Engagement with mobile health interventions for depression: A systematic review☆

Anthony Molloy, Page L. Anderson[*]

Department of Psychology, Georgia State University, Urban Life Bldg, 11th Floor, 140 Decatur Street, Atlanta, GA 30303, USA

ARTICLE INFO

ABSTRACT

Background: Depressive disorders are a major public health problem, and many people face barriers to accessing evidence-based mental health treatment. Mobile health (mHealth) interventions may circumvent logistical barriers to in-person care (e.g., cost, transportation), however the symptoms of depression (low motivation, concentration difficulties) may make it difficult for people with the disorder to engage with mHealth.
Objective: The aim of this systematic review is to examine assessment and reporting of engagement in clinical trials of mHealth interventions for depression, including objective engagement (e.g., number of times program is used), subjective engagement (e.g., qualitative data on users' experiences), and associations between engagement and other clinically important variables (e.g., symptom improvement, participant characteristics).
Methods: Three electronic databases (PsycINFO, Web of Science, PubMed) were searched in February 2020 using search terms for mHealth and depression. Studies were included in the review if they tested a mHealth intervention designed for people with depressive disorders or elevated depression symptoms.
Results: Thirty studies met inclusion criteria and were reviewed. Most studies reported objective engagement ($N = 23$, 76.7%), approximately half reported subjective engagement ($N = 16$, 53.3%), and relatively few examined associations between engagement and clinical improvement, participant characteristics, or other clinically relevant variables ($N = 13$, 43.3%).
Conclusions: Although most studies in this small but rapidly growing literature report at least one measure of engagement, there is substantial heterogeneity. Intentional, theory-driven, and consistent measurement of engagement with mHealth interventions for depression may advance the field's understanding of effective engagement to facilitate clinical improvement, identify dose-response relationships, and maximize generalizability for underserved populations.

1. Introduction

Depressive disorders have an enormous impact on global health and quality of life, affecting over 250 million people worldwide, ranking as the third leading cause of global disability (James et al., 2018), and being associated with unemployment, poor physical health, poor social function, and suicide (Hawton et al., 2013; World Health Organization, 2017). There are effective medications and psychotherapies that improve depressive symptoms, but there are not enough trained mental health professionals to deliver them (Liu et al., 2017; World Health Organization, 2017).

Mobile health, or "mHealth," is viewed as a promising way to overcome well-documented barriers to in-person treatment and increase access to mental health services, particularly among underserved communities. mHealth refers to "medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants and other wireless devices" (van Heerden et al., 2012). Delivering treatment via smartphone creates a substantial opportunity to expand access to mental health treatment, as there are an estimated 3.5 billion smartphone users worldwide (Statista, 2019) and relatively low disparities in smartphone ownership along racial and socioeconomic lines in the U.S. (Pew Research Center, 2019).

Meta-analyses examining clinical trials of smartphone-based mHealth programs for depression have demonstrated that they

significantly reduce depression symptoms (Firth et al., 2017; Weisel et al., 2019), but attrition and low engagement with these programs are a significant concern. Studies of publicly accessible mHealth programs for mental health find that many people stop using these programs shortly after downloading them, before they are likely to achieve any clinical benefit (Lattie et al., 2016). This is broadly true of commercially available smartphone apps, which typically lose about 70% of users within one week of download (Sigg et al., 2016). A better understanding of factors that influence engagement in mHealth interventions for depression is needed to fully realize their potential.

Engagement with digital health interventions is a complex, multifaceted construct. Perski et al. (2017) conducted a systematic review from an interdisciplinary perspective to create a conceptual framework explaining how engagement with digital interventions leads to behavior change. Drawing from the computer science and behavioral science literatures, they define engagement with digital interventions as "the extent (e.g. amount, frequency, duration, depth) of usage and (2) a subjective experience characterised by attention, interest and affect" (p. 261). They emphasized that engagement can be understood and measured objectively, by recording user behavior, and subjectively, by evaluating self-reported qualitative dimensions of users' experiences while engaging with an intervention. The model also purports that engagement is influenced by the intervention itself (e.g., content, delivery mechanism) and by context, which includes individual characteristics of the population using the intervention and their sociocultural environment.

