

# ***Draft Technical Brief***

---

**Number XX**

## **Evaluation of Mental Health Mobile Applications**

**Prepared for:**

Agency for Healthcare Research and Quality  
U.S. Department of Health and Human Services  
5600 Fishers Lane  
Rockville, MD 20857  
[www.ahrq.gov](http://www.ahrq.gov)

This information is distributed solely for the purposes of pre-dissemination peer review. It has not been formally disseminated by the Agency for Healthcare Research and Quality. The findings are subject to change based on the literature identified in the interim and peer-review/public comments and should not be referenced as definitive. It does not represent and should not be construed to represent an Agency for Healthcare Research and Quality or Department of Health and Human Services (AHRQ) determination or policy.

**Contract No.** To be included in the final report.

**Prepared by:**

To be included in the final report.

**Investigators:**

To be included in the final report.

**AHRQ Publication No. xx-EHCxxx**  
**<Month Year>**

This report is based on research conducted by the <EPC> Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. <#>). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

**None of the investigators have any affiliations or financial involvement that conflicts with the material presented in this report.**

The information in this report is intended to help healthcare decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of healthcare services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

This report is made available to the public under the terms of a licensing agreement between the author and the Agency for Healthcare Research and Quality. This report may be used and reprinted without permission except those copyrighted materials that are clearly noted in the report. Further reproduction of those copyrighted materials is prohibited without the express permission of copyright holders.

AHRQ or U.S. Department of Health and Human Services endorsement of any derivative products that may be developed from this report, such as clinical practice guidelines, other quality enhancement tools, or reimbursement or coverage policies may not be stated or implied.

Persons using assistive technology may not be able to fully access information in this report. For assistance contact [EPC@ahrq.hhs.gov](mailto:EPC@ahrq.hhs.gov).

**Suggested citation:** To be included in the final report.

## Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of healthcare in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions and new healthcare technologies and strategies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

This EPC evidence report is a Technical Brief. A Technical Brief is a rapid report, typically on an emerging medical technology, strategy, or intervention. It provides an overview of key issues related to the intervention—for example, current indications, relevant patient populations and subgroups of interest, outcomes measured, and contextual factors that may affect decisions regarding the intervention. Although Technical Briefs generally focus on interventions for which there are limited published data and too few completed protocol-driven studies to support definitive conclusions, the decision to request a Technical Brief is not solely based on the availability of clinical studies. The goals of the Technical Brief are to provide an early objective description of the state of the science, a potential framework for assessing the applications and implications of the intervention, a summary of ongoing research, and information on future research needs. In particular, through the Technical Brief, AHRQ hopes to gain insight on the appropriate conceptual framework and critical issues that will inform future research.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the healthcare system as a whole by providing important information to help improve healthcare quality.

If you have comments on this Technical Brief, they may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 5600 Fishers Lane, Rockville, MD 20857, or by email to [epc@ahrq.hhs.gov](mailto:epc@ahrq.hhs.gov).

David Meyers, M.D.  
Acting Director  
Agency for Healthcare Research and Quality

Arlene Bierman, M.D., M.S.  
Director  
Center for Evidence and Practice Improvement  
Agency for Healthcare Research and Quality

Craig A. Umscheid, M.D., M.S.  
Director  
Evidence-based Practice Center Program  
Center for Evidence and Practice Improvement  
Agency for Healthcare Research and Quality

David W. Niebuhr, M.D., M.P.H., M.Sc.  
Task Order Officer  
Center for Evidence and Practice Improvement  
Agency for Healthcare Research and Quality

## **Acknowledgments**

To be included in the final report.

## **Key Informants**

To be included in the final report.

## **Peer Reviewers**

To be included in the final report.

# Evaluation of Mental Health Applications

## Structured Abstract

**Background.** Mental health mobile applications (apps) have the potential to expand the provision of mental health and wellness services to traditionally underserved populations. There is a lack of guidance on how to choose wisely from the thousands of mental health apps without clear evidence of safety, efficacy, and consumer protections.

**Purpose.** This technical brief proposes a framework to assess mental health mobile applications with the aim to facilitate selection of apps. The results of the application of the framework will yield summary statements on the strengths and limitations of the apps, and are intended for use by providers, and patients/caregivers.

**Methods.** We reviewed systematic reviews of mental health apps; published and gray literature on mental health app frameworks; and we conducted four Key Informant discussions to identify gaps and key framework criteria. These reviews and discussions informed a draft framework to assess apps. Iterative testing and refinement of the framework was done in five successive rounds through double application of the framework to a total of 27 apps. Items in the framework with an interrater reliability under 90% were discussed among the evaluation team for revisions.

**Findings.** Our review of the existing frameworks identified key gaps in the assessment of risks that users may face from the apps, such as privacy and security disclosures, and regulatory safeguards to protect the users. Key Informant discussions identified priority criteria to include in the framework including efficacy and safety of mental health apps. The Framework to Assist Stakeholders in Technology Evaluation for Recovery (FASTER) to Mental Health and Wellness was developed and comprises three sections: Section 1. Risks and Mitigation Strategies: assesses the integrity and risk profile of the app; Section 2. Function: is focused on descriptive aspects related to accessibility, costs, developer credibility, evidence and clinical foundation, privacy/security, usability, functions for remote monitoring of the user, access to crisis services, and artificial intelligence; and Section 3. Mental Health App Features: focuses on specific mental health app features such as journaling, mood tracking, etc.

**Conclusion.** FASTER may be used to help appraise and select mental health apps. Future application, testing and refinements may be required to determine the framework's suitability and reliability across multiple mental health conditions, as well as to account for the rapidly expanding applications of AI, gamification, and other new approaches.

# Contents

Introduction.....	1
Background .....	1
Guiding Questions .....	6
Methods .....	7
Identification and Abstraction of Existing Frameworks .....	7
Identification of Gaps .....	8
Development of the Framework .....	9
Refinement of the Framework .....	11
Findings .....	15
Gaps in Existing Frameworks.....	15
Key Informants Interviews .....	16
Testing and Refinement of the Framework During Pre-pilot and Pilot Rounds .....	18
Overview of the Framework to Assist Stakeholders in Technology Evaluation for Recovery (FASTER) to Mental Health and Wellness.....	21
Summary of Changes Made to the Framework During Pilot Testing Rounds.....	27
Summary and Implications.....	30
Next Steps .....	30
References .....	32

## Tables

Table 1. Categorization of mental health disorders.....	12
Table 2. Summary of the number of changes made and inter-rater reliability in each round .....	20
Table 3. App integrity levels .....	23
Table 5. Risk levels .....	24

## Boxes

Box 1. Key Informants' expertise.....	16
Box 2. Key components of framework from the Key Informants .....	18

## Figures

Figure 1. Framework development process.....	7
Figure 2. Apps screening process .....	19
Figure 3. Framework to Assist Stakeholders in Technology Evaluation for Recovery (FASTER) to Mental Health and Wellness.....	21
Figure 4. Categories of Questions in Each Section .....	22

## Appendixes

Appendix A. List of Frameworks	
Appendix B. Framework Criteria	
Appendix C. Modifications in uMARS Questions	
Appendix D. Sample of 10 apps for the pre-pilot round	
Appendix E. Pilot Rounds Results	
Appendix F. Framework Training Guide	
Appendix G. Framework	

# Evidence Summary

## Main Points

- We developed and pilot tested a framework to assess mental health mobile applications (apps).
- The Framework to Assist Stakeholders in Technology Evaluation for Recovery (FASTER) to Mental Health and Wellness assesses the rigor, technical functionality, and mental health features of apps.
- FASTER can be used by advocacy organizations, payors, and others to inform selection of mental health apps.

## Background and Purpose

People suffering from behavioral health conditions may not seek or receive care owing to stigma, provider shortages, long wait times, proximity to mental health providers, or other accessibility issues. Mental health mobile applications (apps) may help to address this gap. With potential app privacy/security and safety concerns in a more vulnerable population of users, the decisional dilemma is “how can consumers, family members and peer supports, providers and health systems select mental health and wellness apps?” The aim of this technical brief was to develop a framework to assess mental health apps for users across different age groups, for different mental health symptoms and disorders, and for general mental wellness.

## Methods

We reviewed existing frameworks, published review articles on frameworks, and the gray literature. We also conducted key informant interviews to identify elements to inform the development of a framework to assess the safety and effectiveness of mental health apps. A second group of key informants provided feedback on the initial framework. We conducted five rounds of testing and revised the framework.

Our protocol is posted on the program’s website. Details of the methodology can be found in the full report.

## Results

We found that the existing frameworks have a heavy emphasis on technology, have limited information about app features specific to mental health, and do not account for potential risks including those posed by advances in machine learning and artificial intelligence.

The Framework to Assist Stakeholders in Technology Evaluation for Recovery (FASTER) to Mental Health and Wellness expands on existing frameworks to address the identified gaps, especially in the area of risk assessment and use of mental health specific features. It comprises three sections: Section 1 - Risks and Mitigation Strategies assesses the integrity and risk profile of the app; Section 2 - Function is focused on descriptive aspects related to accessibility, costs, developer credibility, evidence and clinical foundation, privacy and security, usability, functions for remote monitoring of the user, informed consent, cultural competency, access to crises services and artificial intelligence; and Section 3 - Mental Health App Features focuses on specific features unique to the primary function of the app.

## **Strengths and Limitations**

The FASTER to Mental Health and Wellness is aimed at facilitating the use of apps for mental health support and recovery through standardized evaluation, screening, and classification of apps. Several of the criteria have been extracted from extant frameworks in the app evaluation and mental health space. However, we identified several gaps in the existing frameworks and addressed them through further prioritization of criteria, the addition of criteria to assess risks and safety of the apps, and assessment of the use of AI and other engagement approaches.

We acknowledge that apps targeted at specific disorders may benefit from an assessment that is quite specific for those types of apps; however, also recognize that developing such disease area specific frameworks is time and resource-intensive and is unlikely to be a practical approach. Lastly, the framework results in summary conclusions on various aspects of the app such as usability, security, etc. As a next step to facilitating adaptation and adoption, it might be valuable to gather user (e.g., providers, patients, etc.) input on the value of these summative conclusions in guiding decisions around the use of the apps.

## **Implications and Conclusions**

We envisage a range of possible applications of this framework. First, the framework can be used by mental health organizations and mental health advocacy agencies to provide a curated and validated mental health app library. Such a library could be used by consumers, family members, peer supports, and health care providers to review and select apps as a resource for patients. Second, the framework can be leveraged by employee health plans, health system leaders, public and private insurance providers, and other entities to review and provide guidance for apps relevant to their members. Lastly, the framework can be used by app developers as guidance to promote transparency in communication about the potential benefits, risks, and evidence regarding their apps.



# Introduction

## Background

Among adults aged 18 or older in the United States, the prevalence of ‘mental illness in the past year’ increased from 17.7 percent (or 39.8 million people) in 2008 to 20.6 percent (or 51.5 million people) in 2019.<sup>1</sup> Of this latter population, 26.0 percent (or 13.3 million people) perceived an unmet need for mental health services. The most common reason given for unmet needs is affordability, with other barriers including mental health stigma and provider shortages or wait time.<sup>2</sup>

According to the WHO, “mental health is a state of well-being in which an individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively and is able to make a contribution to his or her community”.<sup>3</sup> Mobile health apps (mHealth) have great potential to provide much needed access and equity in mental health care and wellness. Mobile health apps are applications that run on smartphones or tablet devices and contain content related to health and wellness. Mental health apps are being used to support diagnosis, treatment, and management for mental health illness, as well as to provide wellness support through meditation and mindfulness. Some mental health apps can provide diagnostic support or assist in the diagnostic pathway by improving time to diagnosis, for example by offering automated standardized mental health assessments. Apps might also facilitate treatment for certain mental health conditions and provide therapeutic support. Firth et al. found in their meta-analysis that apps targeting depression through self-management and alleviating symptoms significantly reduced symptoms of depression when compared to controls.<sup>4, 5</sup> The use of mental health apps can also reduce anxiety when compared to controls.<sup>6</sup>

Apps provide patients with mental health information and help enhance care by building skills and supporting patients in between visits with healthcare providers. It also adds the convenience of accessing mental health care remotely. Mental health apps are likely to lower the costs of traditional mental health care. Price et al. argue that mHealth holds a lot of potential to provide more accessible mental health care, concluding that mental health apps can be a tool to provide better patient education and engagement, assist within the treatment plan, and can augment post-treatment care as well. Improved care coordination and focusing on patient-centered care is another potential benefit of mental health apps.<sup>7</sup> Some apps deliver greater efficiency by providing feedback loops and assessments for providers in between appointments and also deliver round-the-clock support and direct access to crisis lines for patients.<sup>8</sup> While fewer mental health apps provide a direct connection to providers, one potential benefit of apps is the ability to extend mental health care beyond the traditional brick and mortar walls of health care providers.<sup>9</sup>

Mental health apps may provide more privacy than in-person mental health visits and could reduce stigma associated with seeking mental health care. At an individual level, mental health apps can be a resource that provides psychoeducation and medication management, skill building, symptom tracking, thereby amplifying the benefits of existing therapy.

