impact of the pandemic on the frontline medical staff [9]. The study measured the association of factors including professional group, age, and sex factors with work stress.

Our study is among the first to highlight the association of diverse demographic and occupational factors with the mental health condition of HCWs during the pandemic in Canada. We also considered the role of access to PPE on HCWs’ mental health conditions. Unlike existing literature, we studied the individual impact of each of the following factors on the mental health of HCWs while holding the rest constant. Demographic factors include age group, sex, province of the workplace, and immigration status. Occupational factors include work status, frontline category, and exposure to confirmed or suspected COVID-19 cases. Finally, we also considered access to several PPE. The findings of the study will be of prime importance to key stakeholders, including mental health support program planners, health care policy makers, HCWs themselves, and researchers in the area. The goal of this paper is to understand which factors were associated with worsened mental health conditions in HCWs after the start of the pandemic.

Methods

Data Sources, Study Procedure, and Participants
We used a data set from a recent cross-sectional survey by Statistics Canada on the impacts of COVID-19 on HCWs. Unlike other traditional Statistics Canada surveys, a random selection of participants was not used. Instead, Statistics Canada sent an email invitation to HCWs across Canada. Then, a snowball sampling procedure was used. The invitation included a link to a web-based survey that was available through Statistics Canada’s web page. Accordingly, 18,139 responses were collected between November 24 and December 13, 2020, across 7 provincial and territorial regions in Canada. No data were collected for this survey beyond these dates. Only responses by HCWs were included in the data set. The responses were completely anonymized by Statistics Canada.

The questionnaire asked HCWs for information related to the job environment, demographics, geography, and information on access to PPE as background information. Adaptive questioning was used. Our study was limited to these 3 factors categories only: demographics, occupational, and access to PPE based on the available data from Statistics Canada. In assessing mental health well-being, the survey asked HCWs: compared to before the COVID-19 pandemic, how would you say your mental health is now? HCWs self-reported their perceived mental health on a 5-point Likert scale: much better now, somewhat better now, somewhat worse now, somewhat better now, much worse now.

Statistical Analysis Strategy
Our study uniquely studied the association of various interesting factors, as depicted in Figure 1, with HCW mental health conditions as compared to before the start of the pandemic. To consider each factor separately, we used a multivariate ordinal logistic regression [17,18]. The model is defined in Multimedia Appendix 1. The dependent variable is the state of mental health conditions in HCWs after the start of the pandemic.
of the respondents. We reduced the mental health state categories from the 5 mentioned above to 3 (improved, same, or worsened) to yield statistically significant model estimates. The independent variables include demographics such as age, sex, immigration status, and the province of the workplace. Occupational variables include work status, occupational group, and exposure to COVID-19 cases. The model also estimated the association of HCWs’ access to a variety of PPE, such as respirators, eye protection, and face shields. Access to PPE is also reported by the HCW on a 5-point Likert scale, which we reduced to 3 categories (always available, sometimes available, or never available) to yield statistically significant model estimates. All responses are self-reported. Analyses were conducted in RStudio (version 1.4.1717; Posit, PBC).

Figure 1. Factors associated with the mental health of HCWs. HCW: health care worker; PPE: personal protective equipment.

**Ethical Considerations**

Data collected by Statistics Canada were reviewed based on the following principles: privacy, security, transparency, accountability, trust, sustainability, data quality, and fairness, as well as well vetted to be harmless to participants or the public. All ethical approvals were considered by Statistics Canada. No further ethical approval to use these data was required by the authors and the survey is made available to the public. This study is exempt from further ethical approval. Furthermore, the authors also did not have access to information related to the development and testing of the survey. The authors are also not aware of any compensation provided to survey participants.

**Results**

**Participant Profile**

From the full data set of 18,139 responses, we considered nonduplicate responses that provided valid answers to all questions of interest to this study (N=13,990). The remaining responses contained missing or invalid responses. For the considered population, Table 1 outlines the number of observations and percentages for different demographic factors such as age group, sex, and immigration status. The number of observations and percentages were also calculated for occupational factors including the province of the primary workplace, work status (full-time vs part-time), frontline work status, occupational group, and exposure to confirmed or suspected COVID-19 cases. Additionally, the last 3 rows of Table 1 outline PPE access levels for 3 different PPE: face shields, eye protection, and respirators. Overall 4261 (31%) HCWs were younger than 35 years. The majority were female participants (n=12,682, 91%). In total, 12,510 (89%) HCWs were nonimmigrant. HCW workers were distributed across 7 regions, with the largest group of 6626 (47%) working in Ontario. The majority (n=10,152, 73%) of HCWs worked full-time and 5511 (39%) were nonfrontline allied health professionals. A total of 8188 (59%) HCWs did not report exposures to confirmed or suspected COVID-19 cases. As for access to PPE, 10,758 (77%) HCWs always had access to face shields, 10,118 (72%) always had access to eye protection, and 6898 (49%) always had access to respirators.
<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Value, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
</tr>
<tr>
<td>&lt;35</td>
<td>4261 (30.5)</td>
</tr>
<tr>
<td>35-44</td>
<td>3962 (28.3)</td>
</tr>
<tr>
<td>45-54</td>
<td>3415 (24.4)</td>
</tr>
<tr>
<td>≥55</td>
<td>2352 (16.8)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1308 (9.3)</td>
</tr>
<tr>
<td>Female</td>
<td>12,682 (90.7)</td>
</tr>
<tr>
<td><strong>Immigration status</strong></td>
<td></td>
</tr>
<tr>
<td>Nonimmigrant</td>
<td>12,510 (89.4)</td>
</tr>
<tr>
<td>Immigrant</td>
<td>1480 (10.6)</td>
</tr>
<tr>
<td><strong>Province of workplace</strong></td>
<td></td>
</tr>
<tr>
<td>Atlantic provinces</td>
<td>1972 (14.1)</td>
</tr>
<tr>
<td>Quebec</td>
<td>832 (5.9)</td>
</tr>
<tr>
<td>Ontario</td>
<td>6626 (47.4)</td>
</tr>
<tr>
<td>Manitoba and Saskatchewan</td>
<td>1308 (9.3)</td>
</tr>
<tr>
<td>Alberta</td>
<td>2018 (14.4)</td>
</tr>
<tr>
<td>British Columbia</td>
<td>1194 (8.5)</td>
</tr>
<tr>
<td>Northern territories</td>
<td>40 (0.3)</td>
</tr>
<tr>
<td><strong>Work status</strong></td>
<td></td>
</tr>
<tr>
<td>Full-time</td>
<td>10,152 (72.6)</td>
</tr>
<tr>
<td>Part-time</td>
<td>3838 (27.4)</td>
</tr>
<tr>
<td><strong>Frontline occupation</strong></td>
<td></td>
</tr>
<tr>
<td>Physician</td>
<td>397 (2.8)</td>
</tr>
<tr>
<td>Nurse</td>
<td>4689 (33.5)</td>
</tr>
<tr>
<td>Emergency medical personnel</td>
<td>223 (1.6)</td>
</tr>
<tr>
<td><strong>Nonfrontline occupation</strong></td>
<td></td>
</tr>
<tr>
<td>Personal support worker</td>
<td>454 (3.2)</td>
</tr>
<tr>
<td>Allied health professional</td>
<td>5511 (39.4)</td>
</tr>
<tr>
<td>Laboratory worker</td>
<td>1267 (9.1)</td>
</tr>
<tr>
<td>Pharmacist</td>
<td>169 (1.2)</td>
</tr>
<tr>
<td>Dental professional</td>
<td>1280 (9.1)</td>
</tr>
<tr>
<td><strong>Exposure to confirmed or suspected cases</strong></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>5802 (41.5)</td>
</tr>
<tr>
<td>No</td>
<td>8188 (58.5)</td>
</tr>
<tr>
<td><strong>Access to PPE</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Face shields</strong></td>
<td></td>
</tr>
<tr>
<td>Always or usually available</td>
<td>10,758 (76.9)</td>
</tr>
<tr>
<td>Sometimes available</td>
<td>1070 (7.6)</td>
</tr>
<tr>
<td>Never available</td>
<td>357 (2.6)</td>
</tr>
<tr>
<td>Skipped answer</td>
<td>1805 (12.9)</td>
</tr>
<tr>
<td><strong>Eye protection</strong></td>
<td></td>
</tr>
</tbody>
</table>
Analysis Results

In the following, we present the results for estimating the associations between the various factors and the likelihood of worsened mental health conditions for HCWs.

Demographic Factors

Model estimates are expressed as odds ratios (ORs), as presented in Table 2. These ORs indicate the odds of worsened mental health conditions as compared to before the start of the pandemic. Based on the ORs in Table 2, HCWs who are younger than 35 years old were more likely (OR 1.14, 95% CI 1.03-1.27; P=.01) to exhibit worsened mental health conditions than the reference group (35 to 44 years old). Furthermore, those aged 45-54 years and 55 years and older were less likely to exhibit worsened mental health conditions than the reference group (OR 0.71, 95% CI 0.64-0.78; P<.001; and OR 0.55, 95% CI 0.49-0.61; P<.001, respectively). Hence, the older the HCW, the lower the likelihood of worsened mental health conditions. As for sex, male HCWs were less likely (OR 0.76, 95% CI 0.67-0.86; P<.001) to exhibit worsened mental health conditions than their female counterparts. Immigrant HCWs were also less likely (OR 0.57, 95% CI 0.51-0.64; P<.001) to exhibit worsened mental health conditions than nonimmigrants. Geographically, HCWs working in Alberta have the highest likelihood of worsened mental health conditions. HCWs living in Alberta were most likely to exhibit worsened mental health conditions.
Table 2. Estimates for the ordinal regression model for various factors associated with the mental health conditions of HCWs\textsuperscript{a}.