Depression is characterized by behavioral avoidance, difficulty concentrating, anhedonia, and negative cognitions (Beck, 2008), all of which could impact engagement with a mHealth intervention. Additionally, depressed people experience greater levels of social impairment, relationship dysfunction, unemployment, and medical comorbidities (McKnight and Kashdan, 2009), contextual factors that should not be ignored in clinical research. In order to understand and specifically target types of engagement that have the greatest impact on clinical improvement for people with depression, clinical researchers should select engagement metrics that shed light on interactions between individual characteristics, context, different types of engagement, and clinical improvement.

Systematic reviews that examine engagement reporting in clinical trials of mHealth programs have found substantial variety in how it is measured, which limits generalizability across studies and progress in this area. For example, Pham et al. (2019) outlined 14 engagement-related constructs that have been used by mHealth researchers (e.g. "use," "adherence," "compliance," "feasibility") across studies of mHealth programs for chronic health conditions. Reviews of mHealth for mental health find that researchers report engagement quite differently across studies (Linardon and Fuller-Tyszkiewicz, 2020; Ng et al., 2019). Additionally, it is uncommon for studies to evaluate relationships between engagement and clinical outcomes, participant characteristics, or other relevant variables, which limits researchers' ability to develop contextualized models of engagement for specific populations.

Researchers that capture engagement using a variety of different metrics can examine relationships between engagement, clinical outcomes, and participant characteristics, such as baseline depression severity or cultural background. These findings could inform and test theoretical models of engagement with mHealth engagement or clinical decisions about the appropriateness of specific mHealth programs for different populations. Comparing engagement between different mHealth interventions, examining changes in engagement over time, and examining associations between different metrics of engagement could inform mHealth program design and the ways that patients are instructed to use programs. It is therefore important to know whether clinical researchers consistently report engagement, the most common ways that engagement is operationalized, and extent to which researchers examine associations between engagement and other variables in clinical trials.

1.1. The current study

Engagement may pose a particular problem for individuals experiencing depression, but no review to date has specifically examined engagement reporting in studies of mHealth interventions for depression. The current systematic review examined measurement and reporting of engagement in clinical trials of these programs. Studies that did not report engagement were included to evaluate the consistency of engagement reporting in the literature. Both objective and subjective metrics of engagement for mHealth interventions were reviewed. Additionally, the review examined which studies tested for associations between metrics of engagement and other variables, given the potential for these associations to inform future research and implementation of mHealth interventions for depression. Findings are discussed as they relate to theoretical models for, improvement of clinical research on, and optimization of mHealth interventions for depression.

2. Methods

2.1. Electronic searches

A systematic review was conducted using the PsycINFO, PubMed, and Web of Science databases. After a review of the literature, search terms were developed for mobile devices, mHealth, and depression and entered on February 9th, 2020. See Appendix A for specific search terms. In keeping with previous systematic reviews of mobile interventions (Donker et al., 2013; Dubad et al., 2018), only studies published 2008 and afterward were included because this is the year that the first mobile applications became publicly available for download. The first author completed the electronic searches, removed duplicates, then screened titles and abstracts for inclusion and exclusion criteria. Following title and abstract review, full texts of articles that seemed to meet criteria based on titles and abstracts were then reviewed by both authors to reach final decisions about inclusion. Disagreements were resolved through in-depth discussion.

2.2. Data extraction

A data extraction form was developed by the first author based on recent systematic reviews of participant engagement in digital mental health interventions (Linardon and Fuller-Tyszkiewicz, 2020; Ng et al., 2019; Pham et al., 2019) and a preliminary review of articles that met inclusion criteria. For each study, the first author first extracted the methods of assessing depression (assessment for specific diagnosis or cutoff on a self-report measure) and the mobile device used for the study intervention (e.g., app). Interventions were then coded as "structured" if they used locked, sequential modules, "unstructured" if they used tools that can be accessed at any time, "hybrid" if they used structured and unstructured components, or "ecological momentary assessment" if they solely prompted users to complete brief assessments of mood or other constructs (EMA; see Shiffman et al., 2008). Information about the demographic characteristics of the sample (e.g., age, race), major components of mHealth programs (e.g., behavioral activation, cognitive restructuring), and presence or absence of coaching were also recorded. Lastly, the first author coded studies for the types of information they reported about user engagement. These are presented in Fig. 1 and fall into three broad categories: objective user engagement, subjective user engagement, and assessment of associations between engagement and other variables.

