HOMEWOOD RESEARCH INSTITUTE

A Framework for Evaluation of Mobile Apps for Youth Mental Health

May 2020

Prepared by:

Yuri Quintana, PhD,
Chief, Division of Clinical Informatics,
Beth Israel Deaconess Medical Center,
Assistant Professor of Medicine,
Harvard Medical School

John Torous, MD,
Chief, Division of Digital Psychiatry,
Instructor in Psychiatry,
Harvard Medical School

Prepared for Homewood Research Institute

Funded by the RBC Foundation

Funding for this report has been generously provided by:

The RBC Foundation

Homewood Research Institute (HRI) is a registered Canadian Charity dedicated to transforming mental health and addiction treatment through research. We partner with leading scientists, universities, patients and clinicians to improve care, services and outcomes. HRI's charitable registration is # 86307 3334 RR0001.

For more information about this report, please contact:

Roy Cameron, PhD, FCAHS
Executive Director
Homewood Research Institute
150 Delhi Street, Riverslea Building
Guelph, ON N1E 6K9
admin@homewoodresearch.org

Suggested Citation: Quintana, Y & Torous, J. A Framework for Evaluation of Mobile Apps for Youth Mental Health. Homewood Research Institute Report, Guelph, Ontario, Canada, May 20,2020.

Note: The opinions and recommendations included in this report are those of the authors, and do not necessarily reflect the views of the funder.

This report can be accessed free of charge at Homewoodresearch.org

Please Consider Making A Donation.
With your support, HRI can continue to conduct research and evaluation in pursuit of our vision of a world where.... *No life is held back or cut short by mental illness and addiction*

Preamble

Homewood Research Institute (HRI) is a nationally registered Canadian Charity dedicated to transforming mental health and addiction treatment through applied research, evaluation, and innovation. We work with Canadian and international scientists and collaborate with people who receive and provide treatment, to improve services and outcomes.

HRI seeks to address issues that can have a large impact across Canada and beyond. The use of technology to deliver treatment services is enormously important to provide broad access to timely, affordable support to those who need it, regardless of where they live.

One of HRI's goals is to help create an environment where trustworthy digital technologies can be used within formal service provision systems. This report represents a key step toward that goal. It presents a Framework that a) identifies critical issues that must be addressed in designing and evaluating apps, and b) outlines criteria and protocols for rigorously evaluating these issues to generate the evidence needed by those looking for credible apps. This type of Framework is novel and groundbreaking.

We envision the Framework being valuable to those who design, scientifically evaluate, use, or invest in mental health apps and to those who are concerned with developing regulations related to such apps.

The approach is not dogmatic but pragmatic. The authors anticipate that "the Framework will evolve over time as it is applied to more apps. This experience will enable a clearer definition of the criteria, refinement of methods for evaluating the criteria, and the introduction of new criteria as the Framework becomes more widely used." This Framework was informed by a previous project entitled Youth Mental Health Apps in the Digital Age: A Scoping Review of Trends and Evaluations. The report of that earlier project is available on HRI's website[1].

We are grateful to the RBC Foundation for its foresight in making digital technologies a focus of their initiative in promoting youth mental health and for their generous support for this project. HRI is eager to partner with like-minded organizations that place value on creating an environment where high-quality technology tools, backed by strong science, can be employed with confidence within service delivery systems. By aligning our efforts, we will accelerate progress and enhance our collective impact in improving outcomes at individual and population levels.

Please get in touch if you see opportunities for collaboration.

Roy Cameron, PhD, FCAHS

Executive Director
Homewood Research Institute

[1] Homewood Research Institute. Youth Mental Health Apps in the Digital Age: A Scoping Review of Trends and Evaluations. Available at URL: https://homewoodresearch.org/

Executive Summary

Lack of access to mental health and addiction services is a problem worldwide. Technologies such as mobile apps hold great potential for addressing the tremendous unmet need in a scalable, affordable manner and reaching people in remote locations. While there is a rapidly growing number of mobile apps for youth mental health, few have been rigorously assessed or scientifically validated. There is a need for robust evidence to guide decisions about which apps are safe and effective to use, and specifically with which populations, and for what purpose.

Several evaluation frameworks or guidelines exist to guide those seeking high-quality apps for personal use or to adopt in their clinical practice. However, most of these frameworks do not identify a reliable method for making unbiased evaluations. Further, reviews of specific apps are outdated by the time they are published. There is a dearth of such evidence that can be used with confidence to guide those seeking apps.

The premise of this report is that there is an urgent need for a framework that lays out criteria for evaluating apps to generate the evidence needed to promote the use of apps that are safe and effective within service delivery systems. A framework is presented that lays out criteria against which apps should be judged, along with protocols for rigorously evaluating the extent to which apps meet these criteria.

In addressing these issues, the report focuses on youth mental health, especially in the Canadian context. This Framework will also be relevant to other populations and contexts using digital technologies for health. This report is well-aligned with federal and provincial priorities in youth mental health, and the need to provide safe and effective programs that optimize the use of digital technologies.

Contents

1.0 Introduction

There is a critical need for innovative solutions for delivering youth mental health services [1-13]. Scaling up face to face treatment is not feasible, given the enormous demand for service [1]. Worldwide, 10-20% of children and adolescents experience mental health disorders [2]. An estimated 49.5% of adolescents in the United States will develop a mental health disorder over their lifetime [3]. In Canada, some 1.2 million children and youth are affected by mental illness [4]. By age 25, about 20% will develop a mental illness.

Given the prevalence of mental health issues among youth, it is not surprising that access to service is a problem. In Canada, an estimated 75% of children with mental health disorders do not access specialized treatment services [5]. If they do, wait times are a problem; average wait times are 6-12 months [9, 10]. This is especially concerning since, in severe illness, the risk of self-harm and suicide increases [11] in adolescents and young adults and is the second leading cause of death in the 15-24 age group [12]. Among underserved groups, the risk is even more alarming; in some First Nations communities, the suicide rate among youth under the age of 15 is almost 50 times greater than among non-indigenous populations [13].

Technologies like apps hold great potential for addressing the tremendous unmet need in a scalable, affordable manner and reaching people in remote locations. Seizing the opportunity requires moving high-quality digital tools into the healthcare system. There is a need for robust evidence to guide decisions about which apps to use, with which populations, for what purpose.

Healthcare providers prescribe medicine or provide therapy based on evidence of safety and efficacy. If apps are to be responsibly and successfully integrated into treatment within the formal healthcare system, they too must have credible evidence supporting safety and efficacy. Without such evidence, apps may simply be consumer products of unknown value or "digital supplements." Worse, in the absence of good data, substandard apps may be adopted within the healthcare system based on insufficient, or even misleading, evidence.

There is a need to make decisions based on sound evidence if the right tools are to be scaled up. This is the key to realizing the full potential of digital tools to have a global mental health impact. The ability to respond to the worldwide need to improve access to effective mental health services depends on this. There is a lot at stake.

At present, it is challenging to judge the quality of existing apps due to limited high-grade research. The current app environment may be described as "The Wild West." There are a growing number of mobile apps for mental health for youth [14-19]. However, most have not been rigorously assessed or validated. The science of

evaluating mental health apps, including those for youth, remains nascent. Previous reviews have concluded that the literature remains preliminary, and there is insufficient evidence to determine which mobile apps are empirically supported [20-24]. Limitations of the existing evidence base are illustrated in the following section.

2.0 Current Evaluation of Apps

In a previous report [25], we reviewed some highly downloaded apps and found severe limitations of the evaluations. Here is a brief recap of illustrative observations. The Headspace app makes numerous claims around efficacy, including that ten days of use will reduce stress by 14%, and three weeks will reduce aggression 57% [26]. However, a meta-analysis of randomized controlled trials of depression apps noted that while the overall effect size of mental health apps appears substantial at g=0.56 (in range with antidepressant medications), this reduces to the more modest g=0.22 when compared to an active control condition [27]. That is to say that the same mental health app studied in comparison to nothing (no control) versus walking (active control) will report a different degree of impact. The Headspace studies suggest that this app may not be as effective as touted. One study used a digital placebo version Headspace and found no benefit of using the app to improve focus in young people [28]. Another study that evaluated Headspace did find that users had reduced stress but found no benefit when compared to playing the game Tetris [29]. Another study reported that for some people, merely listening to woodland sounds might be more effective in reducing stress compared to listening to a guided meditation, one feature of Headspace [30]. The results are likely not unique to Headspace. Evaluations must include appropriate control groups.

A review paper of the mindfulness app Calm reported that the app reduces stress in students [31], but the control group was instructed not to partake in any mindfulness activities; this group was compared to the group who received mindfulness for stress reduction via the Calm app. With such an unbalanced control group, the results of the study only show that mindfulness can reduce stress compared to no mindfulness – which is a well-known fact. What is more interesting about this study is that only 56% of participating students used the app as directed, indicating that 44% were non-adherent despite volunteering to partake in the study. This lack of use of the app suggests yet another challenge around mental health apps.