It is well established that racial and ethnic disparities exist in access to and use of healthcare services and similar differences are also seen in access to and utilization of mental health care services.<sup>10, 11</sup> While some studies explain that disparities exists due to a lack of access to care including insurance coverage, others, when comparing similar sociodemographic factors, suggest that the differences may be more related to mistrust of traditional mental health services, discrimination, language, and cultural factors.<sup>12</sup> Black and Asian individuals are screened for

depression at a lower rate when compared to white individuals, and for those that receive screening, Black individuals, Latino males, and Asian individuals receive mental health care at a lower rate when compared to white counterparts.<sup>13</sup> Disparities between urban and rural populations' access to mental health exist as well. Up to 65 percent of non-urban counties lack psychiatrists while more than 60 percent of Americans in rural areas live in a Mental Health Provider Shortage Area.<sup>14</sup> Apps could help address disparities in mental health services and provide an option for those who do not trust health professionals or have not benefitted from traditional mental health services. Many mental health apps are affordable, offer a way to reach rural populations and traditionally hard-to-reach groups, and can reduce patients' feelings of stigmatization or discomfort with traditional treatment.<sup>15</sup>

Mental health mobile apps can be readily disseminated with minimal need for staff training or resource investment, and thereby reduce the burden on healthcare settings. Furthermore, technology-enhanced healthcare approaches are increasingly being reimbursed by insurers, making apps more likely to be integrated into the toolkits of healthcare providers.

The market for mental and behavioral health is booming with turbocharged funding through private and venture-back investors. In 2020, \$2.4 billion was raised for startups working in mental health, equal to 17 percent of all funding for digital health in 2020.<sup>16</sup> Segmenting the space into only apps, \$1.2 billion were raised in 2021, which is a 50 percent increase as compared to the prior year.<sup>17</sup> Some driving factors for funding may be the huge increase in demand for mental health services due to the detrimental effects of the COVID-19 pandemic. The isolation, job loss, economic and financial hardships have increased the stress and anxiety level in the general population. For the first time, during the pandemic, the FDA began approving digital mental health therapeutics solutions. Mobile apps, which may fall into the Software as a Medical Device (SaMD) category for the FDA follow the same regulatory guidelines as other medical devices using the same 3 classifications for approval. While there are about 20,000 mental health apps on the market in the Apple App Store or Google Play Store, only 5 of them have FDA approval.<sup>18</sup> The first prescription digital therapeutic for mental health was Pear Therapeutics's reSET therapy in 2017 which is powered through an app.<sup>19</sup> The FDA has taken a "hands-off" approach towards regulating mental health apps that do not fall into the realm of "device software functions". The FDA has also introduced a Software Precertification (Pre-Cert) Pilot Program that provides streamlined regulatory oversight of software-based medical devices developed by certain manufacturers that have consistently demonstrated quality and organizational excellence. The goal of the program is to have a less burdensome regulatory oversight for these organizations to establish trust with the FDA so that they can develop high quality SaMD products.<sup>20, 21</sup>

Rapid proliferation of health apps has resulted in both, haphazard as well as sub-optimal use of these apps, with potential dangers to patients and end-users, since the health system and physicians may not be well-versed enough regarding the purpose, safety, and efficacy of these apps to recommend or prescribe them. There is no existing roadmap for informing the selection of mental health apps for patients and for clinicians. "Prescribing" an app as an adjunct to treatment is restricted owing to the limited evidence on the efficacy of the apps, as well as emerging concerns about the usability, privacy, and safety risks.<sup>22</sup> Apps that claim to provide a diagnosis or that target individuals who may be vulnerable financially or socially owing to their mental health condition, can pose a serious risk.<sup>22</sup> Mental health apps may pose a high risk to users with serious or acute mental disorders such as depression and suicide ideation, especially if the apps do not provide easy without access to crisis lines in case of emergency. Mental health

apps may pose a high risk to users with depression and suicide ideation, especially without access to crisis lines in case of emergency. According to Martinengo et al., there have been more than 2 million downloads of mental health apps without or with inaccurate suicide crisis phone numbers.<sup>23</sup> Privacy and security, particularly related to private health information, is also a concern for mental health apps. Dehling et al. ranked the level of damage to users due to privacy infringement and found that 95 percent of apps pose at least some threat.<sup>24</sup>

There are many frameworks that are being used to evaluate digital health apps, including those that focus on mental health apps.<sup>25-29</sup> Several of these frameworks are being used by advocacy agencies and online health resource platforms to recommend apps to interested users. However, most existing frameworks are geared towards evaluating specific aspects of health apps (e.g., such as usability), and do not adequately reflect concerns around assessment of risks posed by the apps, as well as recent advancements in artificial intelligence (AI), and their use by apps for automating certain diagnostic (e.g., Ada) or counselling protocols (e.g., Woebot, Replika).

## **Objectives of This Technical Brief**

Given the uncertain evidence-base for most mental health apps coupled with potential app safety/privacy concerns in a more vulnerable population of users, the decisional dilemma is “how can consumers, family members and peer supports, and providers and health systems select mental health and wellness mobile apps?” The aim of this technical brief is to develop a framework to assess the safety and effectiveness of mental health and wellness apps for users across different age groups.

## Guiding Questions

This Technical Brief was guided by the following questions:

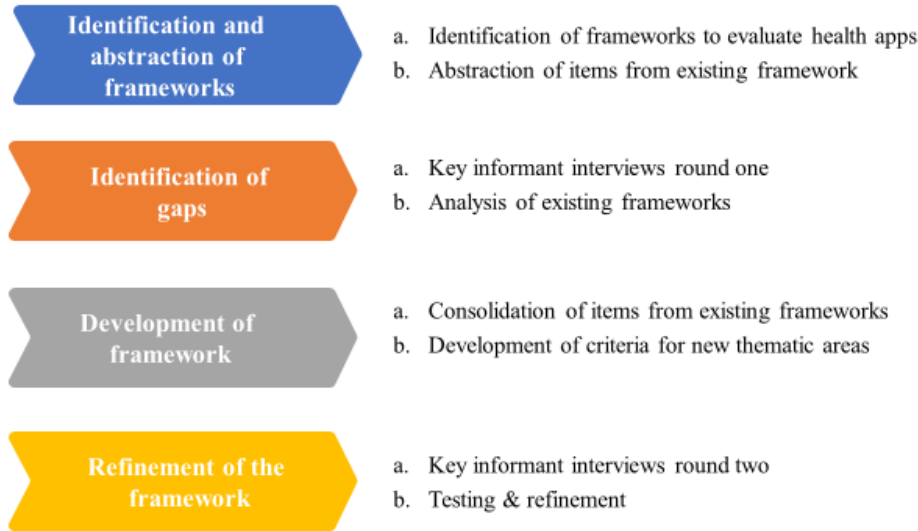
1. What characteristics and minimal standards of available mental health mobile apps need to be analyzed to assess the appropriateness (to various stakeholders) and effectiveness of available apps, to include but not be limited to:
  - Accessibility, including ease of use, health literacy, 508 compliance, digital equity, cost;
  - App background including funding source and purpose;
  - Security features and privacy policy such as data ownership/usage;
  - Clinical foundation and linkage to current evidence-base;
  - Usability, including interoperability across platforms and stability;
  - Therapeutic goals, linkage to the provider, crisis warning notification/alert system?
  
2. Identify or develop an assessment framework for mental health apps and apply the framework to help consumers, family members and peer supports, and providers and health systems select apps. The framework will take into account current FDA status on the use and classification of risks of apps in healthcare.

# Methods

The methods for this Technical Brief follow the Content and Procedures Guide for the Evidence-based Practice Center Program. The protocol was posted on AHRQ’s Effective Health Care website (<https://effectivehealthcare.ahrq.gov/products/mental-health-apps/protocol>). The Technical Brief is organized by two Guiding Questions that address development of a framework for assessing mental health apps.

Figure 1 displays the process of developing the framework.

**Figure 1. Framework development process**



## Identification and Abstraction of Existing Frameworks

### Identification of Frameworks to Evaluate Health Apps

We identified a number of frameworks that focused on the assessment of general health apps, as well as mental health apps. Our review of existing frameworks was guided by our knowledge of frameworks, a rapid literature search for mental health app assessment frameworks using PubMed, and the identification of additional documents through a reference review of included documents. Examples of frameworks for the assessment of general health app assessment include MyHealthApps,<sup>30</sup> Healthy Living Apps Guide,<sup>31</sup> Digital Technology Assessment Framework from National Health Service UK,<sup>32</sup> and Digital Therapeutics Alliance.<sup>33</sup> Examples of frameworks that are focused on mental health app assessment include One Mind Psyber Guide,<sup>25</sup> American Psychiatric Association Initiative,<sup>26</sup> Kaiser Permanente,<sup>27</sup> VeryWellMind,<sup>28</sup> and HealthNavigator.<sup>29</sup> Other notable frameworks reviewed include M-Health Index and Navigation Database (MIND) and the end-user version of the Mobile Application Rating Scale (uMARS). MIND is an operational and flexible framework based on American Psychological Association (APA) App Assessment Framework that includes 105 questions that have been harmonized from 79 frameworks.<sup>34</sup> The end-user version of the Mobile Application Rating Scale (uMARS) provides a comprehensive set of questions about engagement and usability.<sup>35</sup>

We also conducted a search for documents related to the regulation of health apps, including by the FDA, as well as by Federal Institute for Drugs and Medical Devices in Germany and the

UK's National Institute for Health and Care Excellence (NICE) Framework.<sup>36, 37</sup> Finally, we identified relevant systematic reviews on the efficacy and limitations of typical features found in mental health apps such as mindfulness, mood, and symptom trackers, journaling, social and peer interaction, psychoeducation, skill-building, to better understand how they might be relevant to a framework to assess mental health apps.

## **Abstraction of Items from Existing Frameworks**

We abstracted items from 11 existing frameworks (see Appendix A for a list of frameworks). This process yielded 300 items/questions. These items typically aimed to capture similar concepts such as aspects of credibility, safety, available features, accessibility features for those with disabilities, clinical evidence foundation, interoperability, usability, efficacy, and data privacy and security and were therefore categorized under common themes. Items that were clear and easier to operationalize were retained. Those that were too technical to be operationalized by individuals without specialized training, were confusing and unclear, or required extensive research to answer were simplified or removed.

## **Identification of Gaps**

To identify areas for further development of an effective framework to evaluate mental health apps, we conducted key informant interviews to identify priorities from various stakeholder perspectives and analyzed categories that were covered in the existing frameworks.

## **Key Informants Interviews Round One**

We conducted two series of key informant (KI) interviews. We selected KIs to represent a range of expertise and perspective from stakeholders. Patients, family members and caregivers of those living with mental health concerns, advocates; clinicians with a background in mental health, primary health care, and emergency medicine; payers and health system representatives were included in the first phase of KI interviews. The family/patient representative was identified from one of the largest mental health advocacy organizations (National Alliance on Mental Illness, NAMI). For the second phase of KI interviews, we recruited app developers and mental health providers with app development expertise. We were also interested in hearing the perspectives of KIs with expertise in developing frameworks to assess mental health apps. Specifically, we worked to identify KIs with some familiarity with FDA regulations around risk stratification for apps and familiarity with existing challenges around "software as a medical device", health, privacy and security risks. We prepared lists of possible KIs and listed alternates for each perspective we wanted represented (i.e., per particular discipline, or specific MH aspect) and obtained input from our partners and approval from AHRQ.