2.3. Study selection criteria

Studies were included if they met the following criteria: 1) original peer-reviewed articles, 2) published in English, 3) participants met criteria for a unipolar depressive disorder (e.g. major depressive disorder, persistent depressive disorder) as assessed by a structured interview

Objective Engagement Metrics:
- Total duration of use (e.g. used app for XX minutes total or XX minutes per week)
- Program use by day or week (e.g. used app at least once on XX days or XX weeks)
- Total number of sessions (i.e. "log-ins")
- Average duration between sessions
- Average duration of sessions
- Use of specific program features (e.g. used a specific tool XX times or for XX minutes)
- Context of use (e.g. times app was accessed at a specific location, time of day, etc.)
- Interaction with coach or therapist (e.g. number of calls, number of messages)
- Assessment of "active use" (number of features used compared to overall duration of app use)
- Completion of structured modules
- Adherence to usage instructions (e.g. XX% of participants used app as instructed)
- Response to EMA prompts

Subjective Engagement Metrics:
- Self-report measure
- Qualitative interview

Assessed relationship between engagement metrics and other variables:
- Comparison between multiple interventions
- Comparison across baseline participant characteristics (e.g. demographics, baseline psychopathology)
- Relationship with clinical improvement in response to intervention
- Relationship between multiple engagement metrics
- Statistical testing of changes in engagement over time

Fig. 1. Metrics of engagement examined in the current study.

or confirmed by medical records, or had elevated depression symptoms established by any cutoff on a validated self-report measure, and 4) examined a digital psychological intervention delivered via a mobile device (e.g. smartphone, tablet) that specifically targets depression and was intended to be used more than once. Studies were excluded for the following reasons: 1) did not report their assessment of depression or cite a resource with this information, such as a published study protocol, 2) included participants without depression (e.g. mixed samples with depression and/or anxiety), unless the non-depressed sample represented a separate study condition and was examined separately, 3) examined an intervention that functioned simply as a means of communication between user and therapist (e.g. videoconferencing, texting), 4) examined an intervention that did not target psychological symptoms (e.g. targeting only sleep or exercise), or 5) examined an intervention that requires no active input from the user, such as programs that exclusively use passive mobile sensors.

Studies examining EMA or mood tracking programs were included in the review when these were conceptualized as interventions, given the evidence that mood tracking alone has the potential to reduce depression symptoms (Dubad et al., 2018). Blended interventions containing a mobile component alongside other components (e.g. web-based intervention, face to face therapy) were included. Studies examining digital psychological interventions that could be completed without a mobile device (e.g., could be completed using a computer) were not included in the current review. These studies were excluded to focus on engagement metrics that are specifically relevant to mobile devices, which may differ from metrics used for interventions that are commonly accessed using desktop computers. Because metrics of engagement are informative in non-controlled studies, studies with and without a comparison group for the active treatment were included. Lastly, secondary analyses of primary studies were included given that these studies have the potential to report information about engagement that was not reported in the primary article.

3. Results

3.1. Study selection

A total of 4473 references were identified through the database search. After duplicates were removed, 3613 articles were reviewed by title and abstracts. The authors reviewed 289 full text articles and 30 were ultimately included in the systematic review. A Preferred Reporting Items for Systematic Reviews and Meta-Analyses (Moher et al., 2009) flow chart of the study selection process is presented in Fig. 2.

3.2. Study characteristics

Table 1 shows the various types of depression assessment, mobile devices, structured interventions, and coach support used in the studies included in this review. Details about individual studies, including the specific mHealth intervention used, participant demographics, whether and how objective/subjective engagement was measured, and whether the study assessed associations between engagement and other variables are presented in Table 2. The majority of studies were published recently, with nine published in 2018 (30.0%) and 12 published in 2019 (40.0%).