Another study of the Calm app found a reduction in depressive symptoms in users. But the results presented in the appendix stratify decreases in depressive symptoms by engagement with the app and reveal that those who never engaged with the app had the same improvements as those who regularly did [32]. This lack of a dose-effect suggests that either an unconventional mechanism of action is responsible for change or that aspects of the digital placebo effect – a term defined by Torous and

Firth [33] as response related to expectations about app use – may be driving study results. The importance of evaluating effectiveness only after controlling for the digital placebo effect is clear. The extent to which apps are actually used must be evaluated, and outcomes assessed in relation to usage to examine dose-response relationships.

Most studies of smartphone apps do not present data on engagement in a consistent manner, as reported in a study by Ng et al. [34], making it difficult to either compare studies or understand trends in engagement and adherence across these apps. Standardized reporting is needed to improve the quality and understand the effectiveness of digital mental health apps on the market today and in the future.

A review of the most extensive studies in terms of the number of patients enrolled showed that most studies do not have interventions beyond four weeks, and the follow-up period is rarely more than eight weeks after the intervention (Table 1). This is not a long enough period to see if behavioral changes are sustained. Most large studies have been done in adults. Few studies have more than 50 patients. The apps also do not report on the cognitive or behavioral models used in the design of the app. Thus, it is not clear if the app is using validated cognitive models or evidence-based therapeutic protocols. It is essential to conduct evaluations with larger numbers of participants, to assess the extent to which the intervention is built on credible therapeutic protocols, and to do studies with meaningful follow-up periods to assess enduring effects.

App, Year [Reference]	Condition	Country	Target	Size of Control Group	Size of Intervention Group	Outcome Measures	Study Period	Follow-up Period	Content Validation	Funder Disclosed	Cognitive Behavior Model Described
Project EVO Arean 2016 [32]	Depression	USA	Adults with mild-moderate depression	206	211 / 209	PHQ-9	12 weeks	None	PHQ-9	NIMH	None
MyCompass Proudfoot 2013 [35]	Depression	Australia	Adults with mild-moderate depression	198	126 / 195	DASS	7 weeks	12 weeks	DASS	Australian Health and Ageing	None
MoodHacker Birney 2016 [36]	Depression	USA	Adults with mild-moderate depression	150	150	PHQ-9	6 weeks	10 weeks	PHQ-9	NIH / NIMH	None
CBM Active Enock 2014 [37]	Social Anxiety	USA	Adults	141	158	DASS	4 weeks	8 weeks	DASS	NR	None
SuperBetter Roepke 2015 [38]	Depression	USA	Adults with significant depression	93	93 / 97	CES-D	4 weeks	6 weeks	CES-D	NR	None
Mobiletype Reid 2013 [39]		Australia	Primary Care patients aged 14-24	46	68	DASS ESA	2-4 weeks	6 weeks	DASS ESA	Telstra Foundation	None
The Toolbox Bidargaddi 2017 [40]	Depression	Australia	Young adults in Australia	195	192	Well-being	4 weeks	up to 6 months	Mental Health Continuum Short Form (MHC-SF)	Young and Well Cooperative Research Centre	None

Table 1 – Evaluations of Mental Health apps with larger sample sizes

Systematic reviews have been done to compare apps for a specific condition. A systematic review is a scientific review of published studies that uses systematic methods to collect and analyze data, critically appraises research studies, and synthesizes findings qualitatively or quantitatively.

Table 2 shows some of the systematic reviews that have been published on youth mental health apps [20-24,41-65]. These reviews highlight the scarcity of studies that have enrolled 100 participants, which is often needed to make significant comparisons. A randomized controlled trial (RCT) is a type of scientific experiment that aims to reduce sources of bias when testing the effectiveness of new applications or treatments. In these studies, subjects are randomly allocated to two or more groups, including treatment and control groups. The aim is to ensure that people in treatment and control groups are comparable (to eliminate bias), and to do everything possible to ensure that controls are designed to eliminate "digital placebo effects" and other sources of bias (e.g., treatment participants receive face to face treatment in addition to the digital intervention, while controls do not). The groups are followed under conditions of the trial design that allow for unbiased observations of how effective the experimental intervention was.

Treatment efficacy is assessed in comparison to the control. In research studies, bias occurs when a systematic error is introduced into sampling (different participants in different groups), treatment (e.g., treatment group receives face to face treatment, controls do not) or testing (e.g., by encouraging one outcome or answer over others in collecting data). Few studies have met the rigorous standard of low bias in all categories of potential bias. More studies rigorously designed to eliminate biases are required.

Effect size is a quantitative measure of the magnitude of a phenomenon. How big an effect does the app have? This question must be addressed in doing evaluations. Cohen's effect is a common measure, and Cohen's effect greater than 0.5 is considered significant. Among all the systematic reviews summarized in Table 2, there are no studies that had over 100 test subjects, with low bias and Cohen's effect greater than 0.5. More extensive studies are needed that have carefully planned observations to minimize bias measure outcomes over an extended follow-up period.

Author-Year	Focus Area	Studies	Subjects N > 100	RCTs	Low bias in all categories	Effect Cohen >0.5	N>100, C>0.5, & Low Bias	Conclusions
Garrido-2019 [63]	Anxiety, Depression Adolescents	68	19	27	8	3	0	Studies comparing DMHI to achieve active controls conditions were not effective
Khan 2019 [60]	young people with neurodevelopmental disorder	10	1	10	Not reviewed	Not reviewed	0	There need to be more studies with larger sample sizes assessing the effectiveness of Web-based interventions for CYP.
Low 2019 [61]	Adolescents and young adults with chronic disease	29	3	5	Not reviewed	Not reviewed	Not reviewed	Adolescents and young adults are receptive to receiving information through a website/app as engagement in their care. Lack of interventional efficacy trials.
Botella 2017 [57]	Virtual reality exposure therapy (VRET)	11	0	11	Not reviewed	Not reviewed	Not reviewed	VRET has acceptable efficacy, an upcoming important tool for this type of psychological treatment. Acceptance by clinicians paramount.
Lau 2017 [56]	Serious games for mental health, all age groups	9	2	9	1	5*	0	Serious games for mental health seem feasible but there is a lack of quality and properly powered RCTs.
Grist 2017 [22]	Mental health mobile apps for preadolescents and adolescents	24	4	3	Not reviewed	Not reviewed	Not reviewed	More well-designed RCTs needed to evaluate the safety, efficacy, and effectiveness of mental health aps.
Firth 2017 [55]	Interventional apps for depression	18	11	18	1	6*	0	Smartphone-based interventions can have a moderate effect on depression symptoms.
Firth 2017 [54]	Interventional apps for anxiety	9	5	9	0	2*	0	These interventions are significantly effective for anxiety when compared with control, with the greatest effect when compared against wait-list controls.

Table 2 – Systematic Review of Mental Health Apps
(** Hodges's test, which is like Cohen's but incorporates bias for small sample sizes)

3.0 Current Evaluation Frameworks

Frameworks have been developed to guide those seeking high-quality apps for their personal use or to adopt in their clinical practice. The problem is that high-quality evidence needed to make relevant judgments does not exist. For instance, if a Framework suggests someone looking for an app consider effectiveness, this is excellent advice. The problem is there is almost undoubtedly insufficient evidence (and perhaps no evidence at all) to inform this judgment. The fundamental problem in the field is a dearth of high-quality evidence to guide people to the apps that are worthy of use.

In the absence of such evidence, Frameworks developed for people looking for a good app may use something other than evidence, such as expert opinion, to help guide decisions. Although we often need to make many decisions in the absence of evidence, there is no substitute for good evidence in clinical decision making.

A growing number of groups are proposing systems for evaluations of apps, and sometimes reviewing specific apps. However, most of these groups do not identify a reliable method for making unbiased evaluations. Further, reviews of specific apps are outdated by the time they are published. These problems are discussed in a 2019 review paper [65], which noted that the average time to review apps was between 109 and 714 days. Another study done in Canada used a set of guidelines to review apps with a group of 25 participants [66]. The results showed a high variation of ratings among reviewers and that 28% of reviewers indicated that they were uncertain of the overall quality of the health apps. This shows that unless there are clear guidelines and criteria for evaluation, supported by objective evidence, ratings will reflect a wide range of contradictory opinions.