As we were extracting and considering criteria from existing frameworks, we conducted the first set of interviews with stakeholders representing: family members of those living with mental illness; clinicians with a background in mental health, primary health care, and emergency medicine; and payers. We asked these stakeholders to provide guidance on their perceptions and experience with mental health apps and the essential features and omissions in existing frameworks that they were familiar with.

## **Analyses of the Existing Frameworks**

We reviewed the categories covered by the existing app evaluation frameworks, as well as other literature highlighting priority considerations for such frameworks. This literature included documents on regulatory and safety considerations for medical software. Based on this assessment, and feedback from the KIs interviews, key gaps were identified in the existing framework (described in further detail in Results).

## **Development of the Framework**

### **Consolidation of Items from Existing Frameworks**

We systematically selected, adapted, and where necessary, modified the abstracted criteria from the frameworks. To the extent possible, we leveraged existing criteria that have been proposed and used by other frameworks to avoid “recreating the wheel.” Where necessary, we simplified the abstracted criteria and standardized the language for clarity. We were guided by the following principles in the abstraction and revision of the criteria: One, the question posed by the criteria should be relatively easily answered based on a review of the app developer website, and through downloading the app; it should not require a systematic literature search or engagement with the app developer. Two, the framework should be usable by someone with some knowledge of technology and mental health, but evaluators do not need to be experts in either domain. Three, the application of the framework should be largely objective with as much specificity in the criteria as possible, to increase reliability and reduce subjectivity in the evaluation process. To further facilitate this third principle, we developed a training guide to guide the application of the framework.

Several criteria from the MIND framework, which operationalizes the American Psychiatric Association (APA) framework, were adapted.<sup>34</sup> The assessment of the usability of apps continues to pose several challenges, several of which are addressed by the uMARS framework, which has been developed and validated for the assessment of engagement, information, quality, and aesthetics.<sup>35</sup> Questions 3, 5, 6, 7, 8, 9, 10, 11, 12, and 13 of the UMARS framework were incorporated in the our framework. Questions on entertainment, interest, interactivity, the quantity of information, and visual information were more subjective and hence omitted. Questions on the credibility of the source were omitted as they are covered in the Risk Assessment section of our framework. Questions on the subjective quality of apps and the perceived impact of the app are covered in the post-administrative section of our framework where the reviewer can provide their subjective assessment in free text.

### **Development of Criteria for New Thematic Areas**

For the identified gaps that were also prioritized in the KI interviews, we developed new criteria based on peer-reviewed literature, and engagement with internal experts. For the assessment of risks posed by the apps we used normative guidance provided by agencies such as National Institute for Health and Care Assessment Framework (NICE)<sup>38</sup> and the FDA Clinical Decision Support Draft Framework and Software as a Medical Device guidance.<sup>21, 39, 40</sup> For the development of criteria on the use of AI, we reviewed literature on issues of safety in the use of AI for health apps and consulted with the ethicist on our team.

Our approach for new areas of development warrant greater detail, as described below.

## Development of Criteria on Risk Assessment

To develop criteria on the assessment of risks of apps, we reviewed key regulatory documents, including policies from FDA's Digital Health Center of Excellence (DHCoE). DHCoE aligns and coordinates digital health work across the FDA and is chartered with developing a comprehensive approach to regulation of digital health technology. The FDA has taken a "hands-off" approach towards regulating mental health apps that do not fall into the realm of "device software functions".

Device software functions may include "Software as a Medical Device (SaMD)" and "Software in a Medical Device (SiMD)". If a software function meets the definition of a device that is deployed on a mobile platform, it is referred to as a "mobile medical app". Some examples of mobile apps that are regulated by the FDA include those that use a mobile platform's built-in features such as light, vibrations, camera, to perform medical device functions, software functions that control the operation of an implantable or body worn medical device, and software functions that are used in active patient monitoring to analyze patient-specific medical device data.

We also reviewed FDA guidance on clinical decision support software (CDS).<sup>41</sup> A draft document around *Software as a Medical Device: Possible Framework for Risk Categorization and Corresponding Considerations* the risk-based approach to regulating CDS.<sup>42</sup> The risk categorization is divided into 4 categories (I, II, III, IV) based on the impact on individual health driven by the stated intent of the CDS to *treat or diagnose individuals, or drive or inform clinical management*. The risk categorization approach in Section 1 of our framework uses this lens by first assessing the level of risk the app might pose, based on its stated objectives, and second by determining an appropriate level of evidence, given the level of risk.

## Development of Informed eConsent Criteria

To develop criteria related to informed consent we studied published literature on ethics of electronic consent and consulted with our ethicist advisor. A book on Digital Contract Tracking Technologies (DCTT) emphasized the importance of "privacy by design"; that is, building privacy and security protections into the design of technology.<sup>43</sup> Consent procedures for apps usually require users to agree to detailed legal consent forms which may be incomprehensible. The DCTT stresses the importance of incorporating meaningful mechanisms to obtain consent which are easily understood by the user.<sup>43</sup> We also looked at a simple open-source smartphone consent module that was developed by Sage Bionetworks for research uses. Its recommendation includes simple and straightforward information, deliberately organized content, multimodal learning (e.g., visual, audio, written), accessibility for disabled users, multilingual text, and engagement through interaction (e.g., swiping to navigate forward and backward).<sup>44</sup>

## App Integrity and Organizational Credibility

The trustworthiness of an app is based not just on its content, usability, and technology, but also on organizational attributes of reputation and brand.<sup>45</sup> Discussions with KIs highlighted the importance of a viable business model so apps could be maintained and updated regularly. Apps had the potential to cause harm that would outweigh the benefits if individuals using it suddenly found that they no longer had access to the app due to it being discontinued or not being updated to be in compliance with the latest system guidelines. Based on this guidance, we developed app integrity and organizational credibility guidelines.



## Cultural Competence

The definition of cultural competence in the framework was adapted from the Health and Human Services' definition of cultural competence.<sup>46</sup> SAMHSA's research shows that cultural responsiveness can improve client engagement in services, therapeutic relationships between clients and providers, and treatment retention and outcomes and that it is critical to reducing disparities in behavioral health.<sup>47</sup> Based on internal discussions, the definition of cultural competence was expanded to include groups with lived experiences such as pregnant teens and survivors of gender-based violence. This definition was further expanded to include gender sensitivity to highlight its importance in healthcare.<sup>48</sup>

## Definition of Vulnerable Populations and Additional Risk Assessment for this group

Discussions with the first set of KIs highlighted of the unique needs of caregivers that were either responsible for a minor or an adult with substantial impairment. In order to ensure security and safety of minors or individuals with substantial impairment, there needed to be some level of monitoring by caregivers. Based on these interactions, we added consent requirements for caregivers of vulnerable populations.

We considered defining a *vulnerable adult* based on the severity of mental health disorders or conditions. The level of disability and subsequent impact on the quality of life may vary by individuals and over time, irrespective of a mental health diagnoses, making it hard to objectively define a vulnerable adult for the purpose of this framework. We based the definition of a vulnerable adult on their level of impairment rather than the severity of any mental health disorder diagnosis.

## Mental Health Categories

At first, we defined mental health symptoms and diagnostic categories from the *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. The DSM-5 is an authoritative source that defines and classifies mental disorders in order to improve diagnosis, treatment, and research. Guided by the symptoms and diagnostic criteria outlined in the DSM-5 and the International Classification of Diseases (ICD), we searched for systematic reviews on the efficacy on mental health apps to support a range of mental health symptoms and conditions. We reviewed recommended features in mental health apps through systematic reviews of different categories of mental health apps such as those for neurocognitive disorders, personality disorders, sleep-wake disorders etc.<sup>23, 49-56</sup> Our analysis of the literature showed the common features in mental health apps to be mindfulness, mood trackers journaling, social and peer interaction, psychoeducation, skill building. etc. At first, we attempted to develop a list of features of mental health apps based on the specific condition the apps address. However, given that features such as mindfulness and mood tracking can be useful across a range of conditions, and the evidence to guide the use of specific features for specific conditions is limited, we developed a comprehensive listing of features that mental health apps might have (without associating these features with a condition).

## Refinement of the Framework

A draft of the framework was then shared with the second round of key informants, and iteratively tested and revised through 5 rounds of applying the framework to a variety of mental health apps.

## Key Informants Interviews Round Two

The second set of interviews included app developers in the mental health space, as well as clinicians who had experience in app development and a psychiatrist working for the FDA. We solicited feedback from this group on the initial draft framework. This group included several health care providers working in primary care, emergency medicine, and psychiatry/psychology; representatives from the Veteran’s Health Administration and public and private insurance payers.

## Testing and Refinement

The draft framework was applied to a range of mental health apps and iteratively refined. The purpose of the iterative testing was to improve the relevance of the items within the framework to mental health apps targeting a range of mental health conditions, to standardize the language used to describe these items so that they are well-understood by evaluators with varying expertise levels, and to develop the accompanying training guide to systematically apply the framework.

## Search Approach for Mental Health Apps

We used mental health symptoms and diagnostic categories from the *Diagnostic and Statistical Manual of Mental Disorders (DSM–5)* to guide our search and selection of mental health apps. We cross-checked the main diagnostic categories with mental health conditions addressed by the current mental health apps in 42Matters and included those addressed by at least one app. In addition to the DSM-5 categories, we added categories for Self-harm and Suicide; Mental Wellness - meditation, mindfulness; Other mental illness – psychotherapy, transdiagnostic (addresses symptoms across multiple disorders); and Other (Table 1).

**Table 1. Categorization of mental health disorders**

Category	Disorders/Description
Anxiety Disorders	Agoraphobia, anxiety, social anxiety - phobia, separation anxiety disorder, generalized anxiety disorder, panic disorder, panic attack
Bipolar, Depressive Disorder, and Related Disorders	Bipolar disorder, major depressive disorder, mood disorder, depression, dysthymia
Mental Wellness	Meditation, mindfulness, stress management
Neurocognitive Disorders	Dementia, Alzheimer's
Neurodevelopmental Disorders	Autism, language disorders, learning disorders, traumatic brain injuries, neurogenetic disorders, motor disorders, intellectual disabilities, communication disorders, ADHD
Obsessive-Compulsive and Related Disorders	OCD, body dysmorphic disorder
Personality Disorders	Borderline personality disorder, personality disorder
Schizophrenia Spectrum and Psychotic Disorders	Schizophrenia, psychosis
Self-Harm	Suicide
Sleep-Wake Disorders	Insomnia, narcolepsy
Substance-Related and Addictive Disorders	Smoking, binge drinking, opioid use disorder, substance use disorder
Transdiagnostic	Symptoms across multiple disorders
Trauma- and Stressor-Related Disorders	PTSD, adjustment disorders, traumatic stress exposure

Category	Disorders/Description
Eating Disorders	Bulimia, anorexia nervosa, binge-eating disorder, eating disorder
Other mental illness	Other mental illness that may not be listed above

Apps were identified using 42matters (available at <https://42matters.com/>). 42matters is a database with names, developers, descriptions, and a range of technical details of apps contained in the Apple iTunes store and Google Play marketplace. Search criteria on 42matters were limited to apps that fell in the Interactive Advertising Bureau (IAB) categories of “Healthy Living” and “Medical Health,” were released after January 1st, 2010, were in English, and were available in the United States.

Search terms used were autism, language disorders, learning disorders, traumatic brain injuries, neurogenetic disorders, motor disorders, intellectual disabilities, communication disorders, attention-deficit/hyperactivity disorder (ADHD), schizophrenia, psychosis, bipolar disorder, mood disorder, depression, dysthymia, agoraphobia, anxiety, social anxiety, phobia, separation anxiety disorder, generalized anxiety disorder, panic disorder, panic attack, obsessive-compulsive disorder (OCD), body dysmorphic disorder, posttraumatic stress disorder (PTSD), adjustment disorders, anorexia, bulimia, binge-eating disorder, eating disorder, insomnia, narcolepsy, addiction, smoking, binge drinking, opioid use disorder, substance use disorder, dementia, Alzheimer’s, borderline personality disorder, personality disorder, self-harm, suicide, mental wellness, meditation, mindfulness, stress, mental illness, and psychotherapy. Searching was conducted on May 31st, 2021.