3.2.1. Participants
Depression was most commonly assessed using a cutoff on a validated self-report measure, either alone ($N = 16$, 53.3%) or in combination with a confirmed diagnosis of a unipolar depressive disorder ($N = 9$, 30.0%). A smaller proportion of studies selected participants based on diagnosis of a depressive disorder without a self-report measure ($N = 5$, 16.7%).

3.2.2. Interventions
Most studies examined smartphone interventions specific to iPhones ($N = 8$, 26.7%) or compatible with multiple operating systems ($N = 7$, 23.3%). A smaller proportion of studies used interventions specific to Android smartphones ($N = 3$, 10.0%) or examined interventions that were delivered on multiple devices (e.g. smartphone with smart watch or tablet; $N = 4$, 13.3%). A number of studies used smartphone interventions with unspecified operating systems ($N = 8$, 26.7%). Studies most commonly examined unstructured interventions ($N = 15$, 50.0%), followed by an equal number of studies that examined structured ($N = 5$, 16.7%), hybrid ($N = 5$, 16.7%), and EMA ($N = 5$, 16.7%) interventions. About a third of interventions were self-guided ($N = 11$, 36.7%) and the rest involved some level of support from a therapist or coach ($N = 19$, 63.3%).

3.3. Metrics of objective engagement

There was a high level of heterogeneity in reporting of objective engagement. Twenty-three studies (76.7%) reported at least one objective metric of engagement. Frequencies of reporting for all metrics of user engagement are presented in Table 3.

3.3.1. Program use by day or week
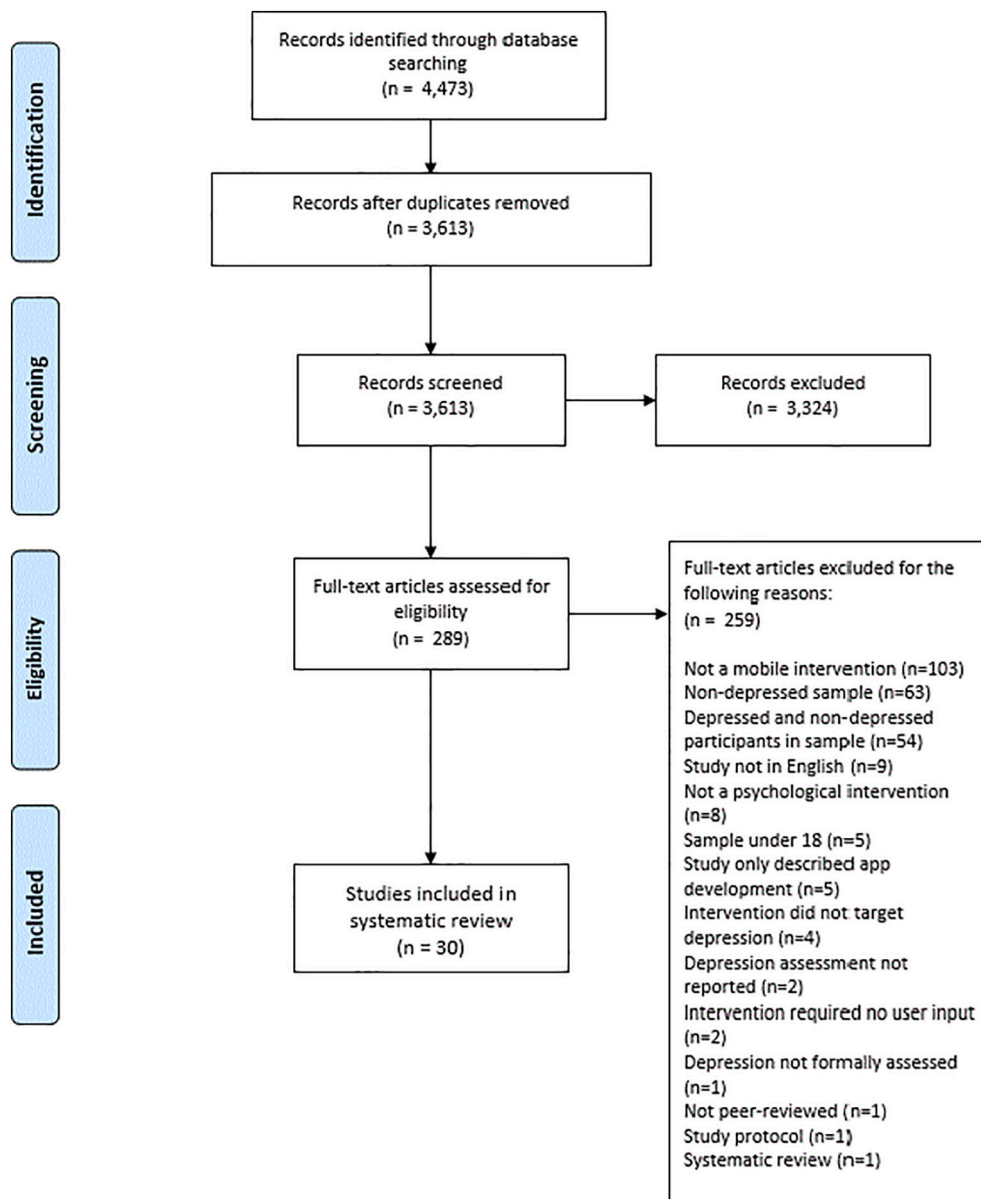This was the most commonly reported metric of user engagement in