The Mental Health Commission of Canada developed an evaluation toolkit for mobile mental health apps [67]. Although it draws attention to critical issues, this toolkit was not designed to evaluate the efficacy of apps scientifically. Most criteria are not specific enough to allow reliable evaluation. The Anxiety and Depression Association of America has webpages with app reviews [68]. Their criteria have several categories but not a clear definition of how to evaluate the app in those categories. For "effectiveness," the website does not specify who is qualified to be a rater and how the rating scale differs for scores from one to five. The Effectiveness category says that it covers Education, Self-Monitoring, and Treatment, but does not describe how to do those evaluations.

An initiative of the American Psychiatric Association, led by John Torous, MD, from our team, has created a framework with more specific category definitions, based on a position paper published in 2019 [69]. This approach has a hierarchical rating system and rubric to make APA members aware of very important information that should be considered when picking an app that is not the same as the information

used to judge a medication or therapy. The evaluation model's foundation rests in the maxim 'do no harm' as well as a risk-benefit analysis. The APA does not explicitly rate apps but provides a useful way to review apps. The four areas comprising the model (beyond gathering basic background information) are Safety/Privacy, Evidence (i.e., effectiveness), Ease of Use, and Interoperability. This a valuable approach, but further ongoing work is needed to develop the scientific method for evaluating each criterion, including therapeutic effectiveness, that can yield reliable results.

In fall 2019, the Clinical Trials Transformation Initiative (CTTI) issued best practices for clinical trial sponsors to guide app studies [70] to generate critical evidence. This work is useful, but details are vague and not operationally defined in terms that could guide app designers and evaluators. For example, the data management section states it is "critical to have a full understanding of the data flow and know who is responsible for the data and accountable for data integrity at each step during the data life cycle." While true, this does not make it clear how to evaluate security and data integrity.

In summary, since most apps are not well evaluated, reviews provided by existing Frameworks cannot be based on sound evidence of safety and effectiveness. Private vendor evaluations do not reveal what data were used for evaluation. Behavioral outcomes need to be independently analyzed with a large enough number of participants and an appropriate follow-up evaluation period and characteristics of the app itself need to be documented to inform decision making.

Existing evaluation frameworks are useful for a preliminary assessment, and often draw attention to key criteria to use when considering adopting an app. The core problem is that the evidence needed to make judgments about how well an app meets criteria is not available. The World Health Organization developed a framework for mental health services, but it does not include a method to evaluate mobile apps [1]. Several review papers have discussed the need for more formal in-depth evaluations [71-76]. A 2019 review of 45 evaluation frameworks for mobile health apps concluded that none of the frameworks could be used unaltered for health technology assessments, and only two of them evaluated the grade of evidence of the app [76]. A paper in 2019 proposed a framework for categories for evaluation [69] that does not score apps but provides a useful approach to review apps.

In short, although there are a growing number of mental health app evaluation frameworks, most of these frameworks have not been scientifically developed or validated. These frameworks may provide misleading information on the effectiveness of apps and may lead to misuse, misdiagnosis, wasted time, and even harm. The premise of the Framework presented below is that to move beyond "The Wild West" what is most urgently needed is a Framework that in a detailed and

systematic way outlines protocols for evaluating apps that will provide the evidence needed to inform decisions about which apps can be used with confidence within the healthcare system.

4.0 A Formal Evaluation Framework

4.1 Framework Summary

The Framework presented below calls for examining and transparently reporting appropriate background information, how the app was developed, how data is managed and controlled, the short-term outcomes of the app within the first month, and longer-term outcomes over six months or more. Behavioral interventions require long-term observations to measure sustained changes over time. These should ideally be done at least over six months and preferably over one year. This Framework a) builds on previous evaluation methods but provides a more detailed approach with clear rigor to guide a more scientific approach to measuring app efficacy and b) addresses pertinent issues that arise in evaluating apps that do not generally pertain to other clinical interventions (e.g., privacy threats inherent in the intervention).

The criteria are all deemed important, so they are not ranked. All of the criteria have a method to test that the criteria have been met. This is one of the critical missing elements from most existing frameworks. Without a method to unambiguously test that the criteria are met, the criteria are subject to misinterpretation and will not serve to reliably evaluate the app. It is difficult to measure "Do No Harm," but to ignore the risk of harm is akin to discovering new severe adverse effects of a drug and not taking action. The Framework will evolve over time as it is applied to more apps. This experience will enable a clearer definition of the criteria, refinement of methods for evaluating the criteria, and the introduction of new criteria as the Framework becomes more widely used.

About App	Design	Data Management	Short-Term Outcomes	Long-Term Outcomes
Intended Use	User Input	User Data Control	Product Usability	Effect Size
Legal Owner	Behavioral Model	Security	User Engagement	Effects over Time
Funding	Prototype Usability	Privacy	User Feedback	Factor Analysis
Cost	Personalization	Data Sharing	Do No Harm	Bias
Content Review	User Consent	Infrastructure	Face Validity	Sensitivity Analyses
Update Cycle	Ethical Principles	Interoperability	Efficacy and Dose-effect	Reproducibility

Table 3 – Summary of Criteria

4.2 Detailed Framework Description

Below is a detailed description of each criterion with guidance on how to test each criterion. The testing protocols are meant as a guide to the level of detail needed, but they are not exclusive. Alternative testing methods may be used if they are consistent with the aim of the criteria and at a similar level of detail.

Criteria	Description	Testing Method
Intended Use	The app clearly identifies the target condition and target audience.	The app must display, prior to download and in the app, the intended audience by target condition, age, and any other demographics or region.
Legal Owner	The app clearly shows the name of the company, organization, or individual that is legally responsible for the app, not just the developer of the app who may have been hired, but the owner of the app.	The app must clearly show, prior to download and in the app, a current mailing address, and an email or phone. Upon contact, the organization responds in a timely manner (within seven days).
Funding	The app needs to disclose the funding source of the app.	The app must declare who has paid for the development of the app. The funding organization must have a verifiable address or individuals who represent the organization.
Cost	The app clearly indicates the cost of usage, before and after any trial period, and methods for canceling service.	The app must clearly show, prior to download and in the app, how to request for cancellation and must respond in a timely way to cancellation requests.
Content Review	The app clearly indicates who reviewed the content and when the content was last updated. Any relevant conflict of interest of contributing authors must be disclosed.	The app must clearly show, prior to download and in the app, the names of the authors and organizations that have reviewed the educational or clinical content. The content must be current and reflect the best-known peer-reviewed clinical practice. Conflicts of interest must be disclosed with transparency about how those conflicts are resolved.
Update Cycle	The app shows the date of the last update, and there must not be more than two years between updates.	The app must clearly show, prior to download and in the app, the date that the app was released and last updated.
User Input	Users have been part of the design process.	Documentation must be available on how many users were consulted in the design process and how the feedback was collected and analyzed.
Behavioral Model	The app uses a validated cognitive and behavioral model.	A clinical expert for the intended therapeutic focus of the app will need to verify that the behavioral model used in the app is appropriate for the condition being treated.
Prototype Usability	In the development of the app, there was usability testing with the intended audience using a formal usability testing method.	Evaluations can be done using a validated heuristic usability scale applied by trained human factors professionals or by observed usage of the app to identify usability problems. Particular focus must be given to different form factors of mobile devices such as screen size and compatibility with previous models of devices and operating systems. The app should work with the most common devices for the intended audience in the market in the last three years.

Personalization	Users must be able to customize communications and content.	The app needs to allow users to select the frequency of notifications via email or push notifications and be able to turn them off. Messaging methods, particularly for youth, need to be culturally and age appropriate.
User Consent	Users must actively consent prior to any data sharing with third parties.	The app needs to have clear methods to show when data is being transferred to a third party and have user consent for any data used for research studies. The consent of minors requires the involvement of parents and must meet national legal requirements.
Ethical Principles	The development of the app should follow ethical principles that guide the design of the app.	The app needs to disclose the ethical principles such as "Health On The Net"principles [77] that guide the development of policies such as privacy, transparency in funding, attributions of content [78-80].
User Data Ownership and Control	Users must be able to delete their data from the app and any servers that back up that data.	The app needs to identify the method by which users can delete or request that their data be deleted from the app and any backup servers. Confirmation of the deletion must be sent to the user.
Security	Any transfer of data must be done using secure methods, and access to the site must be secure.	Login must be done by password, preferably two-factor authentication. Data must be transferred using end-to-end encryption using industry-standard communication protocols. Vulnerability testing must be done to determine if the app and servers can be penetrated by automated means.
Privacy	The app must have an accessible privacy policy that complies with national regulations and is written at a reading level appropriate for the intended users.	The privacy policy must be available before accessing the app and must indicate what state the company is based in and what legal jurisdiction it follows. The reading level should not be higher than grade twelve and preferably lower if intended for minors. There must be a way to contact the company for privacy violations.
Data Sharing	Any sharing of user data with external third parties must be clearly disclosed to users and only done with the prior consent of the user.	The terms and conditions policy must describe any data sharing. Updates to the terms and conditions must be sent to users. Users need to be able to turn off data sharing with third parties. Testing needs to be done to see if the app transmits data packets to external sites when data sharing is turned off versus turned on.
Infrastructure	The app identifies any infrastructure needed to use the app.	The app clearly discloses the technical requirements needed, including Internet connectivity, bandwidth, and any servers and technical support system.
Interoperability	The app should have a way to extract user data from the app or send the data to external systems.	Transmission of data should be done in open standards such as Fast Healthcare Interoperability Resources (FHIR) [81] for exchange with a medical record system.
Product Usability	For the final version app, there has been usability testing with existing users.	Data must be available in a peer-reviewed publication of the usage of the app analyzed by evaluators using coding techniques of the interactions to identify usability problems.
User Engagement	Data must be available to show the level of sustained usage by users after the initial month of usage.	Data must be available in a peer-reviewed publication that shows the number and proportion of users who continued to use the app after 30 days.

User Feedback	Users are able to submit feedback for the improvement of the app.	Data must be available that summarizes the user feedback on the app, and there is evidence that it has been used to make improvements. Positive user feedback does not indicate face validity.
No Harm	The app has been shown not to do harm.	There is evidence that the app has been evaluated for potential harm. If the app is discovered to cause harm, the owners of the app will take corrective actions such as removing the app from use.
Face validity	Data must be available that shows that the app addresses what the app is trying to treat.	There is expert opinion that confirms the app's usage in an intervention is credible and appropriately operationalized, as well as suited to the target condition and target audience.
Efficacy and Dose-effect	Evaluation data is available that shows efficacy based on controlled studies and examines dose-effect with the appropriate comparison group.	Data must be available in a peer-reviewed publication that shows the app is efficacious in unbiased studies that describe the control group (active or passive) and justify the measurement of dose-effect.
Effect Size	The app benefit has to be evaluated for effect size.	Published studies will show effect size using measures such as Cohen's effect size.
Effects over time	Evaluation data must be available that shows that the app is effective over time.	Effects that can be attributed to the app need to be measured at least six months after initial usage.
Factor Analysis	Evaluation of the app must take into account other concurrent therapeutic activities and be able to measure the relative effect of the app versus other factors that may influence outcomes.	Published studies must show the relative effect of other therapeutic activities in both quantitative and qualitative assessments.
Bias	Evaluation data must be available that shows that the app evaluations have been conducted without bias.	Published studies must report the potential sources of bias using the Cochrane framework [82] or similar framework.
Sensitivity analyses	The app must have multiple evaluation studies that have been reviewed using sensitivity analyses.	Using methodologies from scientific groups such as Cochrane Collaboration, sensitivity analysis needs to be done to measure fixed-effect and random-effects. For dichotomous outcomes, the evaluation must have appropriately used odds ratios, risk ratios, or risk differences. For continuous outcomes, where several scales have assessed the same dimension, an evaluation needs to be done to determine if standardized mean difference across all scales or as mean differences individually for each scale should have been used.
Reproducibility	Evaluation data must be available that shows that the app has been evaluated more than once and has consistently reproduced positive outcomes.	There must be more than one independently conducted evaluation with the app published with the CONSORT framework [83-84] or similar reporting standards used in peer-reviewed journals.

Table 4 – Detailed Description of Criteria

4.3 Discussion

About the App

The app needs to disclose the objectives and intended population of users. Without a clear objective and target audience, it is not possible to evaluate the effectiveness or enable users to decide if the app is appropriate for them. Knowing the legal owner and funding sources is essential to know who can be held accountable for the app and if there are potential conflicts of interest. Clear transparency in costs is necessary to understand if it is feasible to use and sustain over time. The content should be developed with current evidence-based knowledge and from authors who do not have conflicts of interest. Knowing the last date of release helps the user know if the content is still current and will ensure that the app still works with current mobile devices.

Design

In order to optimize the design of apps, it is essential to work with youth from the outset to understand if the design objectives and interaction design are appropriate for the intended audience. Youth have particular preferences for styles of interaction and communication, so these design criteria aim to optimize the design from an early stage. Various approaches can be used to engage and co-design with youth [85-88], such as Design Thinking [85]. Design Thinking is a non-linear, participatory, iterative process that seeks to understand users, challenge assumptions, redefine problems, and create innovative solutions to prototype and test. After the prototype is built, the app should be tested for usability [89-92]. This can be done with a validated usability survey instrument evaluated by human factors experts, preferably with user testing of the prototype with youth in an observational study. Having a reference ethical framework shows a coherent strategy in the approach to design and development. Any evaluation of the prototype of the app does not ensure the final version of the app will be validated for usability.

Data Management

Mobile apps can collect an enormous amount of information about the users, and users must provide consent for the collection of data. Data management should be clear to the user and allow users to delete their data from their app and any external servers that have a copy of the user's data. Users need to be able to decide what data they wish to share with any external third parties and have the ability to stop any sharing of data with third parties. The app should transfer data using secure methods at all times. The ability to transfer data to the health system, with the user's consent, is important to facilitate access to additional services such as telemedicine consults or integration of data with clinical records.

Short-Term Outcomes

The app must have a clear evaluation of usability, user engagement, usage, and efficacy. Many apps provide preliminary evaluations based on user feedback; these are valuable but subjective and cannot be used to evaluate the overall performance of the app. The app should have usability testing with observational methods and validated analytical methods. Any evaluation of the app with twenty or fewer users should only be considered preliminary evidence since it does not have enough users to gain statistical significance on usability problems analyzed due to limited representation of users. The app should not cause harm, and app owners should immediately remove the app if, for example, data breaches are detected in how the app collects and transmits data. It is difficult to anticipate all the possible harmful unintended consequences, but this criterion needs to be there to hold developers and app owners accountable for taking reasonable steps to monitor and appropriately act on evidence of adverse effects. The initial evaluation determines how many users continue to use the app beyond the initial days since most apps are not used beyond the first month. Short term outcomes need to be assessed using well-designed trials to eliminate biases; outcomes should be measured beyond the initial month of usage. The extent to which Hill's criteria for inferring causation [93] have been met can be considered in attributing outcomes to the use of the app, with the dose-response relationship of particular concern. Evaluations should be done arm's length from those with a stake in the app, with full findings transparently reported in peer-reviewed journals.

Several issues deserve highlighting in this section. First, apps may be used as stand-alone self-help tools or as part of a clinically supervised treatment program. This distinction is important to bear in mind in assessing efficacy. The same app may produce different outcomes across these two approaches to delivery. The latest evidence suggests that the efficacy of apps will be optimized if they are used in conjunction with a clinician and if the app offers users feedback on their progress [27]. Future efficacy studies should distinguish between apps used as purely self-help tools versus incorporated as an integral component of a clinician-led program, and/or used as an adjunct to professional treatment.

It is also important to recognize that apps may have the potential to do harm. There is an ongoing debate about the impact of screen time itself on mental health, but a lack of conclusive evidence. New research is emerging, such as the NIH-funded Adolescent Brain Cognitive Development (ABCD) study that endeavors to observe 12,000 healthy 9 and 10-year-old children at 21 sites around the United States over the next decade, measuring their mental and physical health and cognition while tracking brain changes, substance use, and digital media habits [94]. Already there is some evidence that screen time may impact memory, attention, and social cognition, although the effect on youth with mental health conditions remains unknown [95].

Ongoing research efforts will illuminate the much-debated effect of digital media use on the brain and the implications for using digital tools for therapeutic purposes.

Apart from screen time, app use may carry additional risks. A study of daily mood tracking among individuals with bipolar disorder, for example, found that mood tracking apps increased depressive symptoms in the intervention group compared to a control group with no access to the app [96]. Moreover, research surrounding digital health interventions for psychosis has suggested that routine self-assessment of symptoms may actually increase rumination on negative experiences [97]. Although more research is needed to probe the nature of such issues, specifically in youth, consideration of such significant risks must factor into future research efforts. The potential for harm warrants a cautious approach and underscores the importance of rigorous standards for the evaluation of mental health app efficacy.

Face validity also warrants elaboration. Face validity is concerned with whether the app actually addresses the issue it purports to treat, Does the app appear to be plausible based on an expert opinion? Does it take a credible approach to do what it claims to do? Is it suited to the intended user population? For instance, if an app purports to provide online CBT, its features and functions should align with validated CBT components. This first condition is critical, as the majority of mental health apps are categorized as "health and wellness" apps and thus not regulated by the FDA. Without an external accreditation body, medical claims can go unchecked. If there is a lack of face validity at the outset, then additional claims of efficacy are moot. In the absence of evidence and regulation, this is a quick way to screen out apps that are clearly substandard or unsuited to the purpose or to population.

Face validity can be assessed by an expert in the field but will not easily conform to a single scoring criterion. Who is an expert is itself a non-trivial question; the expertise needed to make face validity judgments will vary by clinical objective, by the app, by region, and by culture.

As discussed earlier, there is a significant difference in reported efficacy when apps are measured against an active versus passive control. Just as with traditional RCTs, efficacy studies for mental health apps should attempt to measure dose-effects—that is, whether different levels of usage are associated with different effects—perhaps including both passive and active controls for a thorough and transparent investigation. Evolving standards for the field may include trials with control, sham, and active apps, as this study design would offer the robust evidence necessary to make informed decisions.

Long-Term Outcomes

The criterion of effect size is a scientific measure that helps assess how much benefit is provided by the treatment. This should be measured as part of efficacy studies, including follow-up studies. Given the novel and rapidly accelerating pace of development in the digital health space, with 50% turnover every 100 days for apps related to depression and bipolar disorder [98], it is rare to find apps supported by a single robust efficacy study, let alone multiple. The evaluation should to be conducted beyond six months to determine if a user's behavioral changes are sustained. Since the app may be used with other complementary therapies, it is necessary to do a factor analysis to measure the relative contribution of the app versus other factors to the overall observed behavior change. The measure of bias is essential to understand if problems in the evaluation may have introduced any unintended influence in the results. Sensitivity analysis is essential in understanding the relationship between observations and measuring levels of uncertainty in the relation between inputs and outputs. Reproducibility of findings supporting efficacy across well-designed, unbiased studies enhances confidence in the app, especially if these studies consistently report large effect sizes in the long term as well as short term.

Ongoing Developmental Evaluation: How can app performance be optimized in practice?

Once an app has been put into practice based on sound evidence of efficacy, ongoing evaluation studies can be invaluable. Such studies can assess the impact of modifications to the app, its performance with different subpopulations of users, etc. The evidence generated can be invaluable for continually guiding improvements to the tool and the way it is used [99-100]. Evidence carefully generated from practice is highly useful for improving practice.

5.0 Conclusions

Research on the efficacy of these digital interventions is desperately needed, given the vast proliferation of mobile apps for mental health. New frameworks, like the American Psychiatric Association's model, have emerged for the evaluation of apps; however, the lack of consistency in assessing efficacy among these frameworks remains a continued source of confusion for patients, providers, and larger health care systems. In the current space, exaggerated claims by for-profit app companies can go unchecked and unsubstantiated.

The core problem in this area is the lack of availability of transparently reported efficacy data required to use existing frameworks to make decisions that are well informed by reliable evidence. We have proposed a set of four guidelines to provide researchers, clinicians, and users with a benchmark for assessing app efficacy to generate better evidence, in response to an urgent need. These guidelines are meant to raise the bar for making claims related to effectiveness and other characteristics of apps and pave the way for more nuanced analyses of apps and their numerous distinct uses.

Technologically-savvy youth are expressing interest in digital solutions, and with the incidence of mental illness climbing steadily upwards, there is an urgent need—and opportunity—to link young people with the resources and support they need. The proposed guidelines provide a consistent set of standards by which the efficacy of various apps can be evaluated, facilitating the integration of mobile health apps into clinical care.

Society has a stake in this field. It is not just users of apps, clinicians, and researchers who need to be concerned with the issues raised in this paper. Healthcare systems, and indeed the health and welfare of societies, depend on creating an environment that promotes the development and use of high-quality apps based on the best possible evidence of their value, safety and ethical integrity. This perspective raises questions about the role of regulation in creating an environment that enables the development and use of valuable apps within the healthcare system and by individual users who wish to use top quality apps while mitigating the potential for harm.

Acknowledgements

Yuri Quintana and John Torous prepared the final report, with contributions from the following people in alphabetical order:

- Darren Fahy, Research Coordinator, Division of Clinical Informatics, Beth Israel Deaconess Medical, Boston, MA, USA.
- Joanna Henderson, PhD, CPsych, Director, Margaret and Wallace McCain Centre for Child, Youth & Family Mental Health; and Senior Scientist in the Child, Youth and Emerging Adult Program, Centre for Addiction and Mental Health, Toronto, Ontario, Canada.
- Alejandro R. Jadad, MD, Professor, Dalla Lana School of Public Health, University of Toronto, Canada.
- Sarah Lagan, Department of Psychiatry, Harvard Medical School, Boston, MA, USA.
- Kimberly H.M. O'Brien, Assistant Professor of Psychiatry, Boston Children's Hospital, Boston, MA, USA.
- Yuri Quintana, PhD, Chief, Division of Clinical Informatics, Beth Israel Deaconess Medical Center, Assistant Professor of Medicine, Harvard Medical School, Boston, MA, USA.
- Charles Safran, MD, Professor of Medicine, Harvard Medical School, Boston, MA, USA.
- John Torous, MD, Chief, Division of Digital Psychiatry, Instructor in Psychiatry, Harvard Medical School, Boston, MA, USA.

During the development of the project we had consultations with the following people:

- Alaa Ali E. Abd-Alrazaq, PhD, Postdoc, Health Informatics, Hamad Bin Khalifa University, Qatar
- Nashva Mohammed Ali, Senior Researcher, Hamad Bin Khalifa University, Qatar
- Mohamed Alarakhia, MD, The eHealth Centre of Excellence, Kitchener, Ontario Canada.
- Elizabeth Borycki, PhD, RN, Professor Health Information Science, Director - Social Dimensions of Health Program, University of Victoria, Victoria, BC, Canada
- Roy Cameron, PhD, Executive Director, Homewood Research Institute, Guelph, Ontario, Canada
- Farooq Naeem, MD, Chief of General and Health Systems Psychiatry, CAMH, and Professor of Psychiatry at the University of Toronto
- Cyndy Forsyth, Chief Development Officer, Homewood Research Institute, Guelph, Ontario, Canada
- Andre Kushniruk, PhD, Professor and Director of the School of Health Information Science, University of Victoria, Victoria, BC, Canada.
- Vimla L. Patel, PhD, Senior Research Scientist and Director, Center for Cognitive Studies in Medicine and Public Health, The New York Academy of Medicine, and Adjunct Professor Department of Biomedical Informatics, Columbia University, New York, NY
- Mowafa Househ, PhD, Associate Professor, Hamad Bin Khalifa University, Qatar
- Brian Rush, PhD - Senior Scientist, Homewood Research Institute, and Scientist Emeritus, Centre for Addiction and Mental Health, Toronto, Ontario, Canada
- Stanley Kutcher, MD, Senator of Canada, Ottawa, Ontario
- David Wiljer PhD, Executive Director, Education Technology Innovation at University Health Network, Toronto, Associate Professor, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada

Please note that the analyses and conclusions in the present document do not necessarily reflect those of the individuals or organizations mentioned above.

References

1. World Health Organization. Child and Adolescent Mental Health. Available at URL: https://www.who.int/mental_health/maternal-child/child_adolescent/en/ [Accessed May 20, 2020].

2. World Health Organization. Adolescent mental health: mapping actions of nongovernmental organizations and other international development organizations. 2012, pp. 1-49. Available at URL:https://www.who.int/mental_health/publications/adolescent_mental_health/en/ [Accessed May 20, 2020].

3. Merikangas KR, He JP, Burstein M, Swanson SA, Avenevoli S, Cui L, Benjet C, Georgiades K, Swendsen J. Lifetime prevalence of mental disorders in U.S. adolescents: results from the National Comorbidity Survey Replication—Adolescent Supplement (NCS-A). J Am Acad Child Adolesc Psychiatry. 2010 Oct;49(10):980-9. doi: 10.1016/j.jaac.2010.05.017. Epub 2010 Jul 31. PubMed PMID: 20855043; PubMed Central PMCID: PMC2946114. Available at URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2946114/

4. Mental Health Commission of Canada. Children and Youth. Available at URL: https://www.mentalhealthcommission.ca/English/what-we-do/children-and-youth [Accessed May 20, 2020].

5. Canadian Institute for Health Information. Care for Children and Youth With Mental Disorders. May 2015. Available at URL: https://secure.cihi.ca/free_products/CIHI%20CYMH%20Final%20for%20pubs_EN_web.pdf Last [Accessed May 20, 2020].

6. Hoffmann JA, Stack AM, Samnaliev M, Monuteaux MC, Lee LK. Trends in Visits and Costs for Mental Health Emergencies in a Pediatric Emergency Department, 2010-2016. Acad Pediatr. 2019 May - Jun;19(4):386-393. doi:10.1016/j.acap.2019.02.006. Epub 2019 Feb 21. PubMed PMID: 30797896. URL: https://www.ncbi.nlm.nih.gov/pubmed/?term=30797896%5Buid%5D

7. Institute for Health Metrics and Evaluation (2015). Global Burden of Diseases, Injuries, and Risk Factors Study, 2013. Available at URL: http://www.healthdata.org/data-visualization/gbd-compare Last Accessed 20 Aug 2019.