We aimed to screen the apps to identify 50 apps from the App Store (iOS) and 50 Google Play Store (Android) per mental health category for each of the 15 categories. Screening was based on a description of the app provided on 42matters and considered apps eligible if the app was either about mental health or wellness and included mental health related therapeutic content.

For each category, the list of apps was deduplicated using Microsoft Excel. Following this, the apps were randomized within the mental health categories, with the goal of identifying six apps from within each mental health category for further evaluation. The apps were then serially downloaded and assessed further for inclusion. If the app was assessed as ineligible, the next app in the bin was selected for evaluation, till the target goal of 6 apps per bucket was reached. Apps were excluded if they could not be downloaded or opened due to technical issues, they were unrelated to mental health, they were solely for the purpose of telehealth, if an access code was required to use them, or if they required a payment of greater than \$100 to access them. Note that apps selected for the pre-pilot round were purposively purposefully selected and did not follow this procedure.

## **Iterative Application and Refinement of the Framework**

To test and refine the framework, we conducted one pre-pilot round of testing with the core team, followed by four additional pilot rounds of testing by external evaluators. A total of 27 apps were evaluated by at least two reviewers. Ten apps were tested during the pre-pilot round and interrater reliability was assessed for Section A during this round. The pre-pilot round was intended to provide the research team with greater clarity on the relevance, flow, and clarity of the framework. Following this several changes were made, and the framework was iteratively applied by external evaluators.

The reviewers were trained on using the framework using the guide (Appendix F) and explanations included in the framework (Appendix G). The initial training took 2 hours. Follow-up training was conducted after the first pilot round of applications, lasting another 60 mins. There were 11 reviewers across the 5 rounds with varying levels of experience from 2 to 30 years in epidemiology, mental health, project management, technology, and public health. Three of the evaluators were undergraduate public health students, two were graduate public health students, and one was a doctoral student. Five reviewers had a background in mental health and one in technology.

For pilot rounds 1 to 4, following the pre-pilot, item-level inter-rater reliability was calculated after each pilot round and averaged for each category of questions in Sections 1 to 3 of the framework. Two-hour meetings were conducted after each pilot round to discuss the items where there was disagreement. Each of the disagreements was discussed and clarified for understanding with reviewers. If the disagreements were not due to oversight or an error, changes were made to either the question, responses, or the explanation was further enhanced to improve clarity. If certain themes were not relevant to certain types of apps, they were considered for omission. The changes made to the framework were categorized in one of seven ways, as “Modified the question Language for clarity”, “Removed question”, “Added question to further capture the concept”, “Added additional guidance notes”, “Added or consolidated response options”, “Added questions that aren’t asked elsewhere”, and “Question rearranged sequentially in the framework”.

## **Peer Review and Public Commentary**

Experts in mental health, app developers, and individuals representing stakeholder and user communities were invited to provide external peer review of this technical brief. AHRQ task order officers and an associate editor also provided comments. We addressed all reviewer comments and revised the framework and technical brief, as appropriate. The peer-reviewed draft technical brief will be posted on the AHRQ website for four weeks to elicit public comment.

# Findings

## Gaps in Existing Frameworks

Most frameworks had common criteria assessing aspects of credibility, safety, available features, accessibility features for those with disabilities, clinical evidence foundation, interoperability, usability, efficacy, and data privacy and security. Based on the input from KIs and review of frameworks, the following gaps were identified:

### Risk and Credibility

As mental health apps gain traction, it is important to understand risks posed by the app and balance it against the available evidence to understand its clinical safety, efficacy, security and privacy, and institutional and financial stability. While the existing frameworks include questions on privacy, security, risk, and evidence, the existing frameworks did not attempt to assess the risk posed by the app and assign a safety and credibility rating to it.

### Vulnerable Populations

Given that the target users of mental health apps might include vulnerable populations and minors, there is a need for additional checks on apps to ensure that they address the unique needs of such populations. Such features may include how and where security and data use procedures are explained, whether there are costs associated with the use of the app that may not be obvious to a user, etc. Currently, there is no clear guidance on the type of checks and balances that should be in place to ensure the safety of vulnerable populations.

### Accessibility Features

While frameworks do talk about accessibility, they do not distinguish between features that are part of the phone's accessibility options versus specialized features that may have been developed within the app.<sup>57</sup> Phones provide a host of accessibility features and settings such as text-to-speech, colorblind color scheme, and text and resolution adjustment. Other features such as screen reader, adapting audio/video content with transcription can be developed by the app developer to improve user experience. Specific questions were added to distinguish between the smartphone's accessibility features versus those added by the app developer.

### Cultural Sensitivity

Increasingly, several mental health apps are being developed considering the unique lived experiences of various populations. Cultural appropriateness can be an important way in which apps distinguish themselves and become valuable for marginalized populations. The frameworks we reviewed do not assess if the app is targeted at specific cultural groups or uses language that is inclusive of certain populations.

### Artificial Intelligence

The use of machine learning and AI has been rapidly growing in mental health apps and poses unique challenges to the safe use of the apps. AI in healthcare has huge potential to transform it for the better and bring about efficiencies and innovation into clinical practice.

However, there are challenges with informed consent for use, security and transparency, algorithmic fairness and bias, effectiveness and privacy.<sup>58</sup> FDA provides guidance but no regulation around mobile apps that use AI. Existing frameworks do not assess if and how Machine Learning/Artificial Intelligence approaches have been used and whether they incorporate user feedback to improve the accuracy and validity of their recommendations and predictions.

## Mental Health Features and Function

There are many common features provided by mental health apps such as mindfulness, safety planning, journaling, automated chatbots, gamification and social and peer group interactions. A user may want to know about specific users supported by the app to see if it is suitable for use by them. Most app assessment frameworks do not provide a comprehensive list of these features.

## Key Informants Interviews

We completed four 1-hour interviews with twelve Key Informants, each with relevant expertise (Box 1).

### Box 1. Key Informants' expertise

<p><b><u>First set of calls</u></b></p> <p><b>Patient/Caregiver/Advocate perspective (n=1)</b></p> <p><b>Clinicians (n= 3)</b> Primary care Psychology Emergency medicine/app developer</p> <p><b>Health System Perspective (n=1)</b> Veterans' Administration System</p> <p><b>Payer Perspective (n=3)</b></p> <p><b><u>Second set of calls</u></b></p> <p><b>Psychiatrists (n=2)</b> <b>Emergency medicine/app developer (n=1)</b> <b>App developers (n=2)</b></p>
--

## Key Informants Interviews Round One

The first set of KI interviews were scheduled at the beginning of the project. The first call was with an individual representing the patient and family member's perspective. The KIs mentioned that they were more comfortable recommending mental health apps that utilize evidence-based practices (e.g., CBT) and those with privacy and consumer/patient protections. They also mentioned that individuals with mild-to-moderate mental health concerns and functional ability that is high enough to engage with apps could potentially benefit from mental health apps.

In summary, key takeaways were: 1) apps can fill a mental health service gap in areas where there are limited or no mental health professionals, for those who lack behavioral health

insurance, and when someone is on a waitlist for care; 2) apps could be ideal for accessing psychoeducation/information to increase mental health knowledge; and 3) some apps are geared toward individuals and others toward groups; 4) certain apps should have a decision alert or clinical monitoring tool to alert the individual, their family support network or medical provider to the need for additional, more intensive mental health or crisis resources. The KIs noted that there is an appeal to the potential ability to use a framework to take a more personalized medicine approach to mental health apps (i.e., will a specific app work for my specific client). Safety and adverse effects of apps were of concern. One barrier to healthcare providers recommending apps is the view that if they prescribe, providers are responsible for risk/harm of apps as well as benefits.

## **Key Informants Interviews Round Two**

The initial draft framework was presented to KIs in a third and fourth call to elicit feedback. The third KI call included two psychiatrists - one the developer of an existing framework and the other a representative from the FDA (Box 1). The KIs mentioned that there is a need to educate patients and clinicians about apps. Engagement metrics are not available to help in making informed decisions; the evidence base is not very strong for the majority of apps. In terms of new framework components, clinical decision support and risk assessment were highlighted, due to possible unintended iatrogenic impacts of technology in some individuals with mental health conditions as well as the rapidly changing or deteriorating aspects of some mental health symptoms and conditions. Ideally, mental health apps could detect the need for more intensive services and/or provide direct linkage to a crisis text line or hotline.

The fourth KI call included app developers who were also clinicians. The KIs mentioned that the draft framework should place more emphasis on the scientific evidence to support specific apps. Also, app users need to know where their data are going, specifically providing the ability to opt-in versus opt-out for sharing certain data. The KIs discussed how the framework could be used. There are at least three different ways providers could use the framework for decision making: responding to a patient who says, "I use this app for x, what do you think, doctor?"; recommending an app; and prescribing an app. Recommending an app is quite different from prescribing an app, and the framework should aim to clarify this difference, as reflected by the level of scrutiny needed to justify a certain approach. Efficacy and effectiveness are important for apps, as is evidence of doing no harm. Currently there are apps in the mental health space as transdiagnostic tools, rather than specific to a particular mental health diagnosis. It would be hard to put many of the existing apps into specific diagnostic buckets. Things may change because the development of mental health apps is a rapidly changing domain (Box 2).

## Box 2. Key components of framework from the Key Informants

- Data Transparency/Privacy and security policy;
- Evidence base (has at least one randomized clinical trial supporting it; does the app seem to be overselling, without any evidence?);
- Credentials and experience of development team;
- Mission and vision of the development organization (e.g., private sector, non-profit, government);
- Commitment to principles of social justice, equity, and inclusion, e.g., reading level, apps in multiple languages or set to a specific educational level, are some apps video-based or audio-based for individuals for hearing or vision impaired individuals, 508 compliance - safe and accessible for people with disabilities;
- Culturally sensitive;
- Whether mental health apps were beta tested by users with mental health conditions or patients;
- Usability and accessibility;
- Developer responsiveness (i.e., is the app being regularly updated and are bugs being addressed?);
- Clear understanding of outcome targets (e.g., does the app clearly identify what changes/outcomes the user may expect?);
- Clear target audience (e.g., adults or adolescents) - apps should be transparent about which types of patients/populations benefit, in which settings (e.g., inpatient, outpatient, etc.), and under what conditions;
- Safety/crisis response [does the app connect with crisis hot line (call) or Crisis Text Line, are risks identified in apps communicated to family member, loved ones, or provider; have there been any adverse effects?];
- Endorsed by trusted mental health professional association or psychiatric association/agency/mental health advocacy group;
- Whether existing Health plans, Employee Health Program/ Employee Assistance Program, service systems, settings offer the app; uses an evidence-based clinical model/framework;
- Allows for personalization and personalized feedback;
- User experience (star ratings, reviews by users);
- Costs;
- How app will be accessed by users: apps sometimes available via health plan platforms like myStrength or Recovery Record or programmatically deployed through an Employee Assistance Program or college campus.