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>( \beta )</th>
<th>SE</th>
<th>OR\textsuperscript{b} (95% CI)</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age group (years; reference: 35 to 44 years)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than 35</td>
<td>0.13</td>
<td>0.05</td>
<td>1.14 (1.03-1.27)</td>
<td>.01</td>
</tr>
<tr>
<td>45-54</td>
<td>-0.35</td>
<td>0.05</td>
<td>0.71 (0.64-0.78)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>55 and older</td>
<td>-0.60</td>
<td>0.06</td>
<td>0.55 (0.49-0.61)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Sex (reference: female)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>-0.28</td>
<td>0.06</td>
<td>0.76 (0.67-0.86)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Immigration status (reference: nonimmigrant)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immigrant</td>
<td>-0.56</td>
<td>0.06</td>
<td>0.57 (0.51-0.64)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Work location (reference: Alberta)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Atlantic provinces</td>
<td>-0.41</td>
<td>0.07</td>
<td>0.66 (0.57-0.77)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Quebec</td>
<td>-0.30</td>
<td>0.09</td>
<td>0.74 (0.61-0.89)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Manitoba and Saskatchewan</td>
<td>-0.09</td>
<td>0.08</td>
<td>0.91 (0.77-1.08)</td>
<td>.27</td>
</tr>
<tr>
<td>Ontario</td>
<td>-0.24</td>
<td>0.06</td>
<td>0.78 (0.70-0.88)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>British Columbia</td>
<td>-0.09</td>
<td>0.09</td>
<td>0.92 (0.77-1.09)</td>
<td>.31</td>
</tr>
<tr>
<td>North Territories</td>
<td>-0.56</td>
<td>0.34</td>
<td>0.57 (0.30-1.15)</td>
<td>.01</td>
</tr>
<tr>
<td><strong>Work status (reference: full-time)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Part-time</td>
<td>-0.16</td>
<td>0.04</td>
<td>0.85 (0.76-0.93)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Job setting (reference: nonfrontline)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frontline</td>
<td>0.23</td>
<td>0.05</td>
<td>1.26 (1.16-1.38)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Contact with patients with COVID-19 (reference: no contact)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exposure</td>
<td>0.44</td>
<td>0.05</td>
<td>1.55 (1.41-1.70)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Access to respirators (reference: did not need)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Always</td>
<td>-0.03</td>
<td>0.05</td>
<td>0.97 (0.88-1.07)</td>
<td>.13</td>
</tr>
<tr>
<td>Sometimes</td>
<td>0.30</td>
<td>0.08</td>
<td>1.34 (1.14-1.59)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Never</td>
<td>0.27</td>
<td>0.10</td>
<td>1.31 (1.07-1.62)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Access to eye protection (reference: did not need)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Always</td>
<td>-0.03</td>
<td>0.06</td>
<td>0.97 (0.85-1.10)</td>
<td>.17</td>
</tr>
<tr>
<td>Sometimes</td>
<td>0.29</td>
<td>0.11</td>
<td>1.33 (1.08-1.65)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Never</td>
<td>0.41</td>
<td>0.13</td>
<td>1.51 (1.17-1.96)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Access to face protection (reference: did not need)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Always</td>
<td>0.26</td>
<td>0.07</td>
<td>1.30 (1.14-1.48)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Sometimes</td>
<td>0.40</td>
<td>0.11</td>
<td>1.50 (1.21-1.86)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Never</td>
<td>0.35</td>
<td>0.15</td>
<td>1.41 (1.05-1.92)</td>
<td>.02</td>
</tr>
</tbody>
</table>

\textsuperscript{a}HCW: health care worker.

\textsuperscript{b}OR: odds ratio.

**Occupational Factors**

Frontline HCWs, such as physicians, nurses, and emergency medical personnel, were more likely (OR 1.26, 95\% CI 1.16-1.38; \( P < .001 \)) to exhibit worsened mental health conditions than nonfrontline workers (personal support workers, allied health professionals, laboratory workers, pharmacists, and professionals). Part-time HCWs, however, were less likely (OR 0.85, 95\% CI 0.76-0.93; \( P < .001 \)) to exhibit worsened mental health conditions. Furthermore, HCWs who reported encountering suspected or confirmed COVID-19 cases were more likely (OR 1.55, 95\% CI 1.41-1.70; \( P < .001 \)) to exhibit worsened mental health conditions.

**Access to PPE**

HCWs who never had access to PPE such as respirators, eye protection, and face shields exhibited the highest likelihood of worsened mental health conditions than those who always had...
access to such PPE. For instance, HCWs who never had access to respirators were more likely (OR 1.31, 95% CI 1.07-1.62; \(P<.001\)) to exhibit worsened mental health conditions than those who did not need this PPE. Similar trends were exhibited with access to eye protection (OR 1.51, 95% CI 1.17-1.96; \(P<.001\)) and face shields (OR 1.41, 95% CI 1.05-1.92; \(P=.02\)).

**Discussion**

**Principal Findings**

The statistical analysis in this study found that HCWs who are younger than 35 years old were found to be more likely to exhibit worsened mental health conditions than HCWs aged 35-44 years. Male HCWs were less likely to exhibit worsened mental health conditions than female HCWs. Immigrant HCWs were also less likely to exhibit worsened mental health conditions than nonimmigrant HCWs. In contrast, HCWs working in Alberta had a higher likelihood of worsened mental health conditions than HCWs working elsewhere (Atlantic provinces, Quebec, Manitoba, Saskatchewan, Ontario, British Columbia, and Northern Territories). Frontline workers were more likely to exhibit worsened mental health conditions than nonfrontline HCWs. Part-time HCWs were less likely to exhibit worsened mental health conditions than full-time HCWs. HCWs who reported encountering COVID-19 cases were more likely to exhibit worsened mental health conditions than HCWs who reported no contact with the disease. As for PPE, HCWs who never had access to respirators, eye protection, and face shields were more likely to exhibit worsened mental health conditions than those who always had access to the same PPE, respectively.

**Comparison With Prior Work**

Previous research has shown that shock events can result in psychological trauma to HCWs [19-21]. To the best of our knowledge, our study is the first to assess the association of a comprehensive variety of factors (demographics, occupational factors, and access to PPE) independently with the mental health conditions of HCWs.

With regard to demographic factors, a study found sex to be associated with worsened mental health conditions of HCWs. Female medical staff exhibited a higher incidence of severe anxiety than their male counterparts [22]. Sex was also found as a predictor of increased anxiety and distress in a study that found female HCWs more vulnerable to such conditions than male HCWs [23]. Our study findings corroborate these previous findings. Similarly, for age, our results demonstrated that older HCWs exhibited less likelihood of worsened mental health conditions. This agrees with findings from a study in the Middle East that found older workers enjoyed better mental health conditions than younger workers [24]. Older HCWs with longer work experience seemed to have handled the pandemic better than their younger counterparts. Another study found that HCWs 40 years or older were less likely to report higher anxiety during the pandemic than younger HCWs [25]. Our results confirm such findings as well.

As for occupational factors, our findings align with a cross-sectional study conducted in Oman, which found frontline workers to be more likely to have anxiety and sleep problems [26]. This was attributed to frontline workers’ increased awareness of the mortality rate of COVID-19 and their fear of contracting the virus. Similar findings were reported by a study conducted in China as well [27]. Exposure to COVID-19 cases has been recognized as a risk factor associated with an increased likelihood of mental health issues in the literature. For instance, in France, female urologists working on the frontline were 1.41 times more likely to feel a degree of stress during their duties. Of this group, those who worked in a department where patients with COVID-19 are treated were 1.85 times more likely to report a degree of stress during work duties [28]. This was attributed to the workers’ fear of infection and the spread of the virus. Direct involvement with COVID-19 care is also found to be highly associated with fear, depression, and anxiety as compared with those working under lower risk conditions [29]. The authors attributed this to the workers’ fear of bringing the virus to their families at home.

As for the work setting, contrary to our findings, the literature reports higher levels of fear and anxiety among part-time HCWs than among full-time [30]. Another study in France found part-time HCWs to exhibit a greater association with distress [31]. Given the small number of part-time HCWs in our study, caution should be observed in interpreting the results. Regarding factors related to access to PPE, access to PPE was found to be associated with better health and less stress [24]. Lack of access to PPE was found to be a major source of HCW stress in a study that surveyed emergency physicians [32]. In Canada, a study found that inadequate PPE supply is associated with increased symptoms of anxiety and depression among HCWs [33]. Workers were concerned about the ability to access sufficient PPE during work hours. These findings are in line with our results. Compared to studies in the literature, our study is the first comprehensive study that assesses the association of a diverse pool of factors including demographics, occupational, and access to PPE, with HCWs’ mental health conditions.

**Implications: Mental Health Programming**

Our comprehensive study sheds light on the association of various factors with HCWs’ mental health after the start of the pandemic. It helps in understanding the vulnerability of various HCW groups to mental health during such events. Certain groups were at substantially higher risk of exhibiting worsened mental health conditions after the pandemic, hence the need for a specialized support program to target this group. As a direct implication, the findings can be used to inform guidelines for mental health support for HCWs during future public health emergencies. Such mental health support may be directed more specifically to more vulnerable groups. Literature has pointed to the importance of an evidence-based approach to designing mental health support programs for HCWs [34,35]. For instance, a study in Alberta during the pandemic pointed to the importance of understanding HCWs’ occupational settings and mental health mitigation techniques [16]. Others have discussed the importance of support programs for HCWs during the pandemic based on need and background [36,37]. More importantly, many studies highlighted the importance of designing personalized mental health support programs based on various factors including demographic and occupational factors [38,39]. Some of these programs used mobile technology to offer mental health

https://www.i-jmr.org/2024/1/e50064

*AbdulHussein et al.*

*Interactive Journal of Medical Research* 2024 | vol. 13 | e50064 | p.111

*(page number not for citation purposes)*
support to HCWs [35,40]. As such, our study aims to inform stakeholders of such factors associated with HCW groups most vulnerable to future events such as the pandemic and health crises.

Limitations and Future Directions
This study has several limitations. First, the scope of the study was limited to assessing mental health in general without measuring different mental condition forms such as depression, anxiety, and fear. This is due to the limited data offered by the survey we used. Second, participants were asked, “ Compared to before the COVID-19 pandemic, how would you say your mental health is now?” That is, HCWs had to objectively compare their mental health state from the time they received the survey to sometime in the past. There may be some variation in how HCWs perceive this comparison. Third, data collected were at a single point in time, which limits the ability to compare mental health assessments to later periods in the pandemic and during subsequent pandemic waves. Fourth, no information about HCW medical history is collected in this survey. In future research, we recommend collecting data at different points in the given event to allow for a time-series analysis and comparison at different crucial points of the event. We also suggest the collection of distress information and whether the conditions persisted beyond the initial shockwave. HCWs’ years of work experience and ethnicity information can also be useful in future studies. We also believe medical history, particularly a preexisting history of mental health disorders, is an important factor associated with the mental health conditions of HCWs and should be considered in future studies.

Conclusions
This study investigated the factors associated with worsened mental health conditions of HCWs in Canada during the pandemic. Our study suggested the association of various factors with the likelihood of HCWs exhibiting worsened mental health conditions as compared to feeling neutral and better. In agreement with the literature, our findings concluded that younger (vs older), female (vs male), nonimmigrant (vs immigrant), full-time (vs part-time), and frontline (vs nonfrontline) HCWs living in Alberta (vs other provinces) exhibited a higher likelihood of worsened mental health conditions than those who felt neutral or better. Those who reported concerns about access to PPE also exhibited the same trend. Such findings can guide the future development of health care programming and inform mental health support planning for HCWs. COVID-19 is a shocking event that introduced uncertainty to the health care system.

Acknowledgments
We would like to thank Statistics Canada for providing the data used in this study.

Data Availability
The data that support the findings of this study is available through Statistics Canada.

Authors’ Contributions
AA processed and cleaned the data and wrote the first draft of the manuscript with guidance from ZAB. BC and SD conducted iterative reviews and suggested revisions to the draft. All authors offered substantial contributions to the discussion and analysis presented in this manuscript. All authors agreed to publish the final draft of this paper.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Logistic model.

References


**Abbreviations**

HCW: health care worker
OR: odds ratio
PPE: personal protective equipment
Factors Associated With Worsened Mental Health of Health Care Workers in Canada During the COVID-19 Pandemic: Cross-Sectional Survey Study

Abdul Hussein A, Butt ZA, Dimitrov S, Cozzarin B

Please cite as:
Abdul Hussein A, Butt ZA, Dimitrov S, Cozzarin B
Factors Associated With Worsened Mental Health of Health Care Workers in Canada During the COVID-19 Pandemic: Cross-Sectional Survey Study
URL: https://www.i-jmr.org/2024/1/e50064
doi: 10.2196/50064
PMID: 38358785

©Ali Abdul Hussein, Zahid Ahmad Butt, Stanko Dimitrov, Brian Cozzarin. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 15.02.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.
A Web-Based Tool to Assess Social Inclusion and Support Care Planning in Mental Health Supported Accommodation: Development and Preliminary Test Study

Sharon Eager, MSc; Helen Killaspy, MBBS, PhD; Joanna C, MSc; Gillian Mezey, MBBS; Peter McPherson, DPsych; Megan Downey, BSc; Georgina Thompson, MSc; Brynmor Lloyd-Evans, PhD

1Division of Psychiatry, University College London, London, United Kingdom
2Camden and Islington NHS Foundation Trust, London, United Kingdom
3Population Health Research Institute, St George’s, University of London, London, United Kingdom

Corresponding Author:
Brynmor Lloyd-Evans, PhD
Division of Psychiatry
University College London
Maple House
149 Tottenham Court Road
London, W1T 7NF
United Kingdom
Phone: 44 (0)20 7679 9428
Email: b.lloyd-evans@ucl.ac.uk

Abstract

Background: Individuals with severe mental illness living in supported accommodation are often socially excluded. Social inclusion is an important aspect of recovery-based practice and quality of life. The Social Inclusion Questionnaire User Experience (SInQUE) is a measure of social inclusion that has been validated for use with people with mental health problems. Previous research has suggested that the SInQUE could also help support care planning focused on enabling social inclusion in routine mental health practice.

Objective: This study aims to develop a web-based version of the SInQUE for use in mental health supported accommodation services, examine its acceptability and perceived usefulness as a tool to support care planning with service users, determine the extent of uptake of the tool in supported accommodation settings, and develop a program theory and logic model for the online SInQUE.

Methods: This study involved a laboratory-testing stage to assess the acceptability of the SInQUE tool through “think-aloud” testing with 6 supported accommodation staff members and a field-testing stage to assess the acceptability, utility, and use of the SInQUE tool over a 5-month period. An implementation strategy was used in 1 London borough to encourage the use of the SInQUE. Qualitative interviews with 12 service users and 12 staff members who used the tool were conducted and analyzed using thematic analysis. The use of the SInQUE was compared with that in 2 other local authority areas, 1 urban and 1 rural, where the tool was made available for use but no implementation strategy was used.

Results: Overall, 17 staff members used the SInQUE with 28 different service users during the implementation period (approximately 10% of all service users living in supported accommodation in the study area). The staff and service users interviewed felt that the SInQUE was collaborative, comprehensive, user-friendly, and relevant. Although some staff were concerned that particular questions might be too personal, service users did not echo this view. Participants generally felt that the SInQUE could help identify individuals’ priorities regarding different aspects of social inclusion by prompting in-depth conversations and tailoring specific support to address service users’ inclusion goals. Some interviewees also suggested that the tool could highlight areas of unmet or unmeetable needs across the borough that could feed into service planning. The SInQUE was not used in the comparison areas that had no implementation strategy.

Conclusions: The online SInQUE is an acceptable and potentially useful tool that can be recommended to assess and support care planning to enable social inclusion of people living in mental health supported accommodation services. Despite this, uptake rates were modest during the study period. A concerted implementation strategy is key to embedding its use in usual care, including proactive endorsement by senior leaders and service managers.
KEYWORDS
social inclusion; supported accommodation; mental health; digital health; care planning

Introduction

Background
Social inclusion refers to an individual’s ability to participate in important societal activities and their sense of community belonging [1,2]. Someone may feel socially excluded if they do not have opportunities for societal involvement and integration, often because of external factors that are beyond their control [3]. Social exclusion is a multifaceted continuum [2], typically signified by poverty, unemployment, inequality, and poor health [4].

People with serious mental illness are thought to be among the most socially excluded groups in society [5]. Individuals with this type of mental health problems often have smaller and less satisfying social networks [6], lower household income [7], and lower levels of employment [8,9] and experience more criminal and violent victimization [10,11] than those in the general population. Social exclusion can be conceptualized as both a cause and a consequence of mental illness [12]. Furthermore, greater social inclusion is associated with better quality of life and lower levels of loneliness among those with severe mental illness, suggesting that social exclusion is an important area for mental health practitioners to try to address [13,14].

Mental health supported accommodation services provide care and support to individuals with particularly severe and complex mental health problems as a way of supporting recovery in the community [15]. It is estimated that there are approximately 100,000 people living in mental health supported accommodation in England. Services are typically staffed by support workers, with additional specialist clinical input provided by National Health Service (NHS) community mental health teams [16]. In England, three main types of supported accommodation are provided: (1) residential care homes for those with the highest needs that comprise 24-hour–staffed communal facilities where placements are not time limited, with meals, supervision of medication, cleaning, and activities provided to service users; (2) supported housing services that provide shared or individual, self-contained, and time-limited tenancies with staff based on-site up to 24 hours a day to assist service users in gaining skills to move on to less supported accommodation; and (3) floating outreach services that provide visiting support for a few hours per week to people living in permanent, self-contained, and individual tenancies with the aim of reducing support over time to zero [16].

Service users living in supported accommodation are often socially isolated, with low levels of employment and little involvement in civil and political processes [17]. Many report feeling lonely and isolated and experiencing a high level of stigma that causes them to become more socially isolated [18]. There is evidence that users of mental health supported accommodation services report a variety of unmet needs, such as accessing employment opportunities and forming intimate relationships [19,20]. However, relatively little research has been conducted to determine the precise needs of service users living in supported accommodation [21], and a greater focus on this group is needed to identify and implement interventions that are likely to be the most useful for them [22].

Supporting service users to work toward desired goals and community engagement is highly congruent with recovery-based practice in mental health. Recovery-based practice recognizes and builds on service users’ strengths and promotes empowerment through collaboration between them and staff to identify and work toward specific goals [23]. Many of the identified goals are markers of social inclusion, such as employment, social network development, and participation in community activities [24]. There is qualitative evidence from a large national research program suggesting that staff working in mental health supported accommodation services operate with a considerable degree of recovery orientation [24,25], and the more recovery-orientated these services are, the more likely people are to move on successfully to more independent settings [26].

People living in mental health supported accommodation have expressed a strong preference for individually tailored services that offer choice and promote autonomy, consistent with a recovery-based approach [18]. Patient-reported outcome measures have been recommended to inform this individualized approach by directly capturing service users’ perspectives on constructs such as goal attainment, quality of life, and social inclusion [27]. Such measures enable service users to make informed decisions about their own support and care planning, in line with World Health Organization recommendations for recovery-based practice in community care provision [28]. Resources delivered across web-based platforms, particularly those that offer guided support, have been established as accessible, acceptable, and effective for use by participants with severe mental illness [29,30]. A tailor-made web-based assessment tool, the Quality Indicator for Rehabilitative Care for supported accommodation, has also been successfully used by managers of supported accommodation services, suggesting that these settings have the required resources and expertise to implement online measures [31].

The Social Inclusion Questionnaire User Experience (SInQUE) was developed as a measure of social inclusion for individuals with severe mental illness [32]. The measure has been validated across a range of mental health populations, has established reliability, is considered acceptable to service users, and has been proposed as being potentially cross-culturally suitable [32-34]. To date, the SInQUE has been used solely in offline research contexts. However, stakeholder feedback from a previous study testing the SInQUE indicated that the measure may be useful in clinical practice to assess social inclusion, facilitate important conversations with service users, and guide care and support planning [34]. Furthermore, a consistent research recommendation from the developers of the SInQUE
tool has been to investigate whether the measure has utility as a care-planning tool to promote social inclusion in routine mental health practice [13,34].

This study aimed to develop a web-based version of the SInQUE for use in mental health supported accommodation services. We sought to examine the acceptability and perceived usefulness of this tool among supported accommodation staff and service users as a means to assess their needs for greater social inclusion and promote care planning.

**Aims**

The study aims were as follows:

1. To develop and refine a web-based version of the SInQUE social inclusion assessment tool tailored for use in mental health supported accommodation settings.
2. To investigate the acceptability and perceived utility of the online SInQUE tool among supported accommodation staff and service users.
3. To determine the extent of uptake of the tool in supported accommodation settings with and without a locally developed implementation strategy to support its use.
4. Informed by the study findings, to develop a program theory and logic model for the online SInQUE specifying its anticipated outcomes, the mechanisms through which they may be achieved, and contextual factors affecting the use and experience of the SInQUE.

**Methods**

**Study Design**

This study comprised two stages conducted in 1 inner London borough:

1. A laboratory-testing stage to assess initial acceptability of the tool and develop it through “think-aloud” testing and semistructured interviews with supported accommodation staff.
2. A field-testing stage to assess wider acceptability, feasibility, and use of the tool. Semistructured interviews were conducted with staff and service users who had used the online SInQUE during this stage.

The 5-month field-testing stage was supported by a local implementation strategy developed in collaboration with local service leads to support the use of the online SInQUE by supported accommodation staff in the participating London borough. The online SInQUE was also made available to supported accommodation services by local service leads in 2 other areas without any accompanying implementation strategy.

**Description of the SInQUE Tool**

The web-based version of the SInQUE [35] can be used to assess social inclusion and inform support and care planning for people with mental health problems. It is designed to be used by staff as part of routine care planning to be completed collaboratively with service users. It can be used on a computer, tablet, or mobile device. Staff are required to register for an account on the SInQUE site using their work email address and details of their organization and can then use the tool for free. No personal data identifying service users are logged or stored on the SInQUE platform. The online SInQUE generates a unique reference number for each new service user, which is retained by the staff member completing the assessment for future reference and to link any repeat assessments.

The online tool developed for laboratory testing in our study included the 46-item version of the validated SInQUE social inclusion questionnaire, which was refined following stakeholder feedback at the end of the previous measure development study [34]. For this study, we removed 1 question from the SInQUE that asked whether the respondent was living alone as this was considered redundant for people living in residential care and supported housing. The SInQUE’s psychometric properties have been established among people with a range of mental health problems receiving input from community mental health services in previous studies [13,32,34]. Although the removal of a question from the original SInQUE questionnaire compromised its established psychometric properties, this minor adaptation is unlikely to have disrupted them substantially. We wanted to minimize changes to maintain the online SInQUE’s similarity to the validated measure and did not aim to make further significant refinements to the content of the tool. Instead, we wanted to gain feedback on its acceptability and feasibility for use in its digital format among staff and users of mental health supported accommodation services to assist in care and support planning.

The online SInQUE questionnaire yields a total score of 0 to 75, with a higher score indicating greater social inclusion. The questions and subscale scores are grouped into 9 different areas of social inclusion: leisure, social relationships, religious and cultural activities, education and employment, transport, health, crime victimization, home life and housing, and civic duties. These areas cover the 5 social inclusion domains of the validated SInQUE (social integration, productivity, consumption, access to services, and political engagement), but the aforementioned 9 areas were considered more immediately understandable for use in practice.

Using the service user’s responses, the online SInQUE generates a list of areas in which the person has said that they would like to be more socially included. It then offers a prompt for the service user and staff member to collaboratively select up to 3 priority areas that they would like to integrate into the person’s support plan. Once the assessment is completed, a summary report is generated. If the assessment is repeated with the same service user in the future, this report will also display changes in their social inclusion over time. The tool can generate management-level summary reports for each organization that is registered with it and commissioner-level summary reports of services using the tool across an entire area (such as a London borough). Multimedia Appendix 1 provides a full description of the SInQUE using the TIDieR (Template for Intervention Description and Replication) checklist [36].

**Setting**

This study took place in mental health supported accommodation services across 1 inner London borough. There are 21 such services in the borough run by 6 different voluntary sector organizations. They offer varying degrees of support to >270
service users who are also supported by local NHS secondary mental health services. In the borough, there are approximately 24 service users living in residential care, 159 living in supported housing, and 89 who receive floating outreach support. Supported housing services offer 24-hour support to 119 people and “9 to 5” support to 40 individuals.

**Laboratory-Testing Stage: Recruitment, Data Collection, and Analysis**

In total, 6 supported accommodation staff members were recruited to provide initial impressions of the online SInQUE tool. We discussed the study with service managers working in 3 different services and asked them to nominate 2 staff members each from their service who were interested in taking part. Participants were purposively sampled to include staff working in floating outreach support, 24-hour supported housing, and residential care. We asked the managers of each of the 3 services to ask for volunteers from their staff teams. Interviews were arranged with the first 2 staff members identified by the manager of each service.

Data were collected between January 2022 and February 2022. All 6 think-aloud interviews with staff members were conducted and recorded using Microsoft Teams (Microsoft Corp). The researcher first discussed the information sheet with each participant and gave them the opportunity to ask questions. Following this, participants’ consent was verbally collected and audio recorded separately from the main part of the interview. Participants were then asked to fill out a short online form providing their demographic information before beginning the interview.

We conducted “think-aloud” testing of the online tool with staff using a semistructured topic guide developed by the study team (Multimedia Appendix 2). Following a process used previously in developing web-based tools [37], participants were asked to complete set tasks using the online SInQUE tool while providing a continuous commentary on their thoughts. They were asked to open the SInQUE website, register for an account, and complete an assessment as they would with a service user. At all stages, they were prompted to share their thoughts as they navigated the website and offer their initial impressions on how easy it was to understand and use and its potential suitability for their work. Once participants had completed the questionnaire, the researcher asked broader questions about their experience using the tool and any ways in which it could be improved. Throughout these interviews, participants were asked to focus on their experience using the SInQUE tool rather than offering specific feedback on individual SInQUE items. This was because we did not intend to make substantial modifications to the SInQUE questions to maintain their scope and similarity to those of the validated SInQUE measure.

Identified problems and suggestions for improvements to the online tool were collated by the researcher following the interviews. They were then reviewed by the study team, decisions about refinements to the online SInQUE were agreed upon, and the tool was revised accordingly.

**Field-Testing Stage: Recruitment, Data Collection, and Analysis**

**Local Implementation Strategy**

The revised version of the tool was made available for use in mental health supported accommodation services across the participating inner London borough. We iteratively developed and implemented a strategy to encourage and support its use by supported accommodation staff in the borough over a 5-month period beginning on May 11, 2022. This implementation strategy was informed by consultation with supported accommodation service managers and clinicians working in the Islington community mental health rehabilitation team and by individual interviews conducted with supported accommodation service users and staff.

**Interviews With Field-Testing Participants**

Participants and Recruitment

Individual interviews with service users (n=12) and staff (n=12) who tried out the SInQUE tool were conducted from late May 2022 to September 2022. This number was chosen to explore the views of staff and service users from a variety of supported accommodation types and service providers. Following our implementation strategy, we asked staff members to alert the study researcher once they had tried the tool in practice. Any staff member or service user in supported accommodation who tried the SInQUE tool was eligible to participate in an individual interview.

Once a staff member informed the study researcher that they had tried the tool, we asked them whether they would like to participate in an individual interview about their experience. We also invited staff to pass on information about the study to the service users with whom they had used the tool and ask them whether they would like to participate in an interview about their experience. If the service user was interested in taking part, the researcher communicated with them either directly or through the staff member they had completed the SInQUE assessment with depending on their preference. Toward the end of the recruitment stage, we recruited the final few staff members and service users purposively to ensure that participants were from a range of supported accommodation types and provider organizations.

One service user interview and 1 staff member interview were conducted online via Microsoft Teams; all other interviews were carried out in person according to participants’ preferences. In-person interviews were conducted by the study researcher at the staff supported accommodation sites, aside from 1 interview with a service user receiving floating outreach support, which was conducted at their home.

**Measures and Procedures**

The researcher first discussed the information sheet with the participants and gave them the opportunity to ask questions about the study. For in-person interviews, informed consent was collected via a paper consent form; for online interviews, verbal consent was audio recorded. Participants were then asked to answer brief demographic questions about themselves and their associated services. Following this, the researcher asked
each participant questions about their experience using the SInQUE; whether there were any ways in which it could be improved; the appropriateness of the online tool for use in their work; and what impacts, if any, they thought it might have on care provision and service users’ experience. The interview topic guides (one for staff participants and one for service user participants) were developed by the study team as semistructured interviews—they are provided in Multimedia Appendix 3. In-person interviews were recorded using a digital voice recorder; online interviews were recorded on Microsoft Teams. Interview audio recordings were transcribed by a professional transcription company with which University College London (UCL) had a data-sharing and privacy agreement. Interview transcripts were then checked by the study researcher for accuracy. Any potentially identifiable text was anonymized. The resulting cleaned transcripts were then securely stored on the UCL university system.

Analysis
The analysis of the interviews comprised 2 stages. First, the study researcher noted any problems experienced by participants and recorded improvements to the online SInQUE they suggested. These issues and the suggested changes were reviewed by the study team, as in the previous laboratory-testing phase. Minor modifications to the online SInQUE were agreed upon, and we made adjustments to the tool in line with this.

Second, transcripts were uploaded to NVivo (version 12; QSR International) for qualitative analysis. As we aimed to develop a program theory for the online SInQUE intervention, we initially coded data into a deductively derived framework that used an intervention-context-actor-mechanism-outcome (ICAMO) configuration, with each component of this ICAMO framework representing a primary theme [38]. Within each of these 5 primary themes, we inductively derived subthemes from the data using thematic analysis. The initial coding was conducted by the lead author (SE) and was then reviewed and adjusted collaboratively by the study team. This included gaining lived-experience perspectives from a researcher with experience of mental health service use (JC) and clinical insights from a senior clinical academic working in the participating borough as a consultant rehabilitation psychiatrist supporting service users who live in supported accommodation (HK). The team brought in further perspectives from those with backgrounds in social work (BLE), clinical psychology (PM), and forensic psychiatry (GM) and from the community rehabilitation team in the borough (MD).

Data Use Monitoring
Data on the uptake and use of the online SInQUE tool were collected from the online SInQUE informatics for the 5-month field-testing period from May 11, 2022, to October 11, 2022. At the start of this period, the study team also contacted local mental health service rehabilitation and housing leads in 2 other areas: another inner London borough and a rural county in the west of England. These service leads contacted local supported accommodation managers and invited them to use the online SInQUE in their service if they wished. The tool was made available to 7 supported accommodation services in the London borough and 10 in the rural county. No further encouragement to use the tool or implementation support was provided. This allowed us to monitor uptake and use of the tool in 2 areas without an associated implementation plan, thus making inferences about the necessity and impact of the strategy we developed.

Logic Model Development
The study team developed a preliminary logic model for the online SInQUE in planning this study. We used the findings of the aforementioned research activities to review and refine this logic model and develop an updated theory about the potential outcomes for service users and organizations from using the online SInQUE; mechanisms through which these outcomes are achieved; and factors influencing the uptake, experience, and impact of the online tool. Factors were related to (1) the intervention itself, (2) the characteristics and attitudes of staff and service users using the online SInQUE, and (3) the broader organizational and societal context. This was summarized in a logic model in the form of an “ICAMO map” [38], which was developed and refined iteratively through discussion with the study team.

Ethical Considerations
The initial laboratory-testing phase of this study (Supporting social inclusion for people with serious mental illness living in supported housing [SUSHI] phase 1) was approved by the UCL Research Ethics Committee (REC) on June 18, 2021 (REC reference 6711/002). The subsequent field-testing phase (SUSHI phase 2) was approved by the London – Camden and Kings Cross NHS REC on November 4, 2021 (REC reference 21/LO/0657). Written or audio-recorded informed consent was obtained from all participants before they took part, and they were clearly informed that they could opt out of the study at any time. All the study data were carefully deidentified. Service user participants were offered a £20 (US $25.14) shopping voucher to thank them for their time.

Results
Participants
We recruited 6 supported accommodation staff members for the “think-aloud” interviews during the laboratory-testing stage. We recruited a further 12 staff members and 12 supported accommodation service users for the individual interviews as part of the field-testing stage. Participant characteristics for both stages are summarized in Table 1.
<table>
<thead>
<tr>
<th>Participant characteristics</th>
<th>Laboratory testing</th>
<th>Field testing</th>
<th>Service users (n=12), n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Staff (n=6), n (%)</td>
<td>Staff (n=12), n (%)</td>
<td>Staff (n=12), n (%)</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>2 (33)</td>
<td>5 (42)</td>
<td>11 (92)</td>
</tr>
<tr>
<td>Female</td>
<td>4 (67)</td>
<td>6 (50)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Nonbinary</td>
<td>0 (0)</td>
<td>1 (8)</td>
<td>1 (8)</td>
</tr>
<tr>
<td><strong>Age group (y)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-30</td>
<td>0 (0)</td>
<td>8 (67)</td>
<td>4 (33)</td>
</tr>
<tr>
<td>31-50</td>
<td>2 (33)</td>
<td>2 (17)</td>
<td>3 (25)</td>
</tr>
<tr>
<td>≥51</td>
<td>4 (67)</td>
<td>1 (8)</td>
<td>4 (33)</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>0 (0)</td>
<td>1 (8)</td>
<td>1 (8)</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian/Asian British</td>
<td>1 (17)</td>
<td>0 (0)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Black/Black British</td>
<td>3 (50)</td>
<td>2 (17)</td>
<td>4 (33)</td>
</tr>
<tr>
<td>White/White British</td>
<td>2 (33)</td>
<td>8 (67)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Mixed/multiple ethnic groups</td>
<td>0 (0)</td>
<td>1 (8)</td>
<td>6 (50)</td>
</tr>
<tr>
<td>Other ethnic background</td>
<td>0 (0)</td>
<td>1 (8)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Sexual orientation</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heterosexual</td>
<td>N/A</td>
<td>N/A</td>
<td>6 (50)</td>
</tr>
<tr>
<td>Gay/lesbian</td>
<td>N/A</td>
<td>N/A</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Bisexual</td>
<td>N/A</td>
<td>N/A</td>
<td>2 (17)</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>N/A</td>
<td>N/A</td>
<td>3 (25)</td>
</tr>
<tr>
<td><strong>Type of supported accommodation lived or worked in</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Floating outreach support</td>
<td>2 (33)</td>
<td>2 (17)</td>
<td>2 (17)</td>
</tr>
<tr>
<td>9-to-5 supported housing</td>
<td>2 (33)</td>
<td>1 (8)</td>
<td>2 (17)</td>
</tr>
<tr>
<td>24-h supported housing</td>
<td>2 (33)</td>
<td>7 (58)</td>
<td>8 (67)</td>
</tr>
<tr>
<td>Residential care</td>
<td>0 (0)</td>
<td>2 (17)</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Length of time worked or lived in supported accommodation (y)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;2</td>
<td>1 (17)</td>
<td>7 (58)</td>
<td>5 (42)</td>
</tr>
<tr>
<td>2-5</td>
<td>3 (50)</td>
<td>3 (25)</td>
<td>5 (42)</td>
</tr>
<tr>
<td>6-10</td>
<td>1 (17)</td>
<td>2 (17)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>≥10</td>
<td>1 (17)</td>
<td>0 (0)</td>
<td>1 (8)</td>
</tr>
<tr>
<td>Prefer not to say</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (8)</td>
</tr>
</tbody>
</table>

aN/A: not applicable; staff were not asked about their sexual orientation.

**Changes Made to the SInQUE**

Following phase 1 laboratory testing and phase 2 field-testing, suggestions that participants made for how the tool could be improved were collated and reviewed by the team. Accordingly, adjustments were made to the online SInQUE after each stage, an overview of which can be found in Table 2. This addressed aim 1 of this study.
Table 2. Changes made to the online Social Inclusion Questionnaire User Experience (SInQUE) following phase 1 and phase 2 testing. All changes were made following the initial laboratory-testing stage unless indicated otherwise.

<table>
<thead>
<tr>
<th>Section of the SInQUE affected</th>
<th>Explanation of the problem</th>
<th>Resolution</th>
<th>Justification for the change</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Registration and use changes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The online SInQUE home page</td>
<td>Some staff members and service users suggested developing additional materials to help explain the SInQUE.</td>
<td>Developed a guidance manual for service managers and commissioners with information on using the SInQUE as well as an informational leaflet and poster aimed at service users about the SInQUE.</td>
<td>It is easier for managers, commissioners, staff, and service users to understand and use the SInQUE.</td>
</tr>
<tr>
<td>The initial page where staff members are asked to register for the SInQUE</td>
<td>Staff were asked to enter the organization they worked for in a free-text box. Some found it confusing to know which organization name they should enter.</td>
<td>Changed response options to a fixed-response drop-down menu with all housing providers in the borough and an “other” free-text option.</td>
<td>Allows for compilation of service-level data and is easier for staff to navigate.</td>
</tr>
<tr>
<td>The page where staff members enter details to set up a new SInQUE assessment</td>
<td>Some staff members thought that further information on the exact purpose of the questionnaire and how it should be administered would be useful in the introductory paragraph describing the assessment.</td>
<td>Additional guidance on how the questionnaire should be administered was added to the introduction paragraph of the SInQUE.</td>
<td>Important contextual information for the questionnaire was explicitly clarified.</td>
</tr>
<tr>
<td>The page where staff members enter details to set up a new SInQUE assessment</td>
<td>Some staff members found the wording of the following question—“Please select the type of accommodation in which the service user is living from the list below”—to be ambiguous and confusing.</td>
<td>Changed the wording of the question to the following: “Please select the type of housing support the service user receives from the list below.”</td>
<td>The clarity of the question improved.</td>
</tr>
<tr>
<td><strong>Changes to the wording of SInQUE questions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The section covering “leisure” questions</td>
<td>Some staff members felt that it was unclear what the “Other” option meant in the context of question 3f: “Over the past year have you been to...Other?”</td>
<td>Changed the wording of this subquestion to the following: “Other leisure activity?”</td>
<td>The clarity of the question improved.</td>
</tr>
<tr>
<td>The section covering “leisure” questions</td>
<td>Some staff members and service users felt that question 7—“Do you spend time in pubs or cafés?”—was worded in a way that was potentially inappropriate for people who do not drink alcohol.</td>
<td>Changed the wording of the question to the following: “Do you go out for a coffee/drink (e.g. to a café or pub, etc) at least once a week?”</td>
<td>The clarity of appropriateness of the question improved.</td>
</tr>
<tr>
<td>The section covering “social” questions</td>
<td>Some staff members were unsure whether question 9—“How many people, outside those in your care team, could you confide in?”—related to a professional or personal care team.</td>
<td>Changed the wording of the question to the following: “How many people, outside the workers in your care team, could you confide in?”</td>
<td>The clarity of the question improved.</td>
</tr>
<tr>
<td>The section covering “home life/housing” questions</td>
<td>Some staff members thought that question 36—“What kind of accommodation do you live in?”—was worded ambiguously.</td>
<td>Changed the wording of the question to the following: “What type of housing support do you receive?”</td>
<td>The clarity of the question improved.</td>
</tr>
<tr>
<td>The section covering “home life/housing” questions</td>
<td>A statement alerting users that question 38 had been omitted from the online SInQUE, which read the following—“Not relevant for supported accommodation contexts -omitted.”—was confusing.</td>
<td>Changed statement to the following: “Question omitted, not included in the online SInQUE.”</td>
<td>The clarity of the question improved.</td>
</tr>
<tr>
<td><strong>Changes to the SInQUE summary outputs</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The SInQUE summary report</td>
<td>Some staff members found the spider graph to be confusing to interpret as the numbers summarizing scores in each section were not standardized and, therefore, it was difficult to tell which domains scored lower than others.</td>
<td>Simplified the spider graph to show percentage of total score in the graph instead of frequency.</td>
<td>It was easier to interpret the graph as sections with different totals became standardized.</td>
</tr>
<tr>
<td>The SInQUE summary report for multiple assessments completed with the same service user</td>
<td>In the section summarizing scores across multiple time points, staff members thought that a visual depiction of this comparison would be useful.</td>
<td>Made comparative bar charts of multiple scores across time with the same service user available on the summary report.</td>
<td>It was easier to understand and relay the results.</td>
</tr>
</tbody>
</table>

aChanges were made following the field-testing stage.
Overall, the changes made to the online SInQUE were relatively few and minor. Following initial laboratory testing, additional information and guidance for users was added, and minor revisions to the wording of questionnaire items were made to improve clarity. Modifications to the visual representation of scores in the summary reports were also made to aid ease of interpretation. During the field-testing stage, very few suggested changes to improve the usability of the online SInQUE were made by staff or service user participants. Further changes made at this stage included a minor wording adjustment to one question to ensure its cross-cultural appropriateness. Changes were made exclusively to the web-based version of the SInQUE and did not affect the existing SInQUE measure. We also developed an additional guidance document for managers and commissioners and an informational leaflet and poster about the SInQUE.

A few participants suggested substantial modifications to the structure and wording of individual items in the SInQUE that were not implemented by the study team. These decisions were made to preserve the broad scope and logical flow of the tool. We also declined to action some suggestions that were outside the remit of the SInQUE tool, such as adding more or free-text response options to some questions. However, where these suggestions indicated important potential barriers to using the SInQUE, they were noted and integrated into the qualitative analysis and logic model development. A summary of all comments and suggestions that were proposed but not implemented after review by the team can be found in Multimedia Appendix 4.

**Interview Thematic Analysis**

**Overview**

The interviews were analyzed using thematic analysis. Primary themes were deductively imposed according to each core element of the ICAMO model: intervention, context, actors, mechanisms, and outcomes [38]. Subthemes were inductively analyzed within each of these primary themes. The resultant thematic framework considering the perceived utility and acceptability of the online SInQUE and addressing aim 2 of the study is presented in Textbox 1. The themes are summarized in the following sections with a selection of illustrative quotes. Multimedia Appendix 5 provides further illustrative quotes for each subtheme.

**Intervention:** combination of program elements or strategies designed to produce behavior changes or improve health status among individuals

- The online Social Inclusion Questionnaire User Experience (SInQUE):
  - Promotion of positive, collaborative discussion
  - Comprehensive and novel questions
  - Ability to repeat over time
  - User-friendly design:
    - Easy-to-navigate website
    - Quick to complete
    - Fixed-response questions
    - Offers options to choose from
    - Web-based format

**Context:** salient conditions that are likely to enable or constrain the activation of program mechanisms

- Relevance of the SInQUE to staff role
- Inconsistency in current assessments used across services
- Absence of comparably specific assessments
- Emergence from the pandemic

**Actors:** the individuals, groups, and institutions who play a role in the implementation and outcomes of an intervention

- Staff:
  - Professional knowledge and skills
  - Professional boundaries
- Staff (service user views about staff):
  - Trusting relationship
  - Proactivity in offering guidance and support
- Service users:
  - Familiarity and comfort with the questions
  - Individual language and cultural differences of service users
- Service users (staff views about service users):
  - Engagement in the assessment
  - Existing mental health needs

**Mechanisms:** any underlying determinants or social behaviors generated in certain contexts

- Using the online SInQUE can accomplish the following:
  - Boost service user proactivity and confidence
  - Identify service users’ priorities on social inclusion
  - Prompt novel, personal conversations
  - Monitor changes in social inclusion over time
  - Identify gaps in support available within the organization and local community

**Outcomes:** behavior changes that follow the immediate knowledge change (intermediate) and changes such as patients’ health status and impact on community and health system (long term)

- Intermediate:
• Improve staff relationship with and understanding of the service user
• Help plan more relevant, targeted support for the service user
• Long term:
  • Borough-level improvements and changes in services to support social inclusion
  • Individual-level benefits for service users’ recovery and social inclusion

**Intervention**

In general, staff and service user participants felt that the tool was user-friendly and collaborative. Many noted the ability to repeat the assessment and the web-based format as being particularly useful and felt that the website was easy to navigate. The short length of the assessment was also discussed as an important advantage, with both staff and service users commenting that it felt quick to fill out. Participants noted that, despite the short assessment length, it still offered a range of interesting, positive, and sometimes unfamiliar questions that felt comprehensive and useful to discuss:

> I think that it wasn’t too just baseline, it was a little bit more than that and I think that’s good. Because it gives the option of, “Okay, you don’t want something, how can we improve and what is it that you do want that could help you while you’re in our service?” [212; staff member]

Interviewees felt that the user-friendliness was aided by accessible questions that were straightforward for service users to answer and that were cross-culturally appropriate for individuals from different ethnic backgrounds. Although some participants from both groups felt that the fixed-response options for the questions were limited, certain staff members thought that this made the questionnaire more accessible to service users who may otherwise struggle with engagement.

**Context**

Staff members largely felt that the tool was fitting and relevant to their role in helping support service users, and most did not already use assessments that were highly similar to the SInQUE. Certain staff members highlighted a lack of continuity of support workers within their service and noted that this often made it difficult to build rapport with service users. Some also commented on an inconsistency in assessments used across different services (in the local context, where 6 different provider organizations provided supported accommodation services across the borough). They noted that individual providers currently make their own recommendations on the tools that staff should use:

> If it was a standard central assessment that we do in all supported housing, that’s similar, like this for example, it might be beneficial in the long run. But each company has their own policy around it. [205; staff member]

One staff member noted that the tool felt particularly relevant following the COVID-19 pandemic as a means to promote engagement among service users after a period of likely sustained social isolation.

**Actors**

There were 2 key actors to consider in the application of the assessment: the staff members who asked the questions and the service users who responded to them.

One staff member felt that the assessment was not particularly relevant in the context of their work in a residential care service, where they had an established relationship with service users and already knew much of the queried information about them. However, this was an outlying view. Although most staff members thought that the online tool could be suitable and useful for their work, they emphasized the importance of using their professional knowledge and skills to pick when and for whom the assessment would be appropriate. They suggested that service users acutely struggling with their mental health may find it difficult to maintain concentration and engagement with the questionnaire and others may feel that the assessment is not relevant to them.

Staff members also raised the importance of maintaining professional boundaries with service users, and some expressed a concern that certain questions may feel invasive or uncomfortable for service users to answer:

> I think there was one quite private like about if they’re in a relationship or something, and that was the only question that made me feel a bit like I’m asking something very personal about a relationship. Because they might not want to say that. [210; staff member]

However, service user participants did not express any similar concerns about intrusive questions. They generally indicated that they felt comfortable with the assessment and that they were used to answering personal questions. Both staff and service users highlighted trust between those performing the assessment as a key factor in promoting engagement with such questions.

Both respondent groups highlighted the cultural diversity of service users within supported accommodation, and many noted that the tool felt appropriate for those from a variety of religious and ethnic backgrounds. Some service users commented on the importance of the staff members being proactive and taking the time to go through the assessment in detail with them, particularly individuals for whom English was a second language, as further explanation was required for some questions. Various staff members suggested that being provided with information about the assessment and its purpose specifically aimed at service users, for instance, a guidance...
leaflet, would be helpful for them to convey the essential information about the assessment.

Mechanisms
Both participant groups discussed how the tool may boost confidence and proactivity for a wide range of service users by highlighting specific, achievable ways in which an individual can improve their social inclusion. They also noted how the assessment encourages service users to open up and enables more profound conversations between them and staff members:

*Yes I found it really interesting, so like because it’s not really topics I would actually talk about. So it gave me a bit of enthusiasm to talk about some of the questions.* [409; service user]

Both groups suggested that it might be particularly useful during key working sessions as a means to get to know an individual better and identify their support preferences soon after moving into supported accommodation. Participants also noted the value of repeating assessments over time, suggesting that this could be a potentially encouraging way to demonstrate service user progress and identify gaps in available support. The most frequently suggested time between assessments was 1 to 3 months, with up to 6 months mentioned as a potential maximum gap.

Outcomes
Interviewees discussed the short- and long-term outcomes that they felt the tool could offer. They discussed how the tool enabled targeted and relevant support that prioritized the service user’s interests. Both groups also mentioned the potential for the tool to improve the relationship and understanding between service users and staff members:

*It asks questions where maybe like for your support worker to get a better understanding of you, like even though the immediate thing is highlight areas you can work on, it gives a general overview of how you are.* [410; service user]

Some staff members also discussed how prolonged use of the tool could highlight the additional borough-level support that may be needed to improve certain gaps in support and could also promote service user recovery toward the goal of more independent accommodation.

Implementation Strategy
Our implementation strategy was developed to encourage the use of the SInQUE in the supported accommodation services in the borough and was updated through consultation with clinical staff working in the borough’s community rehabilitation team and supported accommodation service managers. Our strategy was further informed by feedback from staff and service user participants during both stages of the study.

Each part of the strategy was developed to target an identified potential barrier to staff using the online SInQUE with service users. Subsequently, we mapped each component of the strategy to the 3 broad domains of the Capability, Opportunity, and Motivation–Behavior framework of behavior change [39] to describe whether each element of the strategy was intended to increase the staff’s capability, opportunity, or motivation to use the online SInQUE. The complete implementation strategy and the Capability, Opportunity, and Motivation–Behavior domain that each component addressed are outlined in Table 3. Strategies were related to enlisting leadership support to encourage supported accommodation staff to use the SInQUE, providing technical guidance and assistance with using the online tool, and developing bespoke summary output reports to reinforce use and increase the organizational benefits of using the SInQUE.
Table 3. Summary of the implementation strategy to support the use of the online Social Inclusion Questionnaire User Experience (SInQUE).

<table>
<thead>
<tr>
<th>Activity</th>
<th>Implementation goal being addressed</th>
<th>COM-B(^a) domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study research assistant (SE) and clinical research staff member (MD) visit all supported accommodation services to introduce the SInQUE tool to staff and offer guidance on its use.</td>
<td>Increase awareness of the SInQUE tool among supported accommodation staff and respond to any of their concerns or other problems.</td>
<td>Motivation, Opportunity, Capability</td>
</tr>
<tr>
<td>NHS(^b) community rehabilitation team manager and clinical researcher (MD) review supported accommodation caseloads to identify suitable service users for a SInQUE assessment and ask their key worker to complete an assessment with those service users.</td>
<td>Lack of accountability after initially asking all staff members to use the SInQUE and staff hesitancy over which service users would be suitable for an assessment</td>
<td>Motivation</td>
</tr>
<tr>
<td>NHS community rehabilitation team manager contacts supported accommodation managers to ask them to support and encourage the use of the SInQUE by identified key workers within a given time frame.</td>
<td>Lack of supported accommodation management prioritization for staff to use the SInQUE</td>
<td>Motivation, Opportunity</td>
</tr>
<tr>
<td>Local authority commissioners contact all supported accommodation managers to encourage the use of the SInQUE within services in the borough.</td>
<td>Lack of supported accommodation management prioritization for staff to use the SInQUE</td>
<td>Motivation</td>
</tr>
<tr>
<td>Study lead (BLE) and research assistant (SE) attend local Housing Forum meetings to update on the study and encourage the use of the SInQUE among all managers and staff members present.</td>
<td>Increase visibility and awareness of the SInQUE among supported accommodation managers and respond to any of their concerns or other problems.</td>
<td>Motivation, Opportunity</td>
</tr>
<tr>
<td>Study research assistant (SE) offers supported accommodation staff technological support with SInQUE registration and use.</td>
<td>Uncertainty about how to manage the technical process of using the SInQUE</td>
<td>Capability</td>
</tr>
<tr>
<td>Study team develops and circulates a leaflet about the SInQUE for supported accommodation staff to give to service users to help explain the purpose of an assessment.</td>
<td>Uncertainty among some supported accommodation staff members about how best to explain the purpose of the SInQUE and engage service users</td>
<td>Capability</td>
</tr>
<tr>
<td>Study team sends summary reports to service managers outlining use of the SInQUE and highlighting the areas of social inclusion that are most frequently prioritized and addressed in their service(^c).</td>
<td>Increased awareness among supported accommodation managers of the value offered by the SInQUE for service planning to encourage them to prompt staff to use it.</td>
<td>Motivation</td>
</tr>
<tr>
<td>Study team sends summary reports to local authority commissioners outlining which services have used the SInQUE the most and highlighting the areas of social inclusion that are most frequently prioritized and addressed across all services in the borough(^c).</td>
<td>Increased awareness of commissioners of the value offered by the SInQUE for service planning and commissioning to encourage them to prompt services to use it.</td>
<td>Motivation</td>
</tr>
</tbody>
</table>

\(^{a}\)COM-B: Capability, Opportunity, and Motivation–Behavior.  
\(^{b}\)NHS: National Health Service.  
\(^{c}\)These actions were planned with service managers and commissioners but not carried out during the 5-month implementation period because of the small number of completed SInQUE assessments.

Usage Data

In total, 27 staff members in the inner London borough registered for an account with the online SInQUE. Of the 27 staff members who registered, 17 (63%) from 6 different supported accommodation providers started or completed a SInQUE assessment with at least 1 service user. This resulted in 30 completed SInQUE assessments with 28 service users in the borough. This represented just >10% of the estimated total number of service users living in supported accommodation in the borough. Of the 28 service users, 4 (14%) were from residential care, 19 (68%) were from 24-hour supported housing, 3 (11%) were from 9-to-5 supported housing, 1 (4%) was from floating outreach services, and 1 (4%) was registered as “other” accommodation type. One staff member from 1 of the local authority areas where there was no specific implementation strategy registered for an account with the SInQUE; however, they did not start or complete a SInQUE assessment. This addressed aim 3 of the study.

Intervention Logic Model

On the basis of the collective study findings, we developed a logic model to summarize the processes involved in using the SInQUE and address study aim 4 (Figure 1).
The logic model was informed by the structure of an “ICAMO map” [38]. The model outlines the key aspects of the intervention (I), including its user-friendliness, its comprehensive nature, and the fact that it was based on a validated measure of social inclusion. It also indicates the potential outcomes (O) from using the SInQUE at both the individual and system levels, including improved support planning, better relationships, and provision of additional support for staff and services, which in turn may improve social inclusion and mental health outcomes as well as care provision more broadly. These operate within the broader societal context (C) of service users often being socially excluded and there being a high turnover of staff within these services and a high degree of variation across services in assessment tools that are recommended and in use, rendering the tool useful and relevant to the staff’s role.

The key actors (A) in implementing the SInQUE are the staff and service users, who require the skills and proactivity to administer the assessment and the motivation and trust to engage with the questions, respectively. Staff may encounter barriers such as a concern that some questions are too intrusive, and service users struggling more severely with their mental health may lack the concentration or motivation to engage with the questions. The potential outcomes operate through certain mechanisms (M), which include increased service user confidence and the prompting of more in-depth personal conversations between service users and staff. The tool also identifies more relevant priorities for service users, which may or may not be chosen as an active priority for support by staff owing to individual or organizational factors. Persistent and wide-ranging use of the tool could, over time, highlight the aspects of social inclusion that are feasible to work on and those that are regularly not being prioritized.

Discussion

Principal Findings

The online SInQUE was generally perceived as acceptable and potentially useful by supported accommodation staff and service users. This is consistent with findings of previous studies that used the SInQUE with other mental health populations [32,34]. Both staff and service users generally found the tool to be user-friendly and relevant and suggested that it could promote more targeted care planning and improve the relationship between staff and service users. Owing to the lower uptake of the SInQUE in residential care and floating outreach services, findings related to the tool’s utility in these settings are less conclusive than those for supported housing, where uptake was highest.

Some staff members expressed a concern that certain questions in the SInQUE could be perceived as intrusive by service users, indicating that they did not feel wholly comfortable asking what they perceived to be highly personal questions. However, this sentiment was not echoed by service users, who generally felt that the questions were appropriate and felt comfortable answering them. This finding is interesting given that supported
accommodation service users have highlighted in previous research the importance of feeling personally understood by staff in their service and have endorsed a process of familiarization with staff [25].

We found that implementation support is essential to promote the use of the tool in services, as evidenced by the lack of use of the tool in the 2 regions where the SInQUE was introduced without a concerted implementation strategy. The most effective steps in our implementation strategy were those during which the use of the tool was actively endorsed by individuals in leadership positions, particularly service managers and local service leaders. However, even with our concerted implementation strategy, uptake of the SInQUE was only achieved with approximately 10% of service users living in mental health supported accommodation in the participating borough within the 5-month study period.

Limitations

We used an established, iterative process of testing and feedback to develop the online SInQUE and determine its real-world acceptability and utility for use in mental health supported accommodation. However, it is important to acknowledge certain limitations of this study.

As mentioned previously, uptake of the SInQUE tool was highest in supported housing compared with residential care homes and floating outreach support. It is unclear whether this discrepancy reflects a greater reluctance from staff or service users in residential care and floating outreach support to use the online SInQUE. As proposed by one residential care staff member, it is possible that the staff in these services perceived the tool as being less relevant to their role. The discrepancy in part reflects the greater number of 24-hour and 9-to-5 supported housing units in the borough compared with residential care and floating outreach services, with approximately 6 times as many service users living in supported housing compared with residential care and nearly twice as many living in supported housing compared with floating outreach support. Through the local health service community mental health rehabilitation team, we also had a more direct connection with supported housing teams compared with other service types, which may have further contributed to the imbalance in services in which the SInQUE was used.

There were no female service user participants in the qualitative analysis; therefore, the findings may not be applicable to women in supported accommodation. It is unclear why it proved more difficult to recruit female participants, although it may reflect the higher proportion of male service users availing of supported accommodation in England—one review suggests that between 68% and 74% of service users are male across all supported accommodation types [40]. Furthermore, as we only tested the SInQUE in 1 London borough, the findings may not be generalizable to other regions.

As the tool was only used with approximately 10% of service users in the borough, the success of our implementation strategy was limited, and the low uptake may limit the wider generalizability of our findings. Owing to the short period and limited scope of the study, it was also not possible to assess whether use of the SInQUE in practice led to improved outcomes for service users or how useful the repeat assessments were over time. As the staff who participated volunteered to do so and they chose which service users to complete the SInQUE with, the findings may have been affected by selection bias and may not accurately reflect all supported accommodation staff members’ and service users’ views.

Finally, we removed 1 question from the original SInQUE questionnaire for our online version as asking people whether they lived alone was considered redundant for people living in supported housing. We also made very minor changes to the wording of 2 other questions in response to users’ feedback (Table 2). We think it is unlikely that these modest changes substantially affected the SInQUE’s psychometric properties. However, revalidation of the SInQUE in its web-based form is desirable in the future to confirm its validity and determine whether the minor wording changes should be retained for all versions of the SInQUE.

Implications for Practice

The SInQUE can be recommended as a potentially useful and acceptable tool for use in mental health supported accommodation settings, particularly supported housing services that offer 24-hour or 9-to-5 support, to provide a thorough assessment of social inclusion and support care planning. The tool may help meet an identified wish from service users for more discussion and support with social inclusion and relationships [41]. It was evident during the study that there is currently no universal tool in widespread use to help with social inclusion in mental health supported accommodation, highlighting the potential gap for an assessment tool such as the SInQUE. If used widely across supported accommodation services, the online SInQUE has the potential to provide benchmarking data and identify service users’ most common priorities for greater social inclusion to inform service planning and evaluation.

Our findings also suggest that, for an assessment tool such as the SInQUE to be widely used, it is essential to have active leadership endorsement and support. For example, it may be required for managers or commissioners to direct staff to use the SInQUE with service users who are willing and reinforce this through team meetings, setting of use targets, or implementation of key performance indicators for its use.

Implications for Research

It is important to hear from staff and service users who chose not to use the online SInQUE to understand their reasons for not using the tool and highlight barriers to using the tool that we may have missed in this study. It would be useful to conduct further testing of the tool in residential care and floating outreach supported accommodation settings to better determine the utility of the SInQUE in these service types. It would also be useful to examine the utility of the SInQUE in other population groups within different service types to determine whether the tool may be useful in additional settings.

Future research is necessary to establish the level of uptake of the SInQUE that can be achieved in supported accommodation over a longer period and potentially establish more effective
means of implementation support. A longer-term study is also needed to establish whether the possible benefits from using the SInQUE that were mentioned by staff and service users are achievable through the use of the tool and how any potential outcomes may vary over time. A hybrid implementation-evaluation study would address these queries to determine the effectiveness of the SInQUE tool as an intervention for social inclusion and establish a precise implementation strategy for widespread uptake of the tool in supported accommodation. Further research using the SInQUE is also warranted to examine service user needs related to social inclusion and identify any additional barriers to addressing these needs in supported accommodation services. Such research could be used to inform the development of a future complex intervention to support social inclusion in supported accommodation services.

Although this study chose to examine the utility of the online SInQUE specifically in supported accommodation, the tool may also be useful in other mental health populations. Previous studies have established that the SInQUE can be used with mental health service users with a wide range of diagnoses [13,32,34]. Therefore, it is reasonable to extrapolate that the online SInQUE may be useful to assess social inclusion and inform support and care planning for other mental health service users, not just those living in mental health supported accommodation.

Acknowledgments

This paper presents independent research funded by the National Institute for Health and Care Research School for Social Care Research. The views expressed are those of the authors and not necessarily those of the National Institute for Health and Care Research or the Department of Health and Social Care.

The authors would like to thank the supported accommodation managers, staff, and service users who supported and participated in this study. They would also like to thank the members of their expert advisory group and the University College London Service User Research Forum for their advice and feedback throughout the study.

Data Availability

The qualitative data generated and analyzed during this study are not publicly available to preserve participant anonymity. The quantitative data generated during this study are available from the corresponding author (BLE) upon reasonable request.

Authors’ Contributions

SE led recruitment, data collection, data analysis, and drafting of the paper. BLE led the study design and project management and supported the data analysis and drafting of the paper. HK, JC, PM, and GM co-designed the study and supervised the project. MD contributed to recruitment and data collection. GT contributed to the development of the study materials. All authors read, critically revised, and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Full description of the online Social Inclusion Questionnaire User Experience using the TIDieR (Template for Intervention Description and Replication) checklist for reporting interventions.

[DOCX File, 24 KB - ijmr_v13i1e45987_app1.docx]

Multimedia Appendix 2

Laboratory testing topic guide.

[DOCX File, 17 KB - ijmr_v13i1e45987_app2.docx]

Multimedia Appendix 3

Field-testing topic guides (for staff and service users).

[DOCX File, 16 KB - ijmr_v13i1e45987_app3.docx]

Multimedia Appendix 4

Explanation of suggested changes during laboratory testing and field-testing that were not made after team review.

[DOCX File, 18 KB - ijmr_v13i1e45987_app4.docx]

Multimedia Appendix 5

Interviews with staff and service users—illustrative quotes for each inductively derived subtheme.


35. SInQUE home page. SInQUE. URL: https://sinque.org.uk [accessed 2022-12-07]


Abbreviations

ICAMO: intervention-context-actor-mechanism-outcome
NHS: National Health Service
REC: Research Ethics Committee
SInQUE: Social Inclusion Questionnaire User Experience
TIDieR: Template for Intervention Description and Replication
UCL: University College London
A Web-Based Tool to Assess Social Inclusion and Support Care Planning in Mental Health Supported Accommodation: Development and Preliminary Test Study
URL: https://www.i-jmr.org/2024/1/e45987
doi:10.2196/45987
PMID:38477978

©Sharon Eager, Helen Killaspy, Joanna C, Gillian Mezey, Peter McPherson, Megan Downey, Georgina Thompson, Brynmor Lloyd-Evans. Originally published in the Interactive Journal of Medical Research (https://www.i-jmr.org/), 13.03.2024. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Interactive Journal of Medical Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.i-jmr.org/, as well as this copyright and license information must be included.