8. Waddell C, McEwan K, Shepherd CA, Offord DR, Hua JM. A public health strategy to improve the mental health of Canadian children. Can J Psychiatry. 2005 Mar;50(4):226-33. Review. PubMed PMID: 15898462. URL: https://www.ncbi.nlm.nih.gov/pubmed/?term=15898462

9. Children's Mental Health Ontario (2016). Ontario's children waiting up to 1.5 years for urgently needed mental healthcare. Available at URL: https://www.cmho.org/blog/blog-news/6519730-let-s-talk-children-s-mental-health [Accessed May 20, 2020].

10. Office of the Auditor General of Ontario (2016). Annual report 2016, volume 1. Toronto, Ontario, Canada: Queen's Printer for Ontario. Available at URL: http://www.auditor.on.ca/en/content/annualreports/arreports/en16/2016AR_v1_en_web.pdf [Accessed May 20, 2020].

11. Hawton K, van Heeringen K. Suicide. Lancet. 2009 Apr 18;373(9672):1372-81. doi: 10.1016/S0140-6736(09)60372-X. Review. PubMed PMID: 19376453. URL: https://www.ncbi.nlm.nih.gov/pubmed/?term=19376453

12. Statistics Canada. Table 13-10-0394-01 Leading causes of death, total population, by age group Available at URL: https://doi.org/10.25318/1310039401-eng [Accessed May 20, 2020].

13. Kumar MB, Tjepkema M. Suicide among First Nations people, Metis and Inuit (2011-2016): Findings from the 2011 Canadian Census Health and Environment Cohort (CanCHEC). National Household Survey: Aboriginal People. Stats Canada. Available at URL https://www150.statcan.gc.ca/n1/pub/99-011-x/99-011-x2019001-eng.htm [Accessed May 20, 2020].

14. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, Davies EB. Annual Research Review: Digital health interventions for children and young people with mental health problems - a systematic and meta-review. J Child Psychol Psychiatry. 2017 Apr;58(4):474-503. doi: 10.1111/jcpp.12663. Epub 2016 Dec 10. Review. PubMed PMID: 27943285. URL: https://www.ncbi.nlm.nih.gov/pubmed/?term=27943285

15. Garrido S, Cheers D, Boydell K, et al. Young People's Response to Six Smartphone Apps for Anxiety and Depression: Focus Group Study. JMIR Ment Health. 2019;6(10):e14385. Published 2019 Oct 2. doi:10.2196/14385

16. Fleming TM, Bavin L, Stasiak K, Hermansson-Webb E, Merry SN, Cheek C, Lucassen M, Lau HM, Pollmuller B, Hetrick S. Serious Games and Gamification for Mental Health: Current Status and Promising Directions. Front Psychiatry. 2017 Jan 10;7:215. doi: 10.3389/fpsyt.2016.00215. eCollection 2016. PubMed PMID: 28119636; PubMed Central PMCID: PMC5222787. Available at URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5222787/

17. Sharifzadeh N, Kharrazi H, Nazari E, Tabesh H, Edalati M, Heidari S, Tara M. Health education serious games: A scoping review. JMIR Preprints. 22/01/2019:13459 DOI: 10.2196/preprints.13459 Available at URL: https://preprints.jmir.org/preprint/13459 [Accessed May 20, 2020].

18. Archangeli C, Marti FA, Wobga-Pasiah EA, Zima B. Mobile Health Interventions for Psychiatric Conditions in Children: A Scoping Review. Child Adolesc Psychiatr Clin N Am. 2017 Jan;26(1):13-31. doi: 10.1016/j.chc.2016.07.009. Epub 2016 Oct 18. Review. PubMed PMID: 27837939. URL: https://www.ncbi.nlm.nih.gov/pubmed/?term=27837939

19. Lui, J. H. L., Marcus, D. K., & Barry, C. T. (2017). Evidence-based apps? A review of mental health mobile applications in a psychotherapy context. Professional Psychology: Research and Practice, 48(3), 199-210. DOI: 10.1037/pro0000122 URL: https://psycnet.apa.org/doiLanding?doi=10.1037%2Fpro0000122

20. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital Mental Health Interventions for Depression, Anxiety, and Enhancement of Psychological Well-Being Among College Students: Systematic Review. J Med Internet Res. 2019 Jul 22;21(7):e12869. doi: 10.2196/12869. PubMed PMID: 31333198; PubMed Central PMCID: PMC6681642. Available at URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6681642/

21. Batra S, Baker RA, Wang T, Forma F, DiBiasi F, Peters-Strickland T. Digital health technology for use in patients with serious mental illness: a systematic review of the literature. Med Devices (Auckl). 2017 Oct 4;10:237-251. doi: 10.2147/MDER.S144158. eCollection 2017. Review. PubMed PMID: 29042823; PubMed Central PMCID: PMC5633292. Available at URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5633292/

22. Grist R, Porter J, Stallard P. Mental Health Mobile Apps for Preadolescents and Adolescents: A Systematic Review. J Med Internet Res. 2017 May 25;19(5):e176. doi: 10.2196/jmir.7332. Review. PubMed PMID: 28546138; PubMed Central PMCID:PMC5465380. Available at URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5465380/

23. 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;107:73-78. doi: 10.1016/j.jpsychires.2018.10.006. Epub 2018 Oct 5. PubMed PMID: 30347316. URL: https://www.ncbi.nlm.nih.gov/pubmed/30347316

24. Larsen ME, Huckvale K, Nicholas J, Torous J, Birrell L, Li E, Reda B. Using science to sell apps: Evaluation of mental health app store quality claims. NPJ Digit Med. 2019 Mar 22;2:18. doi: 10.1038/s41746-019-0093-1. eCollection 2019. PubMed PMID: 31304366; PubMed Central PMCID: PMC6550255. Available at URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6550255/

25. Homewood Research Institute. Youth Mental Health Apps in the Digital Age: A Scoping Review of Trends and Evaluations. Available at URL: https://homewoodresearch.org/app-evaluation-project/ Accessed May 20, 2020.

26. Headspace App. Available at URL: https://www.headspace.com/science/meditation-research [Accessed May 20, 2020].

27. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, Sarris J. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry. 2017 Oct;16(3):287-298. doi: 10.1002/wps.20472. PubMed PMID: 28941113; PubMed Central PMCID: PMC5608852. Available at URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5608852/

28. Noone C, Hogan MJ. A randomised active-controlled trial to examine the effects of an online mindfulness intervention on executive control, critical thinking and key thinking dispositions in a university student sample. BMC Psychol. 2018 Apr 5;6(1):13. doi: 10.1186/s40359-018-0226-3. PubMed PMID: 29622047; PubMed Central PMCID: PMC5887193. Available at URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5887193/

29. Collins E, Cox A, Wilcock C, Sethu-Jones G. Digital Games and Mindfulness Apps: Comparison of Effects on Post Work Recovery. JMIR Ment Health. 2019 Jul 18;6(7):e12853. doi: 10.2196/12853. PubMed PMID: 31322125; PubMed Central PMCID: PMC6670275. Available at URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6670275/

30. The Guardian. Woodland sounds help relaxation more than meditation apps – study. 12 Sep 2019. Available at URL: https://www.theguardian.com/uk-news/2019/sep/13/woodland-sounds-help-relaxation-more-than-meditation-apps-study [Accessed May 20, 2020].

31. Huberty J, Green J, Glissmann C, Larkey L, Puzia M, Lee C. Efficacy of the Mindfulness Meditation Mobile App "Calm" to Reduce Stress Among College Students: Randomized Controlled Trial. JMIR Mhealth Uhealth. 2019 Jun 25;7(6):e14273. doi: 10.2196/14273. PubMed PMID: 31237569; PubMed Central PMCID: PMC6614998. Available at URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6614998/

32. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, Anguera JA. The Use and Effectiveness of Mobile Apps for Depression: Results From a Fully Remote Clinical Trial. J Med Internet Res. 2016 Dec 20;18(12):e330. doi:10.2196/jmir.6482. PubMed PMID: 27998876; PubMed Central PMCID: PMC5209607. Available at URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5209607/

33. Torous J, Firth J. The digital placebo effect: mobile mental health meets clinical psychiatry. *Lancet Psychiatry*. 2016;3(2):100-102. doi:10.1016/S2215-0366(15)00565-9

34. Ng MM, Firth J, Minen M, Torous J. User Engagement in Mental Health Apps: A Review of Measurement, Reporting, and Validity. Psychiatr Serv. 2019 Jul 1;70(7):538-544. doi: 10.1176/appi.ps.201800519. Epub 2019 Mar 27. PubMed PMID: 30914003. URL: https://www.ncbi.nlm.nih.gov/pubmed/?term=30914003

35. Proudfoot J, Clarke J, Birch MR, Whitton AE, Parker G, Manicavasagar V, Harrison V, Christensen H, Hadzi-Pavlovic D. Impact of a mobile phone and web program on symptom and functional outcomes for people with mild-to-moderate depression, anxiety and stress: a randomised controlled trial. BMC Psychiatry. 2013 Nov 18;13:312. doi: 10.1186/1471-244X-13-312. PubMed PMID: 24237617; PubMed Central PMCID: PMC4225666. Available at URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4225666/

36. Birney AJ, Gunn R, Russell JK, Ary DV. MoodHacker Mobile Web App With Email for Adults to Self-Manage Mild-to-Moderate Depression: Randomized Controlled Trial. JMIR Mhealth Uhealth. 2016 Jan 26;4(1):e8. doi: 10.2196/mhealth.4231. PubMed PMID: 26813737; PubMed Central PMCID: PMC4748138. Available at URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4748138/

37. Enock, P.M., Hofmann, S.G. & McNally, R.J. Attention Bias Modification Training Via Smartphone to Reduce Social Anxiety: A Randomized, Controlled Multi-Session Experiment. Cognitive Therapy and Research. (2014) 38: 200. URL: https://doi.org/10.1007/s10608-014-9606-z

38. Roepke AM, Jaffee SR, Riffle OM, McGonigal J, Broome R, Maxwell B. Randomized Controlled Trial of SuperBetter, a Smartphone-Based/Internet-Based Self-Help Tool to Reduce Depressive Symptoms. Games Health J. 2015 Jun;4(3):235-46. doi:10.1089/g4h.2014.0046. Epub 2015 Feb 19. PubMed PMID: 26182069. URL: https://www.ncbi.nlm.nih.gov/pubmed/?term=26182069

39. Reid SC, Kauer SD, Hearps SJ, Crooke AH, Khor AS, Sanci LA, Patton GC. A mobile phone application for the assessment and management of youth mental health problems in primary care: health service outcomes from a randomised controlled trial of mobiletype. BMC Fam Pract. 2013 Jun 19;14:84. doi:10.1186/1471-2296-14-84. PubMed PMID: 23782796; PubMed Central PMCID: PMC3716527. Available at URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3716527/

40. Bidargaddi N, Musiat P, Winsall M, Vogl G, Blake V, Quinn S, Orlowski S, Antezana G, Schrader G. Efficacy of a Web-Based Guided Recommendation Service for a Curated List of Readily Available Mental Health and Well-Being Mobile Apps for Young People: Randomized Controlled Trial. J Med Internet Res. 2017 May 12;19(5):e141. doi: 10.2196/jmir.6775. PubMed PMID: 28500020; PubMed Central PMCID: PMC5446666. Available at URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5446666/

41. Gainsbury S, Blaszczynski A. A systematic review of Internet-based therapy for the treatment of addictions. Clinical psychology review. 2011 Apr 1;31(3):490-8.

42. Donker T, Petrie K, Proudfoot J, Clarke J, Birch MR, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. Journal of medical Internet research. 2013;15(11):e247.

43. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. General hospital psychiatry. 2013 Jul 1;35(4):332-8.

44. Free C, Phillips G, Galli L, Watson L, Felix L, Edwards P, Patel V, Haines A. The effectiveness of mobile-health technology-based health behaviour change or disease management interventions for health care consumers: a systematic review. PLoS med. 2013 Jan 15;10(1):e1001362.

45. DeSmet A, Van Ryckeghem D, Compernolle S, Baranowski T, Thompson D, Crombez G, Poels K, Van Lippevelde W, Bastiaensens S, Van Cleemput K, Vandebosch H. A meta-analysis of serious digital games for healthy lifestyle promotion. Preventive medicine. 2014 Dec 1;69:95-107.

46. Musiat P, Tarrier N. Collateral outcomes in e-mental health: a systematic review of the evidence for added benefits of computerized cognitive behavior therapy interventions for mental health. Psychological medicine. 2014 Nov;44(15):3137-50.

47. Sunyaev A, Dehling T, Taylor PL, Mandl KD. Availability and quality of mobile health app privacy policies. Journal of the American Medical Informatics Association. 2015 Apr 1;22(e1):e28-33.

48. Larsen ME, Nicholas J, Christensen H. A systematic assessment of smartphone tools for suicide prevention. PloS one. 2016;11(4).

49. Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and evaluation of mindfulness-based iPhone apps. JMIR mHealth and uHealth. 2015;3(3):e82.

50. Shen N, Levitan MJ, Johnson A, Bender JL, Hamilton-Page M, Jadad AA, Wiljer D. Finding a depression app: a review and content analysis of the depression app marketplace. JMIR mHealth and uHealth. 2015;3(1):e16.

51. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR mental health. 2016;3(1):e7.

52. Creber RM, Maurer MS, Reading M, Hiraldo G, Hickey KT, Iribarren S. Review and analysis of existing mobile phone apps to support heart failure symptom monitoring and self-care management using the Mobile Application Rating Scale (MARS). JMIR mHealth and uHealth. 2016;4(2):e74.

53. Huguet A, Rao S, McGrath PJ, Wozney L, Wheaton M, Conrod J, Rozario S. A systematic review of cognitive behavioral therapy and behavioral activation apps for depression. PloS one. 2016;11(5).

54. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. Journal of affective disorders. 2017 Aug 15;218:15-22.

55. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, Sarris J. 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.

56. Lau HM, Smit JH, Fleming TM, Riper H. Serious games for mental health: are they accessible, feasible, and effective? A systematic review and meta-analysis. Frontiers in psychiatry. 2017 Jan 18;7:209.

57. Botella C, Fernández-Álvarez J, Guillén V, García-Palacios A, Baños R. Recent progress in virtual reality exposure therapy for phobias: a systematic review. Current psychiatry reports. 2017 Jul 1;19(7):42.

58. Larsen ME, Huckvale K, Nicholas J, Torous J, Birrell L, Li E, Reda B. Using science to sell apps: evaluation of mental health app store quality claims. NPJ digital medicine. 2019 Mar 22;2(1):1-6.

59. Lu AS, Kharrazi H. A state-of-the-art systematic content analysis of games for health. Games for health journal. 2018 Feb 1;7(1):1-5.

60. Khan K, Hall CL, Davies EB, Hollis C, Glazebrook C. The Effectiveness of Web-Based Interventions Delivered to Children and Young People With Neurodevelopmental Disorders: Systematic Review and Meta-Analysis. Journal of medical Internet research. 2019;21(11):e13478.

61. Low JK, Manias E. Use of Technology-Based Tools to Support Adolescents and Young Adults With Chronic Disease: Systematic Review and Meta-Analysis. JMIR mHealth and uHealth. 2019;7(7):e12042.

62. Jeminiwa RN, Hohmann NS, Fox BI. Developing a Theoretical Framework for Evaluating the Quality of mHealth Apps for Adolescent Users: A Systematic Review. J Pediatr Pharmacol Ther. 2019 Jul-Aug;24(4):254-269. doi:10.5863/1551-6776-24.4.254. Review. PubMed PMID: 31337988; PubMed Central PMCID: PMC6633275.

63. Garrido S, Millington C, Cheers D, Boydell K, Schubert E, Meade T, Nguyen QV. What Works and What Doesn't Work? A Systematic Review of Digital Mental Health Interventions for Depression and Anxiety in Young People. Front Psychiatry. 2019 Nov 13;10:759. doi: 10.3389/fpsyt.2019.00759. eCollection 2019. PubMed PMID: 31798468; PubMed Central PMCID: PMC6865844.

64. Carlo AD, Hosseini Ghomi R, Renn BN, Areán PA. By the numbers: ratings and utilization of behavioral health mobile applications. NPJ Digit Med. 2019 Jun 17;2:54. doi: 10.1038/s41746-019-0129-6. eCollection 2019. PubMed PMID: 31304400; PubMed Central PMCID: PMC6572775. Available at URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6572775/

65. Punukollu M, Marques M. Use of mobile apps and technologies in child and adolescent mental health: a systematic review. Evidence-Based Mental Health. 2019 Jul 29:ebmental-2019.

66. Zelmer J, van Hoof K, Notarianni M, van Mierlo T, Schellenberg M, Tannenbaum C. An Assessment Framework for e-Mental Health Apps in Canada: Results of a Modified Delphi Process. JMIR Mhealth Uhealth. 2018 Jul 9;6(7):e10016. doi: 10.2196/10016. PubMed PMID: 29986846; PubMed Central PMCID: PMC6056739.

67. Canadian Mental Health Commission. Toolkit for e-Mental Health Implementation. August 2018. Available URL: https://www.mentalhealthcommission.ca/English/media/4137 [Accessed May 20, 2020].

68. The Anxiety and Depression Association of America. App reviews. Available at URL: https://adaa.org/finding-help/mobile-apps [Accessed May 20, 2020].

69. Henson P, David G, Albright K, Torous, J. Deriving a practical framework for the evaluation of health apps. The Lancet Digital Health Volume 1, Issue 2, June 2019, Pages e52-e54 Available at URL: https://www.thelancet.com/journals/landig/article/PIIS2589-7500(19)30013-5/fulltext [Accessed May 20, 2020].

70. Coran P, Goldsack JC, Grandinetti CA, Bakker JP, Bolognese M, Dorsey ER, Vasisht K, Amdur A, Dell C, Helfgott J, Kirchoff M. Advancing the Use of Mobile Technologies in Clinical Trials: Recommendations from the Clinical Trials Transformation Initiative. Digital Biomarkers. 2019;3(3):145-54.

71. Chan S, Torous J, Hinton L, Yellowlees P. Towards a Framework for Evaluating Mobile Mental Health Apps. Telemed J E Health. 2015 Dec;21(12):1038-41. doi:10.1089/tmj.2015.0002. Epub 2015 Jul 14. PubMed PMID: 26171663. URL: https://www.ncbi.nlm.nih.gov/pubmed/?term=26171663

72. Shen N, Levitan MJ, Johnson A, Bender JL, Hamilton-Page M, Jadad AA, Wiljer D. Finding a depression app: a review and content analysis of the depression app marketplace. JMIR Mhealth Uhealth. 2015 Feb 16;3(1):e16. doi:10.2196/mhealth.3713. Review. PubMed PMID: 25689790; PubMed Central PMCID: PMC4376135.URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4376135/

73. Singh K, Drouin K, Newmark LP, Lee J, Faxvaag A, Rozenblum R, Pabo EA, Landman A, Klinger E, Bates DW. Many Mobile Health Apps Target High-Need, High-Cost Populations, But Gaps Remain. Health Aff (Millwood). 2016 Dec 1;35(12):2310-2318. PubMed PMID: 27920321. URL: https://www.ncbi.nlm.nih.gov/pubmed/?term=27920321

74. Huckvale K, Prieto JT, Tilney M, Benghozi PJ, Car J. Unaddressed privacy risks in accredited health and wellness apps: a cross-sectional systematic assessment. BMC Med. 2015 Sep 7;13:214. doi: 10.1186/s12916-015-0444-y. PubMed PMID: 26404673; PubMed Central PMCID: PMC4582624. Available at URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4582624/

75. Ranasinghe M, Cabrera A, Postel-Vinay N, Boyer C. Transparency and Quality of Health Apps: The HON Approach. Stud Health Technol Inform. 2018;247:656-660. PubMed PMID: 29678042. URL: https://www.ncbi.nlm.nih.gov/pubmed/?term=29678042

76. Moshi MR, Tooher R, Merlin T. SUITABILITY OF CURRENT EVALUATION FRAMEWORKS FOR USE IN THE HEALTH TECHNOLOGY ASSESSMENT OF MOBILE MEDICAL APPLICATIONS: A SYSTEMATIC REVIEW. Int J Technol Assess Health Care. 2018 Jan;34(5):464-475. Epub 2018 Sep 11. PubMed PMID: 30201060. URL: https://www.ncbi.nlm.nih.gov/pubmed/?term=30201060

77. Health On the Net Foundation. Discover the 8 principles of the HONcode in 35 languages. 2019. Available at URL: https://www.hon.ch/cgi-bin/HONcode/principles.pl Accessed May 20, 2020.

78. COPE (Committee on. Publication Ethics) Code of Conduct and Best Practice Guidelines for Journal Editors. Available at URL: http://publicationethics.org/files/Code_of_conduct_for_journal_editors_Mar11.pdf Accessed May 20, 2020.

79. International Committee of Medical Journal Editors. Conflicts of interest / Competing interests. 2020 Available at URL: https://publicationethics.org/competinginterests Accessed May 20, 2020.

80. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly work in Medical Journals. Available at URL: http://www.icmje.org/recommendations/ Accessed May 20, 2020.

81. Health Level Seven International. HL 7 FHIR Release 4. Available at URL: http://hl7.org/fhir/ Accessed May 20, 2020.

82. Cochrane Collaboration. Cochrane Handbook for Systematic Reviews of Interventions Available at URL: https://training.cochrane.org/handbook/current Accessed May 20, 2020.

83. The CONSORT Group. Consolidated Standards of Reporting Trials Available at URL: https://www.consort-statement.org Accessed May 20, 2020.

84. Eysenbach G. CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011;13:e126. doi:10.2196/jmir.1923.

85. IDEO Inc. Design Thinking for Educators Toolkit. Available at URL: https://designthinkingforeducators.com/ Accessed May 20, 2020.

86. Carlson, Cindy. "The Hampton experience as a new model for youth civic engagement." *Journal of Community Practice* 14, no. 1-2 (2006): 89-106.

87. Heffernan, Olivia S., Tyson M. Herzog, Jordana E. Schiralli, Lisa D. Hawke, Gloria Chaim, and Joanna L. Henderson. "Implementation of a youth-adult partnership model in youth mental health systems research: Challenges and successes." *Health Expectations* 20, no. 6 (2017): 1183-1188.

88. Hawke, Lisa D., Jacqueline Relihan, Joshua Miller, Emma McCann, Jessica Rong, Karleigh Darnay, Samantha Docherty, Gloria Chaim, and Joanna L. Henderson. "Engaging youth in research planning, design and execution: practical recommendations for researchers." *Health Expectations* 21, no. 6 (2018): 944-949.

89. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: A scoping review. *Int J Med Inform*. 2019;126:95-104. doi:10.1016/j.ijmedinf.2019.03.018 Available at URL: https://doi.org/10.1016/j.ijmedinf.2019.03.018 Accessed May 20, 2020.

90. Parsazadeh, Nadia, Rosmah Ali, Mehran Rezaei, and Sanaz Zolfaghar Tehrani. "The construction and validation of a usability evaluation survey for mobile learning environments." *Studies in Educational Evaluation* 58 (2018): 97-111. Available at URL: https://doi.org/10.1016/j.stueduc.2018.06.002 Accessed May 20, 2020.

91. Kushniruk AW, Borycki EM. Development of a Video Coding Scheme for Analyzing the Usability and Usefulness of Health Information Systems. Stud Health Technol Inform. 2015;218:68-73.

92. Kushniruk AW, Borycki EM. Low-cost rapid usability engineering: designing and customizing usable healthcare information systems. Healthc Q. 2006;9(4):98-102.

93. Hill AB. The environment and disease: association or causation? Proc R Soc Med. 1965;58:295–300

94. Abbasi J. With Neuroimaging, Large NIH Study Could Shine a Light on the Adolescent Brain. JAMA. Published online August 14, 2019. doi:10.1001/jama.2019.9981.

95. Firth J, Torous J, Stubbs B, Firth JA, Steiner GZ, Smith L, Alvarez-Jimenez M, Gleeson J, Vancampfort D, Armitage CJ, Sarris J. The "online brain": how the Internet may be changing our cognition. World Psychiatry. 2019 Jun;18(2):119-29.

96. Faurhold-Jespen M, Frost M, Ritz C, Christensen EM, Jacoby AS, Mikkelsen RL, Knorr U, Bardram JE, Vinberg M, Kessling LV. Daily electronic self-monitoring in bipolar disorder using smartphones—the MONARCA I trial: a randomized, placebo-controlled, single-blind, parallel group trial. Psychological Medicine. 2015; 45(13):2691-2704.

97. Bradstreet S, Allan S, Gumley A. Adverse event monitoring in mHealth for psychosis interventions provides an important opportunity for learning. Journal of Mental Health. 2019 September; 28(5):461-466.

98. Larsen ME, Nicholas J, Christensen H. Quantifying App Store Dynamics: Longitudinal Tracking of Mental Health Apps. JMIR Mhealth Uhealth 2016; 4(3):e96.

99. Price, M., Lau, F. The clinical adoption meta-model: a temporal meta-model describing the clinical adoption of health information systems. BMC Med Inform Decis Mak 14, 43 (2014). https://doi.org/10.1186/1472-6947-14-43

100. Pratap A, Neto EC, Snyder P, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. *NPJ Digit Med.* 2020;3:21. Published 2020 Feb 17. doi:10.1038/s41746-020-0224-8