## Testing and Refinement of the Framework During Pre-pilot and Pilot Rounds

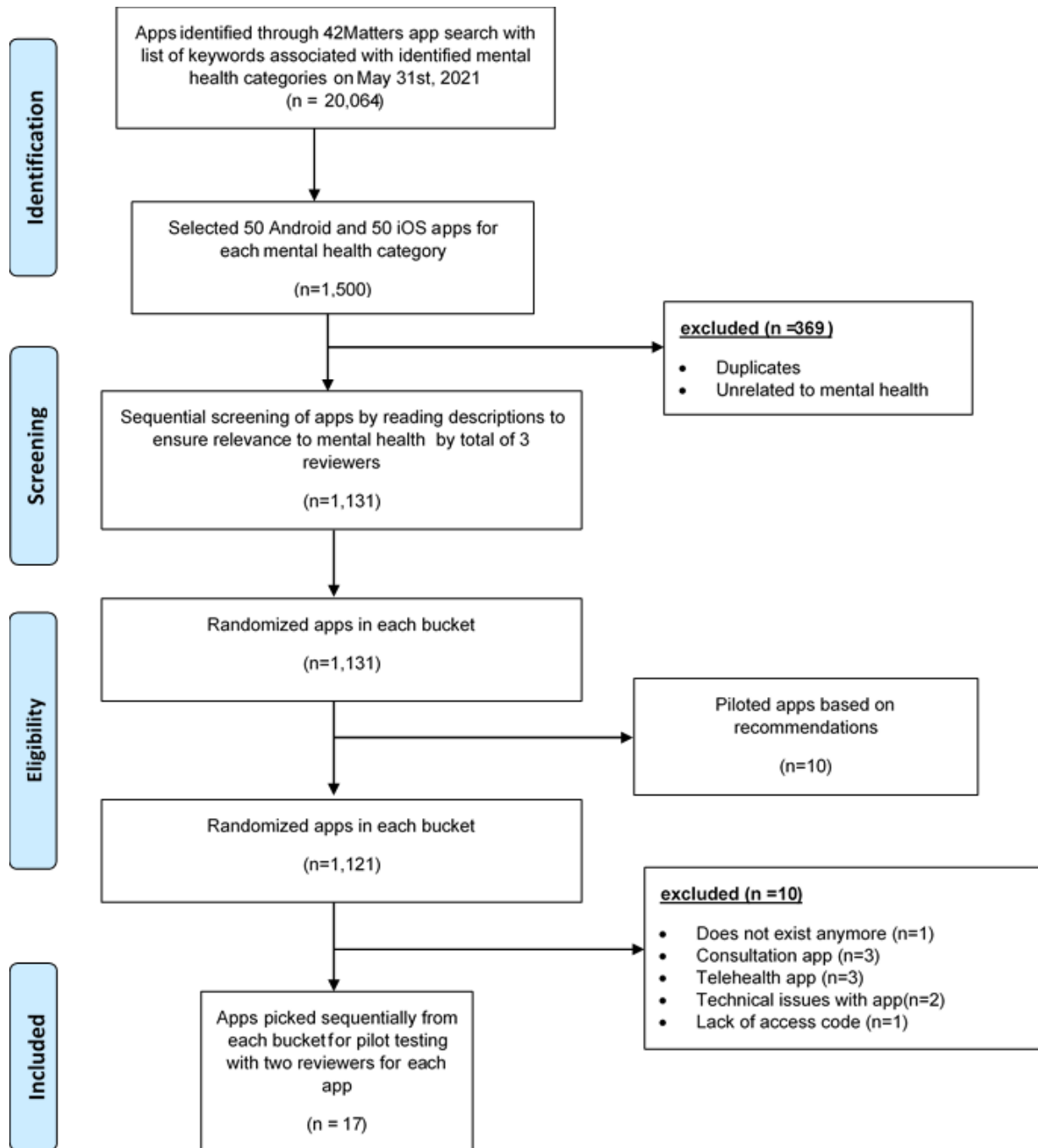
### Results of the App Search Process on 42Matters

We conducted a pre-pilot test of the draft framework using a convenience sample of 10 apps. These 10 apps were selected to represent diversity in the target age group, mental health conditions or symptoms, and app functions (see Appendix D). Free versions of apps that did not require permission from an employer, healthcare professional, or insurance agency were chosen for assessment.

Figure 2 describes the results from the app search and screening process for the subsequent testing. Our search yielded 26,064 apps. We sequentially screened 1,500 apps. For some mental health categories, we did not reach the target of 50 apps each from Apple and Google store. For example, schizophrenia spectrum and psychosis only had 46 apps from the Google store and 27 from the Apple store, anxiety disorder had 42 apps from the Apple stores, eating disorders had 32 from the Apple store and 36 from the Google store, wake-sleep disorder had 37 apps from the Google store, and personality disorders only had 41 apps from the Apple store and 44 apps from the Google store. In total we screened 1,131 apps. Including the pre-pilot and pilot rounds one to four, we completed dual independent review of 27 apps.



**Figure 2. Apps screening process**



n = number of apps

## Testing and Refinement

Our core team applied the initial draft framework to 10 apps to acquire a better understanding of the ease of using the framework, the relevance of various criteria, the clarity of the instructions and the flow of the framework. After a training session that included review of the draft training guide, each app was assessed by two reviewers and inter-rater reliability was calculated for the risk assessment section of the framework. We used the basic inter-rater reliability measure, which is a percent of agreement between reviewers when the correct response is not known. For criteria where inter-rater reliability was less than 80 percent, and/or based on a discussion with the team about the differences in understanding of the proposed criteria, the framework was revised (Appendix E).

After the pre-pilot round, the framework was further iteratively modified in four pilot rounds of testing by evaluators external to the core team. The changes made after each round are summarized in appendices (Appendix E).

We made the most modifications to the Privacy and Security, App Integrity, and Risk Assessment sections, with 10, 9, and 9 changes, respectively. We also rearranged 20 questions between the pre-pilot and pilot round one in order to create a more sequentially seamless framework.

After pilot round one we made 48 changes. Changes were made to questions under app integrity and privacy and security, with 6 and 8 modifications respectively. During this round, we had 11 modifications for additional guidance and adding or consolidating response options each. We also added 9 questions to help us to further understand certain concepts, for example, “If the app allows for sharing of personal health information (PHI) such as assessments, treatment plan, and messaging with the provider, is it made clear to the user that this data is secured and kept private?” to understand if the use of data is clear for users.

After pilot round two, we had half of the modifications compared to the number we had after pilot round one. Overwhelmingly, our modifications were for adding guidance and adding or consolidating response options, with 8 modifications each. All other types of modifications equaled 8. Lastly, after pilot round 3 we had only 5 modifications to make, with 3 of them for adding or consolidating response options (Table 2).

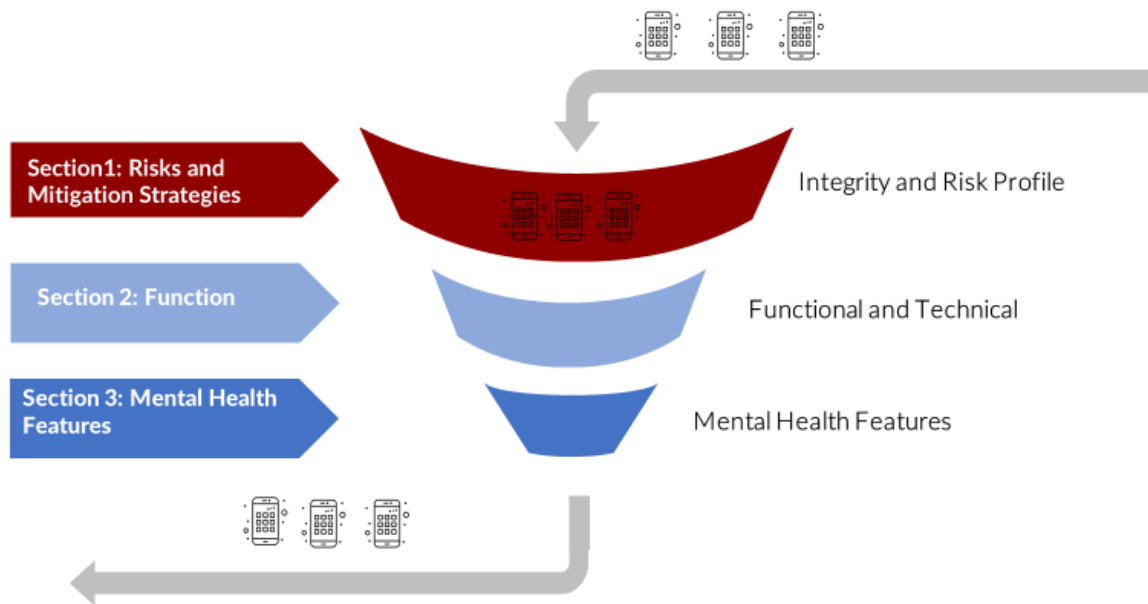
**Table 2. Summary of the number of changes made and inter-rater reliability in each round**

Round	Number of apps reviewed	Average inter-rater reliability (%)	Number of changes to framework
Pre-Pilot	10	--	106
Pilot Round 1	5	68.36	48
Pilot Round 2	5	62.93	24
Pilot Round 3	4	82.38	5
Pilot Round 4	3	87.51	11

# Overview of the Framework to Assist Stakeholders in Technology Evaluation for Recovery (FASTER) to Mental Health and Wellness

The Framework to Assist Stakeholders in Technology Evaluation for Recovery (FASTER) to Mental Health and Wellness comprises three sections with an initial and concluding set of administrative questions: Section 1: Risks and Mitigation Strategies, Section 2: Function, and Section 3: Mental Health App Features (See Appendix F and G). Within each of these sections, there are a series of questions related to the assessment of specific categories that were considered critical based upon the literature review and key informant interviews. The questions are organized in a systematic order so that a reviewer can start the assessment as they search for and download an app from the app store. Section 1: Risks and Mitigation Strategies is uniformly applied to all apps undergoing a review. It facilitates an assessment of the risk profile and integrity of the apps and serves to flag any apps that do not meet basic safety, evidence, and security checks. An app that is flagged for failing Section 1: Risks and Mitigation Strategies may not need to be further assessed. The rationale for the criteria within each of these sections is provided in Figure 3, and brief descriptions and justifications for the criteria are provided in Figure 4.

**Figure 3. Framework to Assist Stakeholders in Technology Evaluation for Recovery (FASTER) to Mental Health and Wellness**



**Figure 4. Categories of Questions in Each Section**

	<b>Risks and Mitigation Strategies</b>	App Integrity	Risk Assessment	Evidence	Linkage to Care	Access to Crisis Resources
	<b>Function</b>	Accessibility Features	App Info	Costs	Developer Credibility	Evidence & Clinical Foundation
		Privacy & Security	Informed eConsent	Cultural Competence	Usability	Remote Monitoring
		Access to Crisis Resources	Artificial Intelligence (AI)			
	<b>Mental Health Features</b>	Mental Health Features				

## Target Audience and Intended Use

The goal of this assessment framework is to support agencies and individuals working in mental health, as well as users of mental health apps, in making informed decisions about using or recommending the use of particular apps. We expect that this framework will be applied by intermediary mental health agencies that have the capacity to train personnel to use this framework to evaluate mental health apps and insurance companies that might have an interest in reimbursing certain health apps. Evaluators who use this framework and accompanying training guide should be individuals with some background in technology and mental health, but it is not required that users of the framework be experts in technology or mental health. The results and summary conclusions of such assessments will be valuable to healthcare professionals as they prescribe apps to patients, and to patients/users/caregivers in search for a mental health support app. The framework might also inform and guide app developers in the development of apps.

The FASTER framework is intended to be applied to assess apps whose primary function is to support mental health and wellness through content and resources within the app. It is not appropriate to use this framework to evaluate apps whose primary function is to facilitate telemedicine (e.g., link users/patients to a mental health professional), or apps that might contain cursory content to support wellness (e.g., a weight loss app that has resources for mindfulness).

The questions in the administrative section precede the formal evaluation of the app and include introductory questions related to the app that is being reviewed: information about the developer, the website, the country of origin, device compatibility (iOS, Android), approval by any regulatory authorities, the app version number, costs, and requirement of authorization for use. These criteria are factual rather than evaluative and aim to gather data on accessibility, currency, and credibility associated with the use of the app. Such information can be important in cataloging and classifying mental health apps.

## Section 1: Risks and Mitigation Strategies

This section aims to assess the integrity and risk profile of the app. This section is evaluative and apps that don't meet the thresholds for risks and integrity are flagged. This section contains only those questions that are used for assessment of the risk and app integrity.