Fig. 2. Preferred reporting items for systematic reviews and meta-analyses flow diagram.

the current review. Studies used different intervals for tracking; most studies tracked program use by the number of active days (i.e. number of days the program was used at least once; $N = 7$), whereas others reported by active week (i.e. number of weeks the program was used at least once; $N = 6$). One study of an unstructured intervention (Caplan et al., 2018) reported the number of participants that used the intervention "several times per week" as their sole objective metric of engagement.

3.3.2. Use of specific program features

Use of specific program features was also one of the most commonly reported objective engagement metrics in included studies. Generally, studies reported the number of times that participants used specific tools, such as setting goals for behavioral activation (Dahne et al., 2019a, 2019b), completing cognitive restructuring exercises (Stiles-Shields et al., 2019), or interacting with peers (Sawyer et al., 2019).

3.3.3. Total number of sessions

Five studies reported the average number of times that participants accessed the intervention. Burns et al. (2011) merged any "log-ins" to

their intervention that occurred within one hour of each other to avoid counting brief sessions that occurred in quick succession.

3.3.4. Interaction with coach or therapist

Of the 19 studies that examined coach or therapist-supported interventions, five studies reported at least one objective metric of interaction with a coach or therapist. There was substantial variety in the ways that coaching was delivered in these interventions and in how it was reported. Economides et al. (2019) reported the number of days that participants were in contact with a therapist and did not specify whether this contact was via messaging or phone (participants had access to both). Other studies reported the number of messages sent to a coach (Ly et al., 2014; Schlosser et al., 2017) or the average amount of time that participants spoke with coaches via phone (Stiles-Shields et al., 2019). Schlosser et al. (2017) examined the construct of "social initiative" by reporting the proportion of interactions between participants and coaches that were initiated by the participant.

3.3.5. Completion of structured modules

Four studies reported completion of structured modules as a metric

Table 1

Proportion of studies using various types of depression assessment, mobile device, structured interventions, and coach support.

Characteristic	N	%
Depression assessment		
Cutoff on self-report measure only	16	53.3
Depressive disorder diagnosis only	5	16.7
Depressive disorder diagnosis and cutoff on self-report measure	9	30.0
Mobile device		
iPhone	8	26.7
Android	3	10.0
Smartphone: Multiple OS	7	23.3
Smartphone: Unspecified OS	8	26.7
Other device or multiple devices	4	13.3
Structure of intervention		
Structured	5	16.7
Unstructured	15	50.0
Hybrid	5	16.7
Ecological Momentary Assessment (EMA)	5	16.7
Coaching support		
Coached	19	63.3%
Self-guided	11	36.7%

of engagement. Two of these were a primary study and secondary analysis that examined a structured intervention (Furukawa et al., 2018a; Mantani et al., 2017). One examined a hybrid intervention (Watts et al., 2013) and one examined an unstructured intervention (Menezes et al., 2019) that included regular behavioral activation sessions which were not in locked sequence.

3.3.6. Total duration of use

Four studies reported the total duration that participants used the study intervention. Duration was reported in average minutes or hours that the program was used per participant. Three studies reported total duration of use throughout the study, whereas one study reported total duration of use per week (Takahashi et al., 2019).

3.3.7. Response to EMA prompts

Of the five EMA studies included in this review, four reported the number of completed EMA prompts. One study reported this as its sole metric of objective engagement (Moukaddam et al., 2019), whereas three reported it in combination with other metrics (Cormack et al., 2019; Hung et al., 2016; Torous et al., 2015).

3.3.8. Average duration between sessions

Three studies reported the average duration between times that participants accessed an intervention (Furukawa et al., 2018a; Mantani et al., 2017; Menezes et al., 2019). All of these studies also reported completion of structured modules and average duration between participants' completion of structured modules.

3.3.9. Average duration of sessions

Three studies examining unstructured and structured interventions reported the average duration of use whenever a participant opened the program (Dahne et al., 2019a, 2019b; Furukawa et al., 2018a). Two of these studies examined similar behavioral activation apps, one of which was adapted for delivery in Spanish.

3.3.10. Adherence to usage instructions

Two studies examining unstructured interventions reported the proportion of participants who adhered to specific recommendations for program usage (Arean et al., 2016; Takahashi et al., 2019). Arean et al. (2016) categorized participants into "none," "suboptimal," and "optimal" usage groups depending on the number of weeks that they used the intervention as instructed. Arean et al. (2016) also reported total number of sessions, whereas Takahashi et al. (2019) also reported average total duration of use per week.

3.3.11. Context of use

Two EMA studies reported the context in which participants responded to EMA prompts (Cormack et al., 2019; Torous et al., 2015). Both studies examined the proportion of prompts to which participants responded across morning, afternoon, and night. Both of these studies also reported overall percentage of response to EMA prompts and the number of days the program was used.

3.3.12. Assessment of "active use"

Schlosser et al. (2017) was the only study in the current review that specifically quantified the extent of participant activity within their intervention as compared to overall duration of use. The authors calculated an "active use rate" by comparing participants' posts, comments, and interactions with coaches and peers within the intervention to the amount of time that participants used it.

3.4. Metrics of subjective engagement

The majority of studies reported at least one metric of subjective user engagement ($N = 16$, 53.3%). Studies that reported subjective engagement used self-report measures ($N = 15$, 50.0%) or qualitative interviews with participants ($N = 5$, 16.7%).

3.4.1. Self-report measures

Fifteen studies used a self-report measure to examine some aspect of participants' subjective experience of an intervention. There was substantial heterogeneity in these measures. Some studies used validated questionnaires like the Credibility Expectancy Questionnaire (Devilly and Borkovec, 2000), User Engagement Scale (O'Brien and Toms, 2010), System Usability Scale (Brooke, 1996), and others. These measures assess a range of constructs including outcome expectancy, focused attention, perception of time during use, and satisfaction. Other studies used questions that were developed by the researchers. Most studies examined subjective engagement at the end of the study, but several assessed it at multiple time points. For example, Caplan et al. (2018) administered three questions about usefulness of their program every two days throughout their study.

3.4.2. Qualitative interviews

Five studies used semi-structured, open-ended qualitative interviews to examine subjective engagement. All studies described highlights of user feedback, although interview content was reported in varying levels of detail. Several studies reported highly detailed interview content, organized content into themes, and included direct quotes from participants.

3.5. Assessment of association between engagement and other variables

Fewer than half of the reviewed studies assessed associations between engagement and other variables ($N = 13$, 43.3%). These studies assessed associations between engagement and the following variables, in order of frequency: clinical improvement ($N = 9$, 30.0%), baseline participant characteristics ($N = 6$, 20.0%), comparison across multiple mobile interventions ($N = 4$, 13.3%), changes in engagement over time ($N = 2$, 6.7%), and association between multiple engagement metrics ($N = 1$, 3.3%).

3.5.1. Clinical improvement

Nine studies examined the association between engagement and participants' clinical outcomes in response to an intervention. Many of these studies used complex statistical models to assess for associations. For example, Economides et al. (2019) used multiple regression models to examine the impact of several objective engagement metrics on symptom reduction. Others categorized participants into responders and non-responders and compared engagement between these groups (Dahne et al., 2019b; Furukawa et al., 2018a). Overall, four studies

Table 2

Characteristics of individual studies.

First author, year	mHealth programs	Key components and treatment target	Sample size	Sample demographics	Objective engagement	Subjective engagement	Assessed association between engagement and other variables
Arean et al. (2016)	Project: EVO iPST	Uses video games designed to increase cognitive control Uses principles of problem-solving therapy to assist with goal-setting and action plans	626	Mean age = 33.95 (SD 11.84); 79.0% Female; 13.7% African-American, 1.0% American Indian, 8.6% Asian, 65.5% White, 10.5% > 1 race, 0.6% Native Hawaiian/Pacific Islander, 12.6% Hispanic	Adherence to instructions; Total number of sessions	None	Compared interventions; Participant characteristics
Burns et al. (2011)	Mobilyze!	Uses behavioral activation strategies, EMA, ecological momentary intervention cued by passive mobile phone sensors, behavioral skills training, didactic content	8	Mean age = 37.4 (SD 12.2); 87.5% female; 13% Hispanic Caucasian, 88% Non-Hispanic Caucasian	Total number of sessions	Self-report measure; Qualitative Interview	None
Caplan et al. (2018)	El Buen Consejo Movil	Provides self-help audio messages based on cognitive-behavioral therapy, encourages social engagement using group forum with messaging and user mood ratings	36	Sample 1: Mean age = 36; 83% Female; 78% from Dominican Republic, 16% from Venezuela, 6% from United States; Sample 2: Mean age = 42; 86% Female; 100% from Dominican Republic	Use by day or week	Self-report measure; Qualitative Interview	None
Cormack et al. (2019)	Cognition Kit	Uses EMA for regular assessment of mood and cognitive function	30	Mean age = 37.2 (SD 10.4); 63.3% Female; Race/Ethnicity not reported	Use by day or week; EMA Prompts; Context of use	Qualitative Interview	Engagement over time; Participant characteristics
Dahne et al. (2018)	Behavioral Apptivation	Uses behavioral activation strategies in conjunction with face-to-face therapy	11	Mean age = 24.91 (SD 11.73); 90.9% Female; 45.50% White, 18.20% Black, 27.30% Asian, 9.10% Other	None	Self-report measure	None
Dahne et al. (2019a)	¡Aptívate!; iCouch CBT	Uses behavioral activation strategies, mood monitoring, and provides social support Uses cognitive restructuring techniques to cope with stressful situations	42	Mean age = 36.05 (SD 11.44); 66.7% Female; 23.8% White, 2.4% Black, 2.4% Native Hawaiian/ Pacific Islander, 7.1% Native American, 11.9% Multiracial, 52.4% Other, 100% Hispanic ethnicity	Total number of sessions; Average session duration; Total duration of use; Use of specific features; Use by day or week	None	Compared interventions; Participant characteristics
Dahne et al. (2019b)	Moodivate Moodkit	Uses behavioral activation strategies, mood monitoring, and provides social support Uses cognitive restructuring techniques to cope with stressful situations	52	Mean age = 43.79 (SD 13.27); 84.6% Female; 40.4% White, 55.8% Black, 3.8% Other, 3.8% Hispanic ethnicity	Total number of sessions; Average session duration; Total duration of use; Use of specific features; Use by day or week	None	Clinical Improvement
Economides et al. (2019)	Ascend	Sequential modules teach skills drawn from mindfulness-based stress reduction, mindfulness-based cognitive therapy, and cognitive-behavioral therapy	197	Mean age = 32.9 (SD 10.3); 77.5% Female; 78.4% from Finland, 21.6% from United States	Use by day or week; Total duration of use; Interaction with coach	None	Clinical Improvement; Engagement over time
Fuller-Tyszkiewicz et al. (2018)	BlueWatch	Sequential modules teach skills drawn from cognitive-behavioral therapy including behavioral activation, