Original Paper

An Assessment Framework for e-Mental Health Apps in Canada: Results of a Modified Delphi Process

Jennifer Zelmer[1*], MA, PhD; Krystle van Hoof[2*], MA; MaryAnn Notarianni[3*], BA, MSW; Trevor van Mierlo[4*], MScCH, MBA, GEMBA, MSc BMR, DBA; Megan Schellenberg[3*], MA; Cara Tannenbaum[5,6*], MD, MSc

[1]Azimuth Health Group & University of Victoria, Toronto, ON, Canada

[2]Canadian Institutes of Health Research, Institute of Gender and Health, Montreal, QC, Canada

[3]Mental Health Commission of Canada, Ottawa, ON, Canada

[4]Evolution Health, Toronto, ON, Canada

[5]Faculties of Medicine and Pharmacy, Université de Montréal, Montreal, QC, Canada

[6]Institut Universitaire de Gériatrie de Montréal, Centre de Recherche, Université de Montréal, Montréal, QC, Canada

[*]all authors contributed equally

Corresponding Author:
Krystle van Hoof, MA
Canadian Institutes of Health Research
Institute of Gender and Health
4565 chemin Queen Mary
Montreal, QC, H3W 1W5
Canada
Phone: 1 5143129019
Email: krystle.van.hoof@criugm.qc.ca

Abstract

Background: The number of e-mental health apps is increasing rapidly. Studies have shown that the use of some apps is beneficial, whereas others are ineffective or do not meet users' privacy expectations. Individuals and organizations that curate, recommend, host, use, or pay for apps have an interest in categorizing apps according to the consensus criteria of usability and effectiveness. Others have previously published recommendations for assessing health-related apps; however, the extent to which these recommendations can be generalized across different population groups (eg, culture, gender, and language) remains unclear. This study describes an attempt by Canadian stakeholders to develop an e-mental health assessment framework that responds to the unique needs of people living in Canada in an evidence-based manner.

Objective: The objective of our study was to achieve consensus from a broad group of Canadian stakeholders on guiding principles and criteria for a framework to assess e-mental health apps in Canada.

Methods: We developed an initial set of guiding principles and criteria from a rapid review and environmental scan of pre-existing app assessment frameworks. The initial list was refined through a two-round modified Delphi process. Participants (N=25) included app developers and users, health care providers, mental health advocates, people with lived experience of a mental health problem or mental illness, policy makers, and researchers. Consensus on each guideline or criterion was defined a priori as at least 70% agreement. The first round of voting was conducted electronically. Prior to Round 2 voting, in-person presentations from experts and a persona empathy mapping process were used to explore the perspectives of diverse stakeholders.

Results: Of all respondents, 68% (17/25) in Round 1 and 100% (13/13) in Round 2 agreed that a framework for evaluating health apps is needed to help Canadian consumers identify high-quality apps. Consensus was reached on 9 guiding principles: evidence based, gender responsive, culturally appropriate, user centered, risk based, internationally aligned, enabling innovation, transparent and fair, and based on ethical norms. In addition, 15 informative and evaluative criteria were defined to assess the effectiveness, functionality, clinical applicability, interoperability, usability, transparency regarding security and privacy, security or privacy standards, supported platforms, targeted users, developers' transparency, funding transparency, price, user desirability, user inclusion, and meaningful inclusion of a diverse range of communities.

Conclusions: Canadian mental health stakeholders reached the consensus on a framework of 9 guiding principles and 15 criteria important in assessing e-mental health apps. What differentiates the Canadian framework from other scales is explicit attention to user inclusion at all stages of the development, gender responsiveness, and cultural appropriateness. Furthermore, an empathy

mapping process markedly influenced the development of the framework. This framework may be used to inform future mental health policies and programs.

Introduction

The number and range of electronic health apps, including those targeting mental health, continues to expand [1,2]. Studies indicate that the use of some apps can be beneficial to mental health and many improve accessibility to mental health services [3]. According to a recent meta-analysis, mobile phone-based mental health apps can have positive effects on depressive symptoms, anxiety, and other mental health conditions [4]. Digital solutions, including mobile phone apps, can also help address some traditional barriers—cost, capacity, geography, and stigma—to mental health services [5]. However, research suggests that some apps are unsafe, ineffective, poorly documented, or do not meet users' privacy and security expectations [6-8].

In Canada, supporting mental health and resilience through appropriate mobile health solutions is an area of growing policy interest, particularly given the rising rates of common mental health conditions [9] and unmet needs for services [10]. For example, *Changing Directions, Changing Lives*, the Mental Health Strategy for Canada, recommends increasing the use of e-mental health to reach more Canadians in need of support [11]. Accordingly, the Mental Health Commission of Canada has explored the use of these solutions in Canada and has also explored the associated opportunities and barriers to their use [5]. Likewise, the *Healthy and Productive Work Signature Initiative* of the Canadian Institutes of Health Research, which aims to support evidence-based interventions that foster healthy, meaningful, and productive work for all workers, recognizes e-mental health as a potential direction for improving the wellness of workers across the mental health continuum [12,13].

Currently, it can be time-consuming and difficult for potential users to assess the quality, safety, and evidence base of available health apps [3,14,15]. The private sector, government, academics, consumer groups, and others are trialing a range of strategies to tap into the benefits of eHealth while also addressing its attendant challenges. Regulatory, accreditation, market influence, educational, informational, and financial interventions are among the approaches that have been explored. For example, the United States Food and Drug Administration has established and updated regulatory guidance for mobile medical apps [16]. The Ontario Telemedicine Network has launched the public-facing Practical Apps website, which presents reviews of various health care apps; primary care providers can use this app to support their patients to better understand and manage their disease [17]. The National Health Service in the United Kingdom established a curated Apps Library, with apps assessed against a defined set of criteria, which was later rolled back following research that showed privacy and security gaps in a large proportion of the included

apps [18]. However, this service has since been relaunched through a beta site [19]. Researchers in Canada and beyond are testing techniques to engage app users in the development process [20,21]. Likewise, Apple Inc. introduced specific requirements for medical apps that are made available through their App Store [22].

Several formal development strategies, rating scales, or assessment frameworks have been published to help raise standards on app quality, and efforts are underway to develop others [1,14,20,23-28]. At their core, most of these efforts depend on a structured assessment of apps against defined criteria. While there is some convergence on the technical criteria considered, there are also important differences between the approaches. The aims, scope, purpose, target audiences, and methods of assessment vary considerably. For instance, some initiatives consider factors such as the characteristics of the app developer or funder; their policies; and features of the app, its performance characteristics, and ongoing maintenance or updating requirements, while others do not. While these efforts have provided information and direction for this study, the authors were not able to find a single scale or framework that addressed the unique combination of cultural and political factors required for the Canadian context.

The well-documented variation in the app quality and safety requires a consistent and transparent assessment framework for apps applied at an organizational level or as a self-assessment for app developers. Results must be meaningful and trustworthy for potential app users with a wide variety of needs and perspectives. In Canada, improving and expanding ready access to—and use of—effective and appropriate e-mental health solutions, including mental health apps, holds promise as a key enabler for addressing mental health. The purpose of this study was to achieve consensus among a broad group of Canadian stakeholders on a comprehensive set of principles to guide the development of a framework for assessing e-mental health apps in Canada as well as future processes to implement such a framework. A secondary objective was to achieve consensus on a complementary set criteria to support and ground these guiding principles by providing informative and evaluative measures that could be applied as part of an assessment process. Methods and results are reported in alignment with the guidance for reporting results from Delphi processes developed by Boulkedid et al [29].

Methods

Study Design

We used a modified Delphi process in a three-stage process with two voting rounds to reach consensus on guiding principles

and criteria for a Canadian e-mental health app assessment framework (Figure 1).

Unlike the open-ended initial phase of a traditional Delphi process [30,31], a modified Delphi process provides a starting point for discussion. Experts are polled for their views on this starting point individually and anonymously through two or more rounds of voting. Results are provided to participants between rounds. The process concludes when predefined stopping points—usually a specified level of consensus—are reached. The mini-Delphi or estimate-talk-estimate approach adapts this technique for face-to-face meetings, allowing experts to interact between iterations of anonymous votes [32]. Figure 1 illustrates this approach as applied to this study. Ethics approval was not required for this study as it falls under an exemption for research conducted by faculty and staff as an outside professional activity (see exemption #8 on the University of Victoria website, the first author's institution) [33].

Participant Recruitment

We used a heterogeneous purposive recruitment strategy to seek diversity on the following three key variables: (1) the breadth of perspective with relevant expertise recognized by the Institute of Gender and Health at the Canadian Institutes of Health Research or the Mental Health Commission of Canada (app developers, app users, health care providers, mental health advocates, lived experience of mental health problems or illnesses either personally or as a family member, policy makers, researchers, and workplace or workforce expertise); (2) sex and gender; and (3) geographic distribution across Canada.

The project steering committee reviewed the distribution of potential participants against a structured template that showed distribution based on the abovementioned criteria. Relevant expertise was loosely defined as a characteristic of individuals
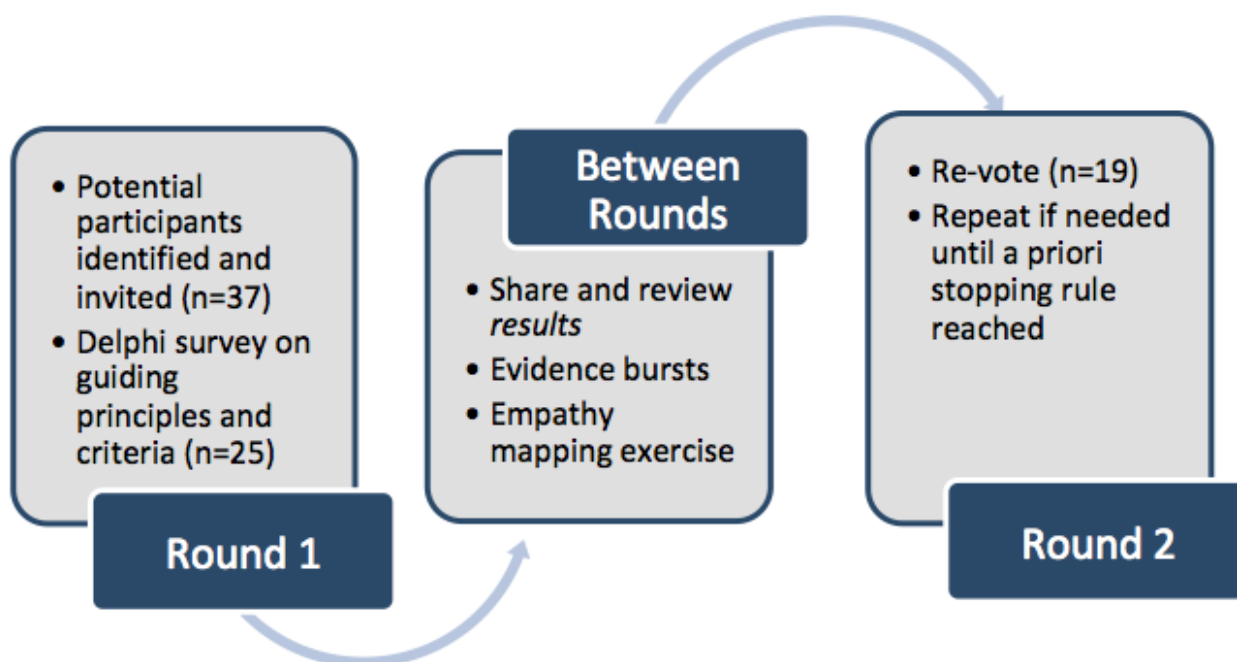
recognized by their peers as having competence and experience in specific areas relevant to this project. For instance, researchers were selected on the basis of having received scientific funding or research productivity in the area of e-mental health; individuals with lived experience were part of a volunteer group that advises the Mental Health Commission of Canada.

Then, we purposefully selected and recruited potential participants via invitations from the Institute of Gender and Health at the Canadian Institutes of Health Research and the Mental Health Commission of Canada to maximize the breadth and depth of experience and expertise available from the participant group [34]. Invitations were distributed via emails, with follow-up by telephone or in-person as required. Those who indicated an interest in taking part then received an email with a link to the website hosting the first round of the Delphi study. This email included the rationale for seeking consensus on an assessment framework for e-mental health apps. Voluntary participation implied consent.

Development of the Initial Set of Draft Principles and Criteria

First, we conducted a rapid review and environmental scan of pre-existing e-mental health app assessment frameworks found in the published and gray literature. Additional resources were identified and supplied by the modified Delphi participants. Alignment with Canadian culture and policies was considered. Gender responsiveness and cultural appropriateness were deemed to be two foundational and nonnegotiable elements, given Canada's strong commitment to gender equality and cultural inclusiveness, particularly of indigenous people [35-37]. Delphi process organizers and the authors of this study felt it was an ethical imperative to include an app's evidence base as a third foundational element of the framework in order to avoid the potential for harm from low-quality e-mental health apps.

Figure 1. The modified Delphi process.

Based on the findings, one of the authors (JZ) constructed a draft list of six other potential principles to guide the development and implementation of an e-mental health app assessment framework. Furthermore, 14 potential criteria were also identified, which aligned with the guiding principles and could be used as part of an app assessment process. The initial list was vetted by the steering committee (MN, MS, and KvH) for the project.

Round 1 of the Modified Delphi Process

In Round 1 of the modified Delphi process, which took place between November 15 and 28, 2016, 25 stakeholders agreed to participate. The survey questions used to assess the initial list of principles and criteria were modeled after an app evaluation Delphi process led by researchers at the Imperial College in England [38] and pre-existing formats extracted from the rapid review and environmental scan of eHealth and e-mental health app assessment frameworks.

The survey was prefaced with three contextual questions:

1. Confirmation of the perspectives that the participant brought to the consensus process, as per the list above (participants could self-identify with multiple perspectives).
2. How they would rate currently available health apps in terms of overall quality, usability or user engagement, use of high-quality evidence from credible sources, and information security or privacy. Respondents were asked to endorse the applicability of each of these criteria using a 4-point scale (poor, fair, good, and excellent) plus a "don't know" option.
3. Whether a framework for evaluating health apps is needed for consumers to identify apps of higher quality using a 5-point Likert scale (strongly disagree, disagree, neither agree nor disagree, agree, and strongly agree).

While gender and geographic diversity were considered in recruitment, we did not ask questions about these dimensions as part of the survey. This decision was taken to preserve the anonymity of responses, an important feature of a Delphi process. Inclusion of demographic variables in the survey would have made it possible to connect votes with the identity of at least some participants, which we sought to avoid. Likewise, we did not undertake a separate participant demographic survey to reduce the respondent burden.

The remaining survey questions assessed the initial list of 6 potential principles to guide the development of an e-mental health app assessment framework and the 14 potential criteria that might be included in such a framework. Criteria were classified as informative or evaluative. Informative criteria were defined as those where one answer is not "better" than another, but the information could be helpful to a user. Informative criteria would usually not be scored or have a minimum threshold for inclusion. Evaluative criteria were defined as those where directionality is clear (eg, whether more or less, presence or absence, is better). Evaluative criteria may have a minimum threshold for inclusion or be scored. We asked participants to rate the principles using a 5-point Likert scale (strongly disagree, disagree, neither agree nor disagree, agree, and strongly agree) and the criteria using a different 3-point scale (not necessary,

desirable, and essential). Participants also had the opportunity to provide comments on and suggest additions to the guiding principles and criteria.

Between Round Discussions, Evidence Presentations, and Empathy Mapping

A face-to-face meeting was held on November 28, 2016, at the Centre for Addiction and Mental Health (Toronto, Canada). Overall, 19 participants of the first modified Delphi process and all authors were in attendance. The event was facilitated by JZ, an experienced health services researcher with expertise in eHealth and a PhD in economics. She has previously led similar consensus-building exercises. Field notes were recorded by a research assistant.

At the beginning of the session, representatives from the Institute of Gender and Health of the Canadian Institutes of Health Research and the Mental Health Commission of Canada provided an overview of the policy context and rationale for this initiative. They specified that while a variety of approaches can be used for e-mental health app assessment, most depend on identifying a set of principles to guide the assessment process. The presenters reinforced that the scope of the exercise should be relevant to apps targeted at individuals and families seeking support to manage their own health, excluding apps targeted specifically at health professionals. Furthermore, clarification was provided that the framework could either be implemented by one or several organizations or as a self-assessment for app developers. Either way, the results of the assessment framework, once implemented, would be aimed at supporting the needs of members of the public as potential app users. App developers were considered a second important audience. Participants agreed that evidence base, gender responsiveness, and cultural appropriateness should be included as 3 foundational guiding principles of any Canadian assessment scale and need not be included in the consensus-building exercise.

Results from Round 1 of the modified Delphi process were presented using frequency distributions to summarize answers to survey questions with ordinal response scales. It was determined *a priori* that only principles and criteria from Round 1 voting with less than 80% agreement would be discussed during the meeting. Agreement was defined as endorsement of "strongly agree" or "agree" for the guiding principles and endorsement of "essential" or "desirable" for the criteria. Summaries of open-ended questions, including suggestions for changes or additions to principles or criteria, were also presented using quotes illustrative of the comments received. A facilitated iterative discussion led to modification, adaptation, removal, and addition to principles and criteria that did not initially secure 80% or greater agreement.

Furthermore, time was allocated for brief evidence presentations or "evidence bursts" from participant experts on opportunities for research related to eHealth apps, existing eHealth app assessment frameworks, and the digital health ecosystem. The purpose of the evidence bursts was to provide additional context and information to all participants on key considerations that might influence the selection of the guiding principles and criteria. Participants were allowed to ask questions after each

presentation and discuss the relevance to the app assessment framework.

We used a facilitated persona empathy mapping exercise to ensure that a broad range of end users' needs and perspectives were considered in the process [39,40]. Four personas were identified in advance, and participants added four more at the meeting (see Appendix 1). The eight personas represented a range of ages, sexes, genders, identities, cultures, geographic locations, and spectrum or severity of mental health conditions found across Canada. For each persona, meeting participants were encouraged to reflect and share the reasons why that individual might use an e-mental health app and what they imagined the use of such an app might achieve for each, according to individualized goals and needs. There was considerable discussion among meeting participants on these topics, including about how they might influence the principles and criteria. Consequently, several changes or additions were made to the text of the principles and criteria.

Round 2 of the Modified Delphi Process

Following the presentations, evidence bursts, and empathy mapping exercise, the participants at the face-to-face meeting re-rated the principles and criteria (in some cases with changes or additions based on the discussion) using the same scales as used in Round 1. Voting was anonymous and took place using electronic tools and on paper. Participants did not have access to each other's votes.

For Round 2 of the modified Delphi process, consensus was defined as 70% agreement, consistent with other studies of this type [41-43]. The intent was for items that did not achieve this consensus threshold to be included in a third postmeeting round of the Delphi process. This round, if required, would be conducted in the same manner as Round 1.

After Round 2 voting, participants shared key advice with policy makers on the directions ahead and provided feedback on the modified Delphi process. Following the session, a summary of the outcomes of the conversation was circulated to participants for review. We received no suggestions for amendments.

Results

Round 1 of the Modified Delphi Process

Participants in Round 1 of the modified Delphi self-identified as one or more of the following: app developers (5/24, 21%), app users (8/24, 33%), health care providers (8/24, 33%), mental health advocates (12/24, 50%), having lived experience of mental health problems as an individual living with such a problem or as a family member (8/24, 33%), policy makers (4/24, 17%), researchers (9/24, 38%), having workplace or workforce expertise (6/24, 25%), or having other relevant roles (6/24, 25%). There was one nonresponder. While participants could self-identify as having more than one role, many of them

were explicitly recruited because of their lived experience of mental health problems.

In response to the contextual questions posed in Round 1, almost half of the participants (12/25, 48%) rated the overall quality of currently available health apps as poor or fair, whereas 24% (6/25) gave a good or excellent rating. The remainder of participants (7/25, 28%) indicated that they were uncertain of the overall quality of currently available health apps.

Guiding Principles

Figure 2 illustrates the consensus ratings from the 6 guiding principles assessed during Round 1 voting. As 3 principles received "agree or strongly agree" ratings of at least 80% after Round 1, these were not discussed in detail prior to Round 2 voting; these included the principles that criteria should be user centered; processes should be open, transparent, and fair; and the research undertaken should reflect ethical norms. In addition, 2 principles representing the concepts that the apps enable innovation and be risk based received ratings of at least 70% but not 80%. Furthermore, the principle of international alignment received a rating of less than 70%.

Criteria

Round 1 voting included 14 potential criteria. Figure 3 lists the initial criteria assessed in Round 1 of the modified Delphi process and their endorsement by Delphi participants. All except one criterion met the essential or desirable inclusion consensus threshold. Only 2 of these—utility and transparency on information security or privacy policies—were endorsed as essential by at least 80% of participants. All participants voted on all criteria in this round, except one. There was one nonresponse for the "utility" criterion. The criterion of "awards that an app has received" was the only one that did not meet the consensus threshold for inclusion and was specifically discussed at the in-person meeting.

Between Round Discussion, Evidence Presentations, and Empathy Mapping

Discussion: Guiding Principles

As per *a priori* decision rules, between rounds at the in-person meeting, participants discussed the 3 guiding principles that did not receive "agree or strongly agree" ratings of at least 80% in Round 1. In addition to the comments received as part of Round 1 open-ended questions, suggestions for revisions to the wording of these principles and the framework emerged throughout the discussion. One suggestion that was endorsed by all was to change the terminology from "evaluation framework" to "assessment framework" to reflect the breadth of the agreed-upon guiding principles and criteria. In addition, participants suggested that the 3 foundational criteria (evidence base, gender responsiveness, and cultural appropriateness) should be explicitly stated and added to the list of guiding principles.

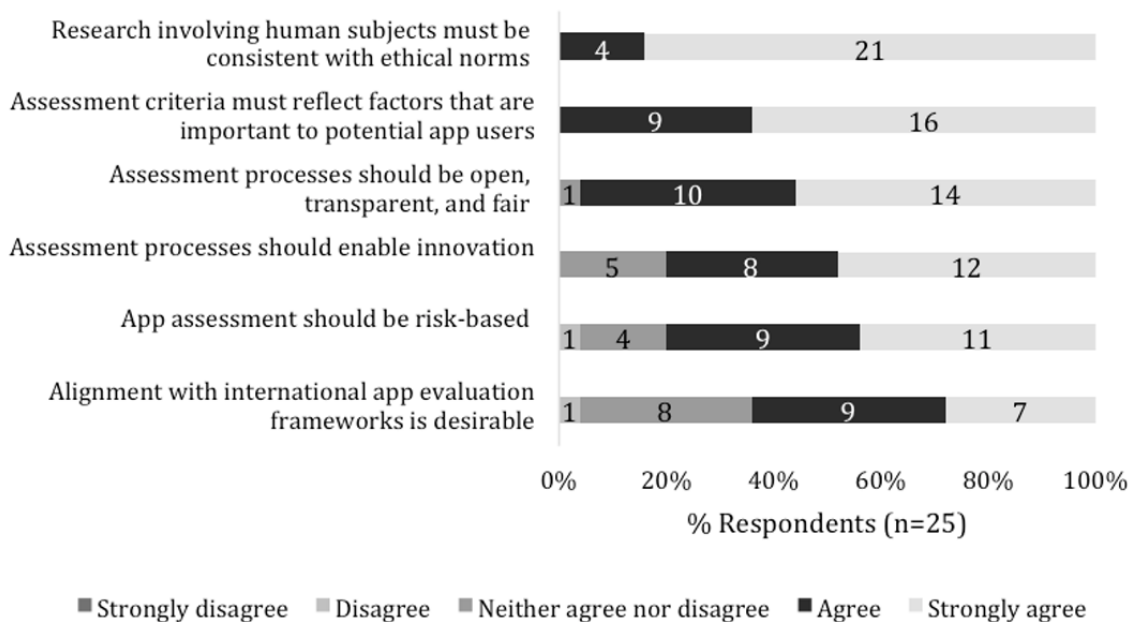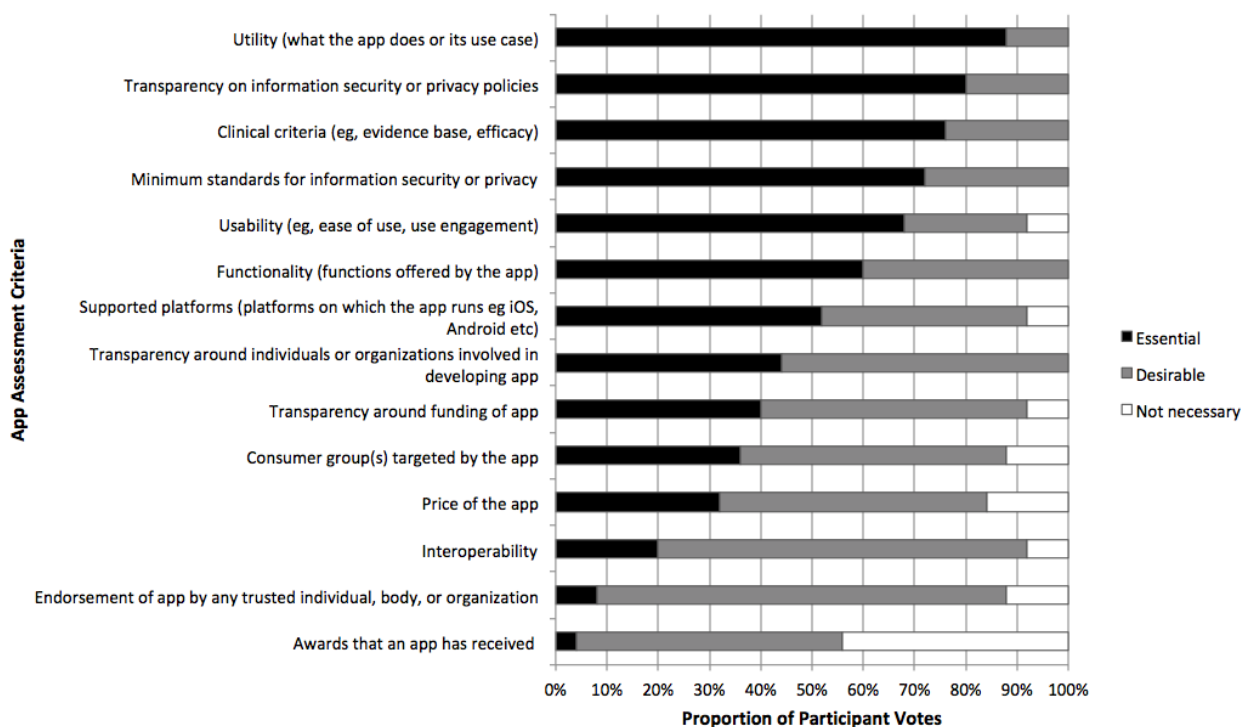**Figure 2.** Rating of the initial guiding principles in Round 1.

Figure 3. Round 1 rating of criteria for an e-mental health evaluation framework.

Discussion: Criteria

Three additional criteria emerged through discussions among participants about the important perspectives and expertise that different stakeholders bring. These 3 additional criteria were added prior to Round 2 voting based on the feedback from participants. These potential criteria included user desirability, ideally presented in a way that can be stratified by the type of user; user inclusion in the development of the apps; and meaningful inclusion of a diverse range of communities in the testing of the apps. Participants underscored that user views must be sought, considered, and reflected throughout the app development process to ensure app quality and relevance. There was general agreement among participants that, at a minimum, stakeholder engagement in the development, implementation, and assessment of e-mental health apps should be expected.

Participants also discussed the appropriateness of the labels used for the list of criteria. There was general agreement that

the "utility" criteria label should be replaced with the word "effectiveness" to clarify its meaning; this change was applied to the criteria list prior to Round 2 voting.

Round 2 of the Modified Delphi Process

Guiding Principles

Following discussion, participants were asked to re-rate the relevant principles, with updated wording applied. All guiding principles reached at least 70% consensus in Round 2 (77%-100% agreed or strongly agreed). As a result, a third round of voting was not required.

A total of 9 guiding principles for an e-mental health app assessment framework were retained (Textbox 1). These included the 3 initial foundational principles: evidence base, gender responsiveness, and cultural appropriateness; 6 principles from the Round 1 list (3 with modifications based on participant input); and 1 new principle introduced on the basis of participant input at the in-person meeting.

Criteria

We asked participants to re-rate the relevant criteria, with updated wording applied. Table 1 shows the distribution of votes endorsing criteria as "essential," "desirable," or "not

necessary" as well as the distribution of votes indicating how criteria should be approached—as either informative or evaluative. All criteria, except 2, reached at least 70% consensus, so a third round of voting was not required. The 2 criteria that were removed were "awards that an app has received" and "endorsement of the app by any trusted individual or body or organization" because participants expressed concern regarding the potential for bias in these criteria. The final criteria to be included in a Canadian e-mental health app assessment framework are listed in Textbox 2.

Reaching Consensus on the Need for a Canadian e-Mental Health App Assessment Framework

In Round 1 and at the end of Round 2, we asked participants about the need for an app assessment framework to help consumers identify e-mental health apps of higher quality. Participants' attitudes changed throughout the process. About two-thirds (17/25, 68%) of respondents in Round 1 agreed or strongly agreed with the need for the abovementioned framework. This number rose to 100% of participants after the in-person meeting, as assessed in Round 2 voting. For logistical reasons, some participants had to leave the in-person meeting prior to voting on this question, which resulted in n=13 for the Round 2 voting on this question.

Textbox 1. The final guiding principles for a Canadian e-mental health app assessment framework.

---

1. Evidence Based (foundational principle)

- Consideration must be given to the apps' evidence base and effectiveness

2. Gender Responsive (foundational principle)

- Apps must take into account sex and gender considerations

3. Culturally Appropriate (foundational principle)

- Appealing to and inclusive of Canada's diverse population

4. User Centered

- Assessment criteria must reflect the needs and expectations of potential app users

5. Risk Based

- App assessment should be risk based
- More detailed assessment is required for interventional apps, for example, drug dosing calculations or Web-based cognitive behavioral therapy, than for those focusing on general wellness support, for example, fitness trackers

6. Internationally Aligned

- App assessment framework should be informed by international experience or frameworks

7. Allows for Development and Continual Improvement

- The app assessment process should not impede the development or continual improvement of available apps

8. Open, Transparent, and Fair

- Assessment processes should be open, transparent, and fair

9. Human Research Consistent with Ethical Norms

- Research involving human subjects must be consistent with ethical norms
- Where an app is provided as part of a research study, Tri-Council Guidelines regarding Ethical Conduct for Research Involving Humans [37] must be followed.

Table 1. Round 2 criteria voting results.

Criteria	Distribution of endorsement			Approach to inclusion	
	Essential	Desirable	Not necessary	Informative	Evaluative
Effectiveness	18	0	0	4	19
Functionality	15	3	1	14	9
Clinical criteria	14	5	0	4	19
Interoperability	3	11	3	11	9
Usability	15	1	2	7	15
Information security transparency	18	1	0	4	19
Information security	16	2	0	4	17
Supported platforms	6	10	1	18	3
Audience	8	8	1	12	7
Developer transparency	9	6	0	18	5
Funding transparency	9	6	1	16	4
App price	6	10	2	19	2
User desirability	9	7	1	9	12
User inclusion	9	6	1	12	9
Meaningful inclusion	5	6	4	14	7
Endorsement of app[a]	3	8	7	6	4
Awards[b]	0	0	17	N/A[c]	N/A

[a]Did not meet the consensus threshold in Round 2.

[b]Did not meet the consensus threshold in Round 1; thus, omitted from informative versus evaluative criteria vote in Round 2.

[b]N/A: not applicable.

XSL•FO
RenderX

Textbox 2. The final criteria to be included in a Canadian e-mental health app assessment framework.

---

1. Effectiveness

- What does the app do? What is its use case?

2. Transparency of Information Security

- Are the app's security and privacy policies transparent and easy to find?

3. Information Security

- Does the app meet minimal standards for information security and privacy?

4. Functionality

- What are the functions offered by the app?

5. Usability

- Is the app easy to use? Do its intended users find it engaging?

6. Clinical Criteria

- What is the app's evidence base? Is there evidence of its efficacy?

7. Developer Transparency

- Is there readily available information regarding the individuals and organizations involved in the development of the app?

8. Funding Transparency

- Is there readily available information regarding who funded the development of the app?

9. User Inclusion

- Were potential users involved in the development of the app?

10. User Desirability

- Have the intended users expressed a desire for the functionality provided by the app?

11. Audience

- Is it clear who the intended consumer group(s) is and what issues the app aims to address?

12. Supported Platforms

- What platforms does the app run on? (eg, iOS, Android, etc)

13. App Price

- How much does it cost to use the app? If it is not free, is it a one-time cost, subscription-based, or other?

14. Meaningful Inclusion

- Is information available on to what extent and how potential end users were involved in the development of the app?

15. Interoperability

- To what extent do users have the ability to move across different platforms (mobile and desktop) while maintaining profile preferences and information?

Discussion

Principal Findings

The Delphi process described in this study demonstrates that consensus was reached on a Canadian set of guiding principles and criteria for assessing the quality, effectiveness, and usability of evidence-informed e-mental health apps that adhere to ethical standards. The purpose of this study is to render transparent the process through which this framework was developed. Specifically, two rounds of a modified Delphi process, including the use of presentations, evidence bursts, and empathy mapping between the two rounds, led to an assessment framework incorporating 9 guiding principles and 15 supporting criteria.

These principles and criteria make up a framework that reflects Canadian priorities associated with the need for evidence-based solutions, transparency, gender responsiveness, cultural appropriateness, and user engagement at all levels of e-mental health app development and testing. Furthermore, iterative and interactive discussions solidified the perception that a framework is essential for meeting the promise and potential of using e-mental health apps as part of a broader strategy to improve mental health in Canada.

While the framework described here was developed with particular attention to the Canadian context and national priorities, the principles and results of our consensus process are generally consistent with other efforts and findings in this area. For example, a 2016 systematic review of methods used to assess mobile health apps found that researchers evaluated the app quality in 6 domains: scientific or clinical basis, functionality, usability, accountability, impact, and popularity; 80% (73/91) participants used measures drawn from one or more of these domains [14]. Table 2 illustrates how we see our framework as mapping onto these domains.

Given the potential global reach of apps and the growing evidence that many apps available today do not conform to the established guidelines or best practices [1], the need for an assessment framework is increasingly salient. Potential users and health care professionals who intend to recommend e-mental health apps will want a simple way of knowing the quality and other characteristics of apps available to them. Assessment frameworks are an important first step in addressing this need, but many challenges exist in achieving a widespread implementation and scale-up.

For instance, the Canadian framework described in this study was designed so that it could be applied at an organizational level or as a self-assessment for app developers. In the former case, a minimal standard of critical appraisal and the resources to undertake the analysis would be needed to research and assess the evidence base for an app, establish the characteristics of an app's developers, and evaluate other criteria, such as cultural appropriateness. Indeed, even defining the process through which evidence base and clinical criteria will be assessed requires further investigation. In addition, an independent accreditation body could be tasked with this process to ensure quality and sustainability. While some Canadian organizations have assumed the aspects of this role (eg, via Canada Health Infoway's certification program for privacy, security, and interoperability of digital health solutions), no organization has undertaken the full scope of assessment involved.

The Mental Health Commission of Canada is working on a fact sheet to disseminate the results of this consensus process broadly to those who might be able to use the information to affect change in a way that mental health apps are developed, recommended, and taken up. Given that health is provincially and territorially regulated in Canada, there is a potential for provincial and territorial health authorities to implement the framework, or an adapted version of it, within their jurisdictions as mental health apps become a more mainstream option to address barriers like affordability and lack of access in remote communities. The 2017 Canadian federal budget included Can $11 billion over 10 years aimed at supporting provinces and territories in improving mental health and home care services [44]. This investment, as well as the spotlight it has put on mental health services, creates an opportune moment to promote this framework as a tool to make e-mental health an effective part of evolving provincial and territorial strategies.

Table 2. How the Canadian framework maps onto 6 domains commonly used by researchers to evaluate the app quality.

Six domains	Canadian framework guiding principles	Canadian framework criteria
Scientific or clinical basis	• Evidence based • Human research consistent with ethical norms	• Clinical criteria • Meaningful inclusion
Functionality	• Allows for development and continual improvement	• Functionality • Supported platforms • Interoperability
Usability	• Gender responsive • Culturally appropriate • User centered	• Usability • User inclusion
Accountability	• Risk based • Open, transparent, and fair	• Transparency of information security • Information security • Developer transparency • Funding transparency
Impact	• Internationally aligned	• Effectiveness • App price
Popularity	N/A[a]	• User desirability • Audience

[a]N/A: not applicable.

Another question that must be addressed is how the assessment results would be presented. For instance, a key decision is whether evaluative criteria should be used to calculate an overall summary score for each app or there should be minimum acceptable thresholds for certain criteria. The *Mobile App Rating Scale* (MARS), developed in Australia and one of the most developed examples of an app evaluation tool, weighs all criteria equally [27]. MARS ratings average the mean scores from the following 5 subscales: engagement, functionality, esthetics, information, and subjective quality. Given that *evidence base* is only 1 of 7 items listed under the information subscale, an app could potentially receive a very high rating on the MARS scale without having a strong evidence base. This method does not appear to be consistent with the findings from our Canadian Delphi process, which highlighted the significance that stakeholders attach to factors such as evidence base, gender responsiveness, and cultural appropriateness. Thus, further research could be helpful in addressing these and related questions regarding how best to obtain assessment results that accurately and appropriately reflect the guiding principles of the framework.

Research could also help inform a number of other practical implementation challenges. For example, the landscape of global mental health apps is evolving rapidly. Participants in the Delphi process emphasized that an assessment framework and process should not impede the development or continual improvement of apps. In this context, one needs to consider factors such as the timeline and criteria for re-assessing apps as they evolve. Likewise, if apps that have been assessed can be accessed via an integrated listing or shared repository, who would curate the collection and how would need to be determined, as would how to encourage its use by the public and health professionals.

Many other countries around the world are grappling with similar issues related to the endorsement of e-mental health apps. Some countries have published recommendations for assessing health-related—including mental health-related—apps, but the extent to which they can be generalized across a variety of populations and their characteristics (eg, culture, gender, and language) remains unclear. This is one of the reasons why we used a process focused on the needs of stakeholders across Canada. While the diversity of our participants is a strength of this study, we solicited their feedback within a defined Canadian scope and perspective. Furthermore, there was a 20% attrition of participants who voted in the first (Web-based) and second (in-person) round of the Delphi process. This may have introduced selection bias for some of the principles and criteria retained in the final framework.

While the methods used in this study can be replicated readily, the outcomes that we obtained by consulting stakeholders from across Canada may or may not be transferable to other countries or contexts. As several of the factors considered in our process were aligned with a related process simultaneously underway in the United Kingdom, it may be possible to explore this question in the future when their results are available.

Conclusions

Improving health gains and reducing risks from the use of e-mental health apps depends on tilting the balance toward solutions that are of higher quality and are more acceptable to potential users. To inform the efforts to achieve this goal, we used an innovative structured Delphi process that incorporated evidence bursts and empathy mapping to ask a diverse group of Canadian stakeholders about what factors were important in assessing e-mental health apps. The group reached consensus on 9 guiding principles and 15 criteria for an e-mental health app assessment framework that could be applied at an individual or organizational level to support them in meeting the needs and expectations of potential app users and other key stakeholders. This consensus has the potential to inform future research, policy, and programs at a still relatively early stage in the evolution of e-mental health solutions.

Conflicts of Interest

JZ received consulting fees from the project funders to plan and facilitate the modified Delphi process described herein, as well as to prepare a summary of the meeting outcomes. TvM is the CEO & Founder of Evolution Health Systems Inc, a company that owns and manages digital health interventions. He is also the acting CSO of VeggieCake LLC, a company developing mental health interventions targeting millennials. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Canadian Institutes of Health Research (CIHR) or the Government of Canada.

Multimedia Appendix 1

Personas for empathy mapping.

[PDF File (Adobe PDF File), 25 KB-Multimedia Appendix 1]

References

1. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental Health Smartphone Apps: Review and Evidence-Based Recommendations for Future Developments. JMIR Ment Health 2016 Mar 01;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]

2. Research2Guidance. mHealth App Developer Economics 2016: The current status and trends of the mHealth app market, 6th annual study on mHealth app publishing. 2016 Oct. URL: http://research2guidance.com/r2g/r2g-mHealth-App-Developer-Economics-2016.pdf [accessed 2018-02-01] [WebCite Cache ID 6wt2batXa]

3. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]

4. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]

5. Mental Health Commission of Canada. Ottawa: Mental Health Commission of Canada; 2014. E-mental health in Canada: transforming the mental health system using technology - A briefing document URL: http://www.mentalhealthcommission.ca/sites/default/files/MHCC_E-Mental_Health-Briefing_Document_ENG_0.pdf[WebCite Cache ID 6wt2lu1yr]

6. S. Bhuyan S, Kim H, Isehunwa O, Kumar N, Bhatt J, Wyant D, et al. Privacy and security issues in mobile health: Current research and future directions. Health Policy and Technology 2017 Jun;6(2):188-191. [doi: 10.1016/j.hlpt.2017.01.004]

7. Nicholas J, Larsen M, Proudfoot J, Christensen H. Mobile Apps for Bipolar Disorder: A Systematic Review of Features and Content Quality. J Med Internet Res 2015 Aug 17;17(8):e198 [FREE Full text] [doi: 10.2196/jmir.4581] [Medline: 26283290]

8. Larsen ME, Nicholas J, Christensen H. A Systematic Assessment of Smartphone Tools for Suicide Prevention. PLoS One 2016;11(4):e0152285 [FREE Full text] [doi: 10.1371/journal.pone.0152285] [Medline: 27073900]

9. Statistics Canada. Ottawa: Statistics Canada; 2016. Table 13-10-0802-01 Mental health characteristics: Ability to handle stress and sources of stress URL: http://www5.statcan.gc.ca/cansim/a26[WebCite Cache ID 6wt64rl37]

10. Sunderland A, Findlay LC. Perceived need for mental health care in Canada: Results from the 2012 Canadian Community Health Survey-Mental Health. Health Rep 2013 Sep;24(9):3-9 [FREE Full text] [Medline: 24258361]

11. Mental Health Commission of Canada. Calgary; 2012. Changing Directions, Changing Lives: The Mental Health Strategy for Canada URL: http://strategy.mentalhealthcommission.ca/pdf/strategy-images-en.pdf[WebCite Cache ID 6wt6B7oqt]

12. Tannenbaum C, Voss P, El-Gabalawy H, Joanette Y. Gender, Work, and Aging. Can J Aging 2016 Dec;35(3):405-411. [doi: 10.1017/S0714980816000416] [Medline: 27418079]

13. Canadian Institutes of Health Research. 2015 Dec 20. Healthy and Productive Work URL: http://www.cihr-irsc.gc.ca/e/47706.html [accessed 2018-01-31] [WebCite Cache ID 6wt6PQioG]

14. Grundy Q, Wang Z, Bero L. Challenges in Assessing Mobile Health App Quality: A Systematic Review of Prevalent and Innovative Methods. Am J Prev Med 2016 Dec;51(6):1051-1059. [doi: 10.1016/j.amepre.2016.07.009] [Medline: 27659122]

15. Shen N, Levitan M, Johnson A, Bender J, Hamilton-Page M, Jadad A, et al. Finding a Depression App: A Review and Content Analysis of the Depression App Marketplace. JMIR mHealth uHealth 2015 Feb 16;3(1):e16. [doi: 10.2196/mhealth.3713]

16. U.S. Food and Drug Administration. 2015 Sep 02. Mobile Medical Applications: Guidance for Industry and Food and Drug Administration Staff URL: https://www.fda.gov/downloads/MedicalDevices/.../UCM263366.pdf [accessed 2018-02-01] [WebCite Cache ID 6wt3j99Xl]

17. Ontario Telemedicine Network. The Practical Apps Approach. 2016. URL: https://practicalapps.ca/approach/ [accessed 2018-01-31] [WebCite Cache ID 6wt43XEWO]

18. Huckvale K, Prieto J, Tilney M, Benghozi P, Car J. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. BMC Med 2015 Sep 25;13(1):A. [doi: 10.1186/s12916-015-0444-y]

19. National Health Service, UK. Apps Library Beta URL: https://apps.beta.nhs.uk/ [accessed 2018-04-04] [WebCite Cache ID 6yQkOJlJ3]

20. Grindrod KL, Li M, Gates A. Evaluating user perceptions of mobile medication management applications with older adults: a usability study. JMIR Mhealth Uhealth 2014 Mar 14;2(1):e11 [FREE Full text] [doi: 10.2196/mhealth.3048] [Medline: 25099993]

21. Avis JV, van MT, Fournier R, Ball GD. Lessons Learned From Using Focus Groups to Refine Digital Interventions. JMIR Res Protoc 2015 Jul 31;4(3):e95 [FREE Full text] [doi: 10.2196/resprot.4404] [Medline: 26232313]

22. Apple Inc. Apple Developer. App Store Review Guidelines URL: https://developer.apple.com/app-store/review/guidelines/ [accessed 2018-01-31] [WebCite Cache ID 6wt4GcGJh]

XSL•FO
RenderX

23. Fiore P. How to Evaluate Mobile Health Applications: A Scoping Review - Studies in Health Technology Informatics. In: Building Capacity for Health Informatics in the Future. Amsterdam: IOS Press; 2017:109-114.

24. Chindalo P, Karim A, Brahmbhatt R, Saha N, Keshavjee K. Health Apps by Design: A Reference Architecture for Mobile Engagement. International Journal of Handheld Computing Research 2016;7(2):34-43. [doi: 10.4018/IJHCR.2016040103]

25. Albrecht U. Transparency of health-apps for trust and decision making. In: J Med Internet Res. Hershey, PA: IGI Global; Dec 30, 2013:e277-e108.

26. Hanrahan C, Dy Aungst T, Cole S. Evaluating Mobile Medical Applications. Bethesda, MD: American Society of Health-System Pharmacists, Inc; 2014.

27. Stoyanov S, Hides L, Kavanagh D, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]

28. Singh K, Drouin K, Newmark LP, Rozenblum R, Lee J, Landman A, et al. Developing a Framework for Evaluating the Patient Engagement, Quality, and Safety of Mobile Health Applications. Issue Brief (Commonw Fund) 2016 Feb;5:1-11. [Medline: 26934758]

29. Boulkedid R, Abdoul H, Loustau M, Sibony O, Alberti C. Using and Reporting the Delphi Method for Selecting Healthcare Quality Indicators: A Systematic Review. PLoS ONE 2011 Jun 9;6(6):e20476. [doi: 10.1371/journal.pone.0020476]

30. Dalkey N, Helmer O. An Experimental Application of the DELPHI Method to the Use of Experts. Management Science 1963 Apr;9(3):458-467. [doi: 10.1287/mnsc.9.3.458]

31. Fink A, Kosecoff J, Chassin M, Brook RH. Consensus methods: characteristics and guidelines for use. Am J Public Health 1984 Sep;74(9):979-983. [Medline: 6380323]

32. Gustafson D, Shukla R, Delbecq A, Walster G. A comparative study of differences in subjective likelihood estimates made by individuals, interacting groups, Delphi groups, and nominal groups. Organizational Behavior and Human Performance 1973 Apr;9(2):280-291. [doi: 10.1016/0030-5073(73)90052-4]

33. University of Victoria. Human Research Ethics. 2018. URL: https://www.uvic.ca/research/conduct/home/regapproval/humanethics/index.php [accessed 2018-02-10] [WebCite Cache ID 70W0yuUCn]

34. Mathison S. Purposeful sampling. In: Encyclopedia of Evaluation. Thousand Oaks, CA: Sage Publications, Inc; 2004.

35. Status of Women Canada. 2017 Oct 03. Government of Canada's Approach: Gender-Based Analysis Plus URL: http://www.swc-cfc.gc.ca/gba-acs/approach-approche-en.html [accessed 2018-02-01] [WebCite Cache ID 6wtBJzQvm]

36. Government of Canada. 2017 Jun 27. Health Portfolio Sex and Gender-Based Analysis Policy URL: https://www.canada.ca/en/health-canada/corporate/transparency/corporate-management-reporting/heath-portfolio-sex-gender-based-analysis-policy.html [accessed 2018-02-01] [WebCite Cache ID 6wtC0qH02]

37. Department of Justice. 2018 Feb 14. Principles respecting the Government of Canada's relationship with Indigenous peoples URL: http://www.justice.gc.ca/eng/csj-sjc/principles-principes.html [accessed 2018-02-01] [WebCite Cache ID 6wtCaNT49]

38. Mobasheri M. Personal communication with Jennifer Zelmer. E-mail 2016 Oct 25.

39. Ferreira B, Williamson S, Oliveira EAJ, Conte T. Designing Personas with Empathy Map. 2015 Presented at: SEKE; July 6-8, 2015; Pittsburgh p. 501-505.

40. Williams I, Brereton M, Donovan J, McDonald K, Millard T, Tam A, et al. A Collaborative Rapid Persona-Building Workshop: Creating Design Personas with Health Researchers. International Journal of Sociotechnology and Knowledge Development 2014 Apr;6(2):17-35. [doi: 10.4018/ijskd.2014040102]

41. Hasson F, Keeney S, McKenna H. Research Guidelines for the Delphi Survey Technique. Journal of Advanced Nursing 2008 Jun 28;32(4). [doi: 10.1046/j.1365-2648.2000.t01-1-01567.x] [Medline: 11095242]

42. Kleynen M, Braun S, Bleijlevens M, Lexis M, Rasquin S, Halfens J, et al. Using a Delphi technique to seek consensus regarding definitions, descriptions and classification of terms related to implicit and explicit forms of motor learning. PLoS One 2014;9(6):e100227 [FREE Full text] [doi: 10.1371/journal.pone.0100227] [Medline: 24968228]

43. Zafar SY, Currow DC, Cherny N, Strasser F, Fowler R, Abernethy AP. Consensus-based standards for best supportive care in clinical trials in advanced cancer. Lancet Oncol 2012 Feb;13(2):e77-e82. [doi: 10.1016/S1470-2045(11)70215-7] [Medline: 22300862]

44. Health Canada. 2018 May 14. Shared Health Priorities URL: https://www.canada.ca/en/health-canada/corporate/transparency/health-agreements/shared-health-priorities.html[WebCite Cache ID 6z4mS5tUq]

45. Greenspun H, Coughlin S. mHealth in an mWorld: How mobile technology is transforming health care. Washington, DC: Deloitte Center for Health Solutions; 2012. URL: https://www2.deloitte.com/content/dam/Deloitte/us/Documents/life-sciences-health-care/us-lhsc-mhealth-in-an-mworld-103014.pdf[WebCite Cache ID 6p25Cf7Da]

46. World HO. mHealth New horizons for health through mobile technologies: Based on the findings of the second global survey on eHealth - Global Observatory for. Geneva: World Health Organization; 2011.

47. Philbrick WC. mHealth and MNCH - State of the Evidence.: mHealth Alliance, UN Foundation; 2013. URL: http://www.mhealthknowledge.org/sites/default/files/15_un_007_evidencegapreport_digital_aaa.pdf [accessed 2017-03-17] [WebCite Cache ID 6p25nq3nf]

48. Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities Research Council of Canada. Tri-Council Policy Statement: Ethical Conduct for Research Involving

Humans. Ottawa: Government of Canada; 2014. URL: http://www.pre.ethics.gc.ca/pdf/eng/tcps2-2014/TCPS_2_FINAL_Web. pdf [accessed 2018-06-28] [WebCite Cache ID 70WB9cjG9]

49. Stoyanov S, Hides L, Kavanagh D, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]

50. Chan S, Torous J, Hinton L, Yellowlees P. Towards a Framework for Evaluating Mobile Mental Health Apps. Telemed J E Health 2015 Dec;21(12):1038-1041. [doi: 10.1089/tmj.2015.0002] [Medline: 26171663]

51. Cuhls K. Delphi Method in Delphi Surveys: Teaching Material for UNIDO Foresight Seminars. Internet. Wien: UNIDO; 2005. URL: http://www.unido.org/fileadmin/import/16959_DelphiMethod.pdf [accessed 2017-02-08] [WebCite Cache ID 6o8CYJlFB]

Abbreviations

CIHR: Canadian Institutes of Health Research
MARS: Mobile App Rating Scale

XSL•FO
RenderX