## App Integrity

Questions in the app integrity category assess whether the app can be trusted for use. This is based on whether the app is being updated regularly, based on user feedback and industry trends; whether it provides warnings and disclaimers about use; and whether it gets users to sign a privacy and security agreement if it collects user data. These questions aim to ensure that if users are using an app for mental health or wellness, it is not suddenly pulled off the market, that personal health and financial information is treated appropriately, and that the app has legal commitment to user privacy and security. The criteria also assess whether an app has been endorsed or is being used by a trusted federal agency (e.g., National Institute of Health), or non-government body (e.g., APA) which would reinforce credibility, as these institutions exercise due diligence before endorsing or making the app available to their employees.

The responses to questions around these concepts determine the level of integrity as High or Low, as defined in Table 3.

**Table 3. App integrity levels**

Integrity Level	Requirements
High	If the app has been updated in the previous 6 months, ensures privacy and security of the user's data (or/and provides disclaimers and warnings), and/or if the app has been endorsed by a trusted organization.
Low	If the app has not been updated in the previous 6 months and/or provides no privacy and security statement, and/or provides no disclaimers and warnings.

## Risk Assessment

Questions in the Risk Assessment category assess the risks posed by a mental health app, as evaluated by alignment of the goals of the app, the target audience, and severity of the mental health condition with available evidence to support the approach, oversight and linkage to care, and privacy and security protocols. Linkage to care is defined as linkage to a healthcare provider who can monitor their patient through an interface in the app or through data being linked to their Electronic Medical Record (EMR) system. Additional risk criteria are defined for vulnerable populations.

Alignment between intended goals, target audience, mental health condition, and evidence/safety measures has been adapted for mental health apps from the model proposed by International Medical Device Regulators Forum “Software as a Medical Device (SaMD): Clinical Evaluation.”<sup>59</sup> Assessment of the level of risk posed by the app is based on the primary function of the app (stand-alone treatment vs. other uses such as psychoeducation, wellness, etc.) and whether the app is intended to treat those living with specific mental health diagnoses and/or mental health impairments that substantially impact their ability to function. For example, if the app is used by someone with a clinically diagnosed condition, such as schizophrenia, then the risk posed by the app is higher than if the person has mild anxiety. Similarly, if the app targets children and adolescents, then the potential risk associated with the app may be higher than if it is an app for use by the general population. If the app is used by someone with a mild or moderate level of functional impairment, then the risk is greater than one that can be used by any member of the general population with no functional impairment. Similarly, if an app claims to provide treatment without the support of a healthcare provider then the risks posed may be greater than those posed by a wellness app that is intended to play a supportive function.

Based on these criteria, apps may be classified under three risk levels (Table 4), each of which have specific requirements for evidence and linkage to caregivers or clinical care.

**Table 4. Risk levels**

<b>Risk level</b>	<b>Evidence requirement</b>
Risk Level 1: Minimal Risk	No requirement for providing evidence or for linkage to care. For example, apps aimed at supporting mindfulness practices would fall into this category.
Risk Level 2: Some Risk	Requires some research support regardless of the experimental design. The app should also leverage an evidence-informed theory to guide its approach. Additionally, it should facilitate remote sharing of information with a provider and provide the user with information on a crisis hotline or other resources. For vulnerable populations, the app should require caregiver permission and facilitate sharing of information with them.
Risk Level 3 Considerable Risk	Requires research support with at least one or more <b>randomized controlled trials</b> that show evidence of impact. The app should also leverage an evidence-informed theory to guide its approach. Additionally, it should facilitate remote sharing of information with a provider and provide the user with information to access a crisis hotline or other resources. For vulnerable populations, the app should require caregiver permission and facilitate sharing of information with them.

At the end of this section, the reviewer will be able to determine whether the app clears the required credibility and risk threshold.

## **Evidence**

Questions in the Evidence category help determine whether the app has a solid clinical evidence foundation. The greater the risk of an app, the greater the burden of evidence. For apps that pose a higher level of risk, the framework requires that there are robust studies assessing the efficacy and risks posed by the apps in order for the app to clear the safety screening.

## **Linkage to Care**

Questions in the Linkage to Care category evaluate the linkages to a healthcare provider who can monitor their patient through an interface in the app or through data being linked to their EMR system. If the app poses a higher level of risk, the framework requires that it also provides resources for linkage to care in order for the app to clear the safety screening.

## **Access to Crisis Services**

Questions in the Access to Crisis Services category evaluate whether the app provides access to emergency sources of information.

## **Section 2: Function**

This section is focused on descriptive aspects related to accessibility, costs, developer credibility, evidence and clinical foundation, privacy/security, usability, functions for remote monitoring of the user, informed consent, cultural competency, access to crisis services, and AI. These criteria are intended to facilitate (1) systematic cataloguing of the functions of the app, so that users may choose an app based on the functionality, (2) describe features offered by the app for users to assess its fit for their needs, and (3) assess how the app may help individuals with a mental health diagnosis.

## **Accessibility Features**

Accessibility features incorporated within an app should facilitate easier use of the technology by individuals with disabilities. Common accessibility features include text-to-speech, closed-captioning, and keyboard shortcuts.

## **App Information**

This section captures details about the platform required by the app (e.g., iOS, Android), and users' perspectives of the apps through the number of reviews and ratings provided by users.

## **Costs**

Increasingly, apps have complex pricing models which, especially in the case of a vulnerable user base with mental health impairments, may pose risks. Any costs associated with the app should be provided upfront. The criteria in the Costs category classify the costs as monthly/annual subscription or as freemium services that require in-app purchasing. Freemium refers to a pricing strategy by which a basic product or service is provided free of charge, but there are charges for additional features or services that expand the functionality of the free version of the app. Many users of mental health apps may not know that freemium apps may require in-app purchases in order to access the full range of functionality provided by the app. Additionally, some apps require a payment from the user, while others are reimbursed by the healthcare provider or through insurance.

## **Organizational Credibility**

The criteria in the Organizational Credibility category assess any documented complaints against the app developers, as well as the organizational health of the app developing organization or company.

## **Evidence & Clinical Foundation**

The questions related to evidence in this category go beyond what was assessed in Section 1. In this category, the reviewer assesses alignment of the content with the claims made by the app, whether the clinical workflows are rooted in evidence and best-practices, and the clarity of the content. Additionally, evidence on the duration of use of the app is also assessed. While some apps are intended to be used for transition periods only, often sustained engagement with apps over time is helpful. This continues to be a significant challenge in the use of apps for healthcare.<sup>60, 61</sup>

## **Privacy/Security**

Security is about the safeguarding of data, including features that provide protection against unauthorized access to data. Privacy is about the safeguarding of the identity of the user. Given the stigma that may be associated with some mental health conditions, privacy of the user is of particular importance for an app that may be developed for vulnerable populations. The criteria to assess privacy and security focus on whether there is transparency about how user data are used for research, quality improvement or commercial purposes, whether identified or de-identified data are shared or sold to other vendors, how the data are stored, whether user data can be deleted in their entirety (including from group posts), and whether any claims of Health Insurance Portability and Accountability Act (HIPAA) or other analogous national standards for protected health information (PHI) are made.

## **Informed eConsent**

Informed consent is a process for getting permission before conducting some form of research using health data or prior to disclosing the users' health and related information. Most apps have a disclosure list that is long and hard to understand. There are best practices for

ensuring that users actually understand what, exactly, they are agreeing to before they “Agree” to the privacy and security practices. The questions in the Informed eConsent category evaluate whether the app follows these best practices.

## **Cultural Competence**

Cultural competence is defined as the ability to understand, appreciate, and account for different cultures or belief systems based on race, ethnicity, income strata, religious beliefs, etc. The criteria in the Cultural Competence category assess whether the app is targeted at, or inclusive of, specific population groups and cultures. If the app is targeted at a specific cultural group, the criteria assess whether the app was tested in that group. The criteria also assess the use of gender inclusive language, and evidence of effectiveness in a non-white population.

## **Usability**

Usability can be described as the capacity of a system to provide a condition for its users to perform tasks safely, effectively, and efficiently. It is important that the user experience be engaging and pleasing, otherwise users are likely to stop using the app. Usability assessments are challenging given the limitations in objectivity, which can be focused on metrics such as font size, etc.

## **Functions for Remote Monitoring of the User**

Remote patient monitoring is a technology to enable monitoring of patients outside of conventional clinical settings, such as in the home or in a remote area, which increases access to care and decreases healthcare delivery costs. For mental health apps, the provider may receive an alert about their patient's health, or they may be able to access the patient's health indicators from within the app. To enable remote monitoring, apps need to adhere to established data standards for interoperability to safely exchange health data, including with wearable devices that may be used to monitor vital parameters or behaviors.

## **Access to Crisis Services**

Apps can either provide information or place users in direct contact (24 hours a day, 7 days a week) with a trained counselor from the Crisis Text Line or the National Suicide Prevention Lifeline, as well as generate a 911 alert.

## **Artificial Intelligence**

Apps often use or claim to use AI for a variety of tasks including automation, problem solving, and prediction. Apps may claim to use AI when they are just using a rule-based system. The questions in the AI category are important because we want to gauge the potential for the apps to cause harm and also to determine whether apps are updating algorithms based on user input.

## **Section 3: Mental Health App Features**

This is a specialized section where the questions may only apply to specific apps. This section aims to match the needs of the user with the specific mental health features certain apps provide. Questions in this section ask what mental health symptom or condition the app addresses, as well as the primary function of the app [e.g., wellness (mindfulness/meditation/relaxation, psychoeducation); skills training; symptom



tracking/monitoring; social support]. The specific mental health features that apps may include are facilitating social interaction, motivation enhancement, planning/alternative strategies/planning for high-risk situations, screening, self-help, skill building, safety planning, and promoting sleep hygiene. The criteria in this section are intended to facilitate the cataloging of specific features and goal-oriented or skill-building resources (e.g., sleep diaries, tracking lifestyle, monitoring mood).

## **Post-Administrative**

The Post-Administrative Questionnaire solicits details about the app and any training that was available and required for using the app. Links to all the evidence analyzed as part of the review are documented here. It also affords an opportunity for the reviewer to provide a subjective evaluation of the app.

## **Summary of Changes Made to the Framework During Pilot Testing Rounds**

### **Development of a Pre-assessment Questionnaire**

Several apps that made it past the initial screening as mental health apps, were focused entirely on telemedicine, or provided features such as physical fitness tracking with a small component that focused on mindfulness and mental health. The FASTER framework is aimed at evaluating mental health apps. A pre-assessment questionnaire was developed to guide the evaluator on whether the app being evaluated has a mental health focus.

### **Endorsements and Usage**

To assess app integrity, one of the items asked whether the app has been endorsed by a government institution (e.g., Veterans Affairs, and Centers for Medicaid and Medicare Services) or trusted mental health or psychiatric associations (e.g., American Psychological Association, Substance Abuse and Mental Health Services Administration, and National Alliance on Mental Illness). Operationalizing this question proved to be inconsistent during the pilot testing. Given the legal liabilities with endorsement, very few apps currently have an endorsement. There are many apps, such as the Stanley-Brown Safety Plan and Calm, that are *used* by employees at academic institutions and government agencies but not *endorsed* by any agency. We modified the question to include both endorsement and usage since it captures the intent of the question. By *usage* we mean that the app may have been made available to employees of the organization. However, we also distinguish that apps do not meet the integrity bar if they are *developed* by (and not used or endorsed by) governments and academic institutions, given that several apps developed within academic settings at often testing in smaller populations and may not have the funding and institutional backing to scale and be sustainably available to populations.

### **Assessment of App Risk Based on “Recommended Age”**

In the initial rounds of pilot testing, we had lower interrater reliability due to mis-categorization of risk based on the age groups targeted by the apps. The age at which an app can be used without parent’s consent varies greatly amongst apps intended for users 13 to 18 years of age. In the pilot testing, we found that the App Store rating for recommended age often differed

from the permissible age for use listed within the app once it was downloaded. For example, the Calm paid app, which has a Calm for Kids option, states 4+ in the Apple App Store, yet the Terms of Service say that users over 13 “are not barred from using the Service”, yet to purchase the app, users “must be 18 years or older and capable of forming a binding contract.” Further research showed that the app store rating just indicated the appropriateness of content for a specific age group.<sup>62</sup> To standardize the evaluation of what age group the apps are targeted towards, we concluded that the Terms of Use, rather than the App Store, is a better source for understanding the app’s intended target populations.

## **“Special” to “Vulnerable” Populations**

After the pilot round, we changed the term “special population” to “vulnerable population” and elaborated more on our definition of vulnerable populations. We made a distinction between “can the app be used” by these populations to “is the app intended for use” by these populations, highlighting that anyone with access to appropriate technology can use the app, but the intention of this question is to assess whether the app targets a particular group (and if so, does it do so with appropriate checks and balances). Details of the group the app targets is typically listed within the app or its website, and therefore, is more amenable to standard interpretation.

## **Privacy Policy and Terms of Agreement**

There is a lack of standardization on the terminology for warnings, disclaimers, security and privacy agreement, and sharing of data. These are distributed under “Terms of Use”, “Terms and Conditions”, “Privacy and security agreement” and other agreements if at all they are provided by the app.<sup>63</sup> The inconsistency in the location of this information required adding additional guidance for the user of the framework. An additional challenge was that sometimes this information was not possible to find in the app itself but could be found on the app website. We chose not to include information found on the app website that was not also in the app because most app users are not going to the website to look up these types of agreements before using the app.

## **Cultural Competence**

Through pilot testing, we broadened the parameters for “cultural competence” to include questions related to groups with lived experiences such as pregnant teens and survivors of gender-based violence. We added clarity around gender inclusivity. We also added a question on whether the app was tested across different ethnic groups and to what level of representativeness any feasibility/efficacy studies of the app included diverse cultural groups.

## **Changes to Usability Criteria**

Of the 20 questions in the uMARS framework, we used 10 questions (Appendix C). One of these questions was adapted: “Quality of information: Is app content correct, well written, and relevant to the goal/topic of the app?” was changed to “Is the app content, well written, and relevant to the goal and/or topic of the app?”

We observed that the lack of interrater reliability on these questions was due to differential the two extremes- i.e., evaluators picked either "good", or “very good”; or they chose either “bad” or “very bad”. They were typically on the same side of the neutral choice.). We made changes to the language used in the responses - in general, replacing the neutral terms such as

“moderate” and “ok” with “satisfactory”.. Therefore, the five response categories was consolidated to 3 during tabulation for calculation of inter-item agreement with the consolidation of the two most positive responses and the two most negative responses.

## **Interaction with a Coach or Healthcare Professional**

An item added based on pilot testing was interaction with a live coach, or the ability to engage with a person via text, voice, video, etc.

## **Categorization of Mental Health Features**

The categorization of mental health features was on a 3 point scale – 1) basic 2) moderate 3) comprehensive. We found that it was hard to standardize the definitions for each of these and reviewers’ agreement on what functionality they considered “basic”, “moderate” and “comprehensive” varied, specifically when differentiating between basic and moderate. To improve inter-rater reliability, we redefined the feature on a two-point scale – 1) not comprehensive 2) comprehensive.

## **Acceptable Research**

Most apps do not have randomized clinical trials to support their claims of efficacy. Given the innovative and dynamic nature of the field of digital health, we wanted to balance the requirements from a safety perspective while widening the acceptable scientifically validated evidence to evaluate efficacy and effectiveness. Hence, we added a response on research support to see if the app included at least one published paper in a peer-reviewed journal that used single-case design or quasi-experimental methods to demonstrate efficacy.

## **Subjectivity Versus Objectivity in Question Responses**

One of the guiding principles for developing the framework was to restrict the questions and response categories to options that can be objectively assessed. Objective information, while observable, quantifiable, and provable, may not convey quality to the same extent as subjective responses that provide a more nuanced evaluation, although biased by the evaluator's beliefs. For example, it is important for the user to know “if the app had advertising that was intrusive and distracting” as well as “if the app had any advertising”. If advertising is a way to support further development of an otherwise free app, it may be acceptable if it isn’t intrusive, and this leaves the user to be able to weigh the pros and cons of the app. We decided to retain some subjectivity for questions with higher inter-rater reliability but either consolidated responses or changed to binary for those with low inter-rater reliability.

## Summary and Implications

Mental health mobile apps can fill a major gap in mental health services. Apps can also alert the family, support network, or health care provider to the need for more intensive mental health or crisis resources. There is appeal to the idea of being able to take a more personalized medicine approach to mental health apps (i.e., will a specific app work for a specific client). At present, the evidence base is not very strong and the potential for risks are not non-zero for many mental health apps. This field is rapidly evolving, and the COVID-19 pandemic has elucidated the critically important role technology can play to support wellbeing and mental health.

FASTER to Mental Health and Wellness is aimed at facilitating the use of apps for mental health support and recovery through standardized evaluation, screening, and classification of apps. Several of the criteria have been extracted from existing frameworks in the app evaluation and mental health space. However, we identified several gaps in the existing frameworks and addressed these through further prioritization of criteria, the addition of criteria to assess risks and safety of the apps, and assessment of the use of AI and other engagement approaches. Prior frameworks to evaluate health apps have been largely technical, focused on the evaluation of the technology specifications, rather than the use of the apps in specific healthcare domains. In contrast, FASTER to Mental Health and Wellness is structured to facilitate an initial screening of apps to align the purpose of, and the mental health condition targeted by, an app with the possible risk and overall risk mitigation features (like foundational evidence, access to crisis hotlines) to determine overall safety of an app given the level of evidence and risk mitigation protocols in place. Additionally, FASTER to Mental Health and Wellness includes some guidance to address the vast and rapidly expanding use of AI in apps targeted at mental health.

## Next Steps

We envisage a range of possible applications of this framework. First, the framework can be used by mental health organizations and mental health advocacy agencies that provide mental health resources to provide a curated and validated mental health app library. Such libraries could be consumer/patient-facing with recommendations tailored to the mental health condition or the type of skill/resource for which the user is seeking assistance. Such a consumer/patient-facing library with recommendations based on the framework could potentially also be used by consumers, family members, peer supports, and health care providers to select apps. Updating and maintaining such a library would necessitate substantial planning and resources for routine curation and screening of apps using the framework. Second, the framework can be leveraged by employee health plans, health system leaders, public and private insurance providers, and other entities like APA to review apps relevant to their members and provide guidance and advice that is specific to mental health. A federal agency or other trusted mental health organization could host, automate, update, and disseminate the proposed framework. Lastly, the framework can be used by app developers as guidance to promote transparency in communication about the potential benefits of, risks from, and evidence for the app.

Assessment and standardization of mental health apps poses some unique challenges that we anticipate will continue to require attention. Assessment of the risks that an app poses to individuals with mental health conditions is challenging. An individual's mental health condition can change quickly which changes the potential risks. We accounted for this in the framework by the inclusion of criteria on appropriate linkage to a provider and other caregivers. Mental health symptoms are transdiagnostic, and typically apps may aim to support a symptom rather than the disease. Section 1: Risks and Mitigation Strategies of the framework assesses risks based on the

type of mental health condition and the aim of the app (i.e., stand-alone treatment vs. supportive/adjunctive care). However, several mental health apps may aim to target symptoms such as anxiety or insomnia which are common across several mental health conditions. Further refinement of the framework may be needed to address applicability across apps that target transdiagnostic symptoms. Additional criteria may be needed to account for potential harm or iatrogenic impacts of an app, based on the severity or other characteristics of specific mental health conditions or which pertain to specific situations, such as a new diagnosis of a mental health condition where acuity may be unclear. As our knowledge of how apps can be effectively used to address mental health conditions grows, this framework would need to be updated to reflect that knowledge. The framework presents considerations that can be generalized for most mental health apps; as such, it does not facilitate a comprehensive assessment of apps for a specific mental health condition. We acknowledge that apps targeted at specific disorders may benefit from an assessment that is quite specific for those types of apps; however, we also recognize that developing such disease area specific frameworks is time and resource-intensive and is unlikely to be a practical approach. Lastly, the framework results in summary conclusions on various aspects of the app such as usability, security, etc. As a next step to facilitating adaptation and adoption, it might be valuable to gather user (e.g., providers, patients, etc.) input on the value of these summative conclusions in guiding decisions around the use of the apps.

We acknowledge that most patients and providers may not be able to review a detailed assessment report on the utility of a certain app and might find a simple “go/no-go” recommendation easier to process. FASTER does not support the development of a nuanced report on the apps and does not provide an explicit recommendation on whether an app should be used. As new governance and regulations for software as a medical device are formulated, the framework should be adapted to include those. Similarly, we recognize that there are ongoing developments in our understanding of what prerequisites for apps from a privacy/security perspective, as well as rapid innovation in the digital health and AI space. We expect that this framework will need to be updated routinely to reflect these areas of growth, especially those specific to AI. In future versions of the framework, it would be important to add greater input from commercial app developers as they can provide insight regarding the app roadmap and challenges in commercializing health apps. Future application, testing, and refinements to this framework may be required to determine its suitability and reliability across multiple mental health conditions, as well as to account for the rapidly expanding applications of AI in mental health apps. Additionally, adaptations to the framework may be needed to account for the privacy and security of user information in regulatory contexts. Ultimately, to facilitate the adoption and sustenance of this framework, it would be necessary to have a system that can train personnel to apply this framework, and screen apps. The results of the review of apps using this framework would ideally be hosted as an interactive webpage that can be used by patients and mental health advocacy agencies. Potential stakeholders to provide such support for the framework include healthcare systems, employee assistance programs, colleges/universities, SAMHSA, American Psychiatric Association, American Psychological Association, National Alliance on Mental Illness (NAMI), Mental Health America, [mentalhealth.gov](http://mentalhealth.gov), HelpGuide, World Health Organization, Anxiety and Depression Association of America (ADAA), Depression and Bipolar Support Alliance (DBSA), International OCD Foundation, National Eating Disorders Association, PTSD Alliance, Schizophrenia and Related Disorders Alliance of America, Treatment Advocacy Center, Active Minds, the Child Mind Institute, American Foundation for Suicide Prevention, and the American Association of Suicidology.

## References

1. Substance Abuse and Mental Health Services Administration. (2020). Key substance use and mental health indicators in the United States: Results from the 2019 National Survey on Drug Use and Health (HHS Publication No. PEP20-07-01-001, NSDUH Series H-55). Rockville, MD: Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. Retrieved from <https://www.samhsa.gov/data/>.
2. Carbonell Á, Navarro-Pérez JJ, Mestre MV. Challenges and barriers in mental healthcare systems and their impact on the family: A systematic integrative review. *Health Soc Care Community*. 2020 Sep;28(5):1366-79. doi: 10.1111/hsc.12968. PMID: 32115797.
3. Mental health: strengthening our response. <https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response>.
4. Mantani A, Kato T, Furukawa TA, et al. Smartphone Cognitive Behavioral Therapy as an Adjunct to Pharmacotherapy for Refractory Depression: Randomized Controlled Trial. *J Med Internet Res*. 2017 Nov 3;19(11):e373. doi: 10.2196/jmir.8602. PMID: 29101095.
5. Firth J, Torous J, Nicholas J, 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):287-98. doi: 10.1002/wps.20472. PMID: 28941113.
6. Firth J, Torous J, Nicholas J, et al. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. *J Affect Disord*. 2017 Aug 15;218:15-22. doi: 10.1016/j.jad.2017.04.046. PMID: 28456072.
7. Price M, Yuen EK, Goetter EM, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. *Clin Psychol Psychother*. 2014 Sep-Oct;21(5):427-36. doi: 10.1002/cpp.1855. PMID: 23918764.
8. Powell AC, Torous JB, Firth J, et al. Generating value with mental health apps. *BJPsych Open*. 2020 Feb 5;6(2):e16. doi: 10.1192/bjo.2019.98. PMID: 32019619.
9. Rodriguez-Villa E, Rauseo-Ricupero N, Camacho E, et al. The digital clinic: Implementing technology and augmenting care for mental health. *General Hospital Psychiatry*. 2020 2020/09/01;66:59-66. doi: <https://doi.org/10.1016/j.genhosppsych.2020.06.009>.
10. Bridges KM. Implicit Bias and Racial Disparities in Health Care. [https://www.americanbar.org/groups/crsj/publications/human\\_rights\\_magazine\\_home/the-state-of-healthcare-in-the-united-states/racial-disparities-in-health-care/](https://www.americanbar.org/groups/crsj/publications/human_rights_magazine_home/the-state-of-healthcare-in-the-united-states/racial-disparities-in-health-care/). 2021.
11. Manuel JI. Racial/Ethnic and Gender Disparities in Health Care Use and Access. *Health Serv Res*. 2018 Jun;53(3):1407-29. doi: 10.1111/1475-6773.12705. PMID: 28480588.
12. Sealy-Jefferson S, Vickers J, Elam A, et al. Racial and Ethnic Health Disparities and the Affordable Care Act: a Status Update. *J Racial Ethn Health Disparities*. 2015 Dec;2(4):583-8. doi: 10.1007/s40615-015-0113-z. PMID: 26668787.
13. Hahn HC, Cook BL, Ault-Brutus A, et al. Intersection of race-ethnicity and gender in depression care: screening, access, and minimally adequate treatment. *Psychiatr Serv*. 2015 Mar 1;66(3):258-64. doi: 10.1176/appi.ps.201400116. PMID: 25727113.
14. Health Resources and Services Administration. Designated health professional shortage areas statistics. Bureau of Health Workforce. U.S. Department of Health and Human Services [Internet], 2019. ([https://ersrs.hrsa.gov/ReportServer/?HGDW\\_Reports/](https://ersrs.hrsa.gov/ReportServer/?HGDW_Reports/)). 2019.

15. Wang K, Varma DS, Prosperi M. A systematic review of the effectiveness of mobile apps for monitoring and management of mental health symptoms or disorders. *J Psychiatr Res.* 2018 Dec;107:73-8. doi: 10.1016/j.jpsychires.2018.10.006. PMID: 30347316.
16. Wang E, Zweig M. A defining moment for digital behavioral health: Four market trends | Rock Health. <https://rockhealth.com/insights/a-defining-moment-for-digital-behavioral-health-four-market-trends/>. 2021.
17. Mercom. Mental Health Apps Raised \$1.2 Billion in 2020. @mercom. <https://mercomcapital.com/mental-health-apps-raised/>. 2021.
18. Need mental health help? There are apps for that, but picking the right one is tough. <https://www.latimes.com/california/story/2021-06-21/mental-health-apps-consumer-challenge-picking-the-right-one>. 2021.
19. Weir K. The ascent of digital therapies. *Monitor on Psychology*, 49(10). <http://www.apa.org/monitor/2018/11/cover-digital-therapies>. 2018.
20. U.S. Food & Drug Administration. Digital Health Software Precertification (Pre-Cert) Program. <https://www.fda.gov/medical-devices/digital-health-center-excellence/digital-health-software-precertification-pre-cert-program>.
21. Software as a Medical Device (SaMD). FDA. Published September 9, 2020. Accessed May 11, 2021. <https://www.fda.gov/medical-devices/digital-health-center-excellence/software-medical-device-samd>.
22. Bergin A, Davies EB. Technology Matters: Mental health apps - separating the wheat from the chaff. *Child Adolesc Ment Health.* 2020 Feb;25(1):51-3. doi: 10.1111/camh.12363. PMID: 32063750.
23. Martinengo L, Van Galen L, Lum E, et al. Suicide prevention and depression apps' suicide risk assessment and management: a systematic assessment of adherence to clinical guidelines. *BMC Med.* 2019 Dec 19;17(1):231. doi: 10.1186/s12916-019-1461-z. PMID: 31852455.
24. Dehling T, Gao F, Schneider S, et al. Exploring the Far Side of Mobile Health: Information Security and Privacy of Mobile Health Apps on iOS and Android. *JMIR Mhealth Uhealth.* 2015 Jan 19;3(1):e8. doi: 10.2196/mhealth.3672. PMID: 25599627.
25. App Guide | PsyberGuide. One Mind PsyberGuide. Accessed May 11, 2021. <https://onemindpsyberguide.org/apps/>.
26. APA App Advisor. Accessed May 11, 2021. <https://www.psychiatry.org/psychiatrists/practice/mental-health-apps>.
27. How to pick employee mental health apps | Kaiser Permanente. Accessed May 11, 2021. <https://business.kaiserpermanente.org/insights/mental-health-workplace/mental-health-apps-workforce-wellness>.
28. Know more. Live brighter. Verywell Mind. Accessed May 11, 2021. <https://www.verywellmind.com/>.
29. New Zealand health information | Health Navigator NZ. Accessed May 11, 2021. <https://www.healthnavigator.org.nz/>.
30. Methodology | myhealthapps.net. Accessed May 11, 2021. <http://myhealthapps.net/methodology>.
31. Healthy Living Apps Guide. Accessed May 11, 2021. <http://www.vichealth.vic.gov.au/media-and-resources/vichealth-apps/healthy-living-apps>.

32. Digital Technology Assessment Criteria (DTAC). NHSX. Accessed May 11, 2021. <https://www.nhs.uk/key-tools-and-info/digital-technology-assessment-criteria-dtac/>.
33. Policymakers. Digital Therapeutics Alliance. Accessed May 11, 2021. <http://dtxalliance.org/value-of-dtx/policymakers/>.
34. Lagan S, Sandler L, Torous J. Evaluating evaluation frameworks: a scoping review of frameworks for assessing health apps. *BMJ Open*. 2021 Mar 19;11(3):e047001. doi: 10.1136/bmjopen-2020-047001. PMID: 33741674.
35. Stoyanov SR, Hides L, Kavanagh DJ, et al. Development and Validation of the User Version of the Mobile Application Rating Scale (uMARS). *JMIR Mhealth Uhealth*. 2016 Jun 10;4(2):e72. doi: 10.2196/mhealth.5849. PMID: 27287964.
36. BfArM - Germany Federal Institute for Drugs and Medical Devices. <https://www.emergobyul.com/resources/europe/germany-federal-institute-drugs-medical-devices>.
37. NATIONAL INSTITUTE FOR HEALTH AND CARE EXCELLENCE EVIDENCE STANDARDS FRAMEWORK FOR DIGITAL HEALTH TECHNOLOGIES. <https://www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/evidence-standards-framework/digital-evidence-standards-framework.pdf>. 2019.
38. Evidence standards framework for digital health technologies | Our programmes | What we do | About. NICE. Accessed May 11, 2021. <https://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standards-framework-for-digital-health-technologies>.
39. Policy for Device Software Functions and Mobile Medical Applications - Guidance for Industry and Food and Drug Administration Staff. :45.
40. Clinical Decision Support Software. U.S. Food and Drug Administration. Published May 6, 2020. Accessed May 11, 2021. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-decision-support-software>.
41. Clinical Decision Support Software. Draft Guidance for Industry and Food and Drug Administration Staff. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-decision-support-software>.
42. IMDRF -International Medical Device Regulators Forum. Software as a Medical Device": Possible Framework for Risk Categorization and Corresponding Considerations. <http://www.imdrf.org/docs/imdrf/final/technical/imdrf-tech-140918-samd-framework-risk-categorization-141013.pdf> 2014.
43. Kahn, J., & Johns Hopkins Project on Ethics and Governance of Digital Contact Tracing Technologies (2020). *Digital Contact Tracing for Pandemic Response: Ethics and Governance Guidance*. Baltimore: Johns Hopkins University Press., doi:10.1353/book.75831.
44. Doerr M, Suver C, Wilbanks J. Developing a Transparent, Participant-Navigated Electronic Informed Consent for Mobile-Mediated Research . Available at SSRN: <https://ssrn.com/abstract=2769129> or <http://dx.doi.org/10.2139/ssrn.2769129>. 2016.
45. van Haasteren A, Gille F, Fadda M, et al. Development of the mHealth App Trustworthiness checklist. *Digit Health*. 2019 Jan-Dec;5:2055207619886463. doi: 10.1177/2055207619886463. PMID: 31803490.
46. SAMHSA -Substance Abuse and Mental Health Services Administration. Resources on Cultural Competency <https://www.samhsa.gov/section-223/cultural-competency/resources>.



47. Substance Abuse and Mental Health Services Administration. Improving Cultural Competence. Treatment Improvement Protocol (TIP) Series No. 59. HHS Publication No. (SMA) 14-4849. Rockville, MD: Substance Abuse and Mental Health Services Administration, 2014. . 2014.
48. Celik H, Lagro-Janssen TA, Widdershoven GG, et al. Bringing gender sensitivity into healthcare practice: a systematic review. *Patient Educ Couns.* 2011 Aug;84(2):143-9. doi: 10.1016/j.pec.2010.07.016. PMID: 20719461.
49. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. *Ann Behav Med.* 2013 Aug;46(1):81-95. doi: 10.1007/s12160-013-9486-6. PMID: 23512568.
50. Riley WT, Rivera DE, Atienza AA, et al. Health behavior models in the age of mobile interventions: are our theories up to the task? *Transl Behav Med.* 2011 Mar;1(1):53-71. doi: 10.1007/s13142-011-0021-7. PMID: 21796270.
51. Riley WT, Martin CA, Rivera DE, et al. Development of a dynamic computational model of social cognitive theory. *Transl Behav Med.* 2016 Dec;6(4):483-95. doi: 10.1007/s13142-015-0356-6. PMID: 27848208.
52. Webb TL, Joseph J, Yardley L, et al. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. *J Med Internet Res.* 2010 Feb 17;12(1):e4. doi: 10.2196/jmir.1376. PMID: 20164043.
53. Ferreira S, Pêgo JM, Morgado P. The efficacy of biofeedback approaches for obsessive-compulsive and related disorders: A systematic review and meta-analysis. *Psychiatry Res.* 2019 Feb;272:237-45. doi: 10.1016/j.psychres.2018.12.096. PMID: 30590278.
54. Sander LB, Schorndanner J, Terhorst Y, et al. 'Help for trauma from the app stores?' A systematic review and standardised rating of apps for Post-Traumatic Stress Disorder (PTSD). *Eur J Psychotraumatol.* 2020;11(1):1701788. doi: 10.1080/20008198.2019.1701788. PMID: 32002136.
55. Aji M, Gordon C, Peters D, et al. Exploring User Needs and Preferences for Mobile Apps for Sleep Disturbance: Mixed Methods Study. *JMIR Ment Health.* 2019 May 24;6(5):e13895. doi: 10.2196/13895. PMID: 31127714.
56. Guo Y, Yang F, Hu F, et al. Existing Mobile Phone Apps for Self-Care Management of People With Alzheimer Disease and Related Dementias: Systematic Analysis. *JMIR Aging.* 2020 Jan 24;3(1):e15290. doi: 10.2196/15290. PMID: 32012045.
57. AARP -Accessibility Tools on Your Smartphone That Can Make Your Life Easier. <https://www.aarp.org/home-family/personal-technology/info-2020/smartphone-accessibility.html>.
58. Gerke S, Minssen T, Cohen G. Chapter 12 - Ethical and legal challenges of artificial intelligence-driven healthcare. In: Bohr A, Memarzadeh K, eds. *Artificial Intelligence in Healthcare.* Academic Press; 2020:295-336.
59. Software as a Medical Device (SaMD): Clinical Evaluation. Accessed June 8, 2021. <http://www.imdrf.org/consultations/cons-samd-ce.asp>.
60. Melcher J, Camacho E, Lagan S, et al. College student engagement with mental health apps: analysis of barriers to sustained use. *J Am Coll Health.* 2020 Oct 13:1-7. doi: 10.1080/07448481.2020.1825225. PMID: 33048626.

61. Li J, Chang X. Improving mobile health apps usage: a quantitative study on mPower data of Parkinson's disease. *Information Technology & People*. 2021;34(1):399-420. doi: 10.1108/ITP-07-2019-0366.
62. Mobile software content rating system.  
[https://en.wikipedia.org/wiki/Mobile\\_software\\_content\\_rating\\_system](https://en.wikipedia.org/wiki/Mobile_software_content_rating_system).
63. Robillard JM, Feng TL, Sporn AB, et al. Availability, readability, and content of privacy policies and terms of agreements of mental health apps. *Internet Interventions*. 2019 2019/09/01;17:100243. doi: <https://doi.org/10.1016/j.invent.2019.100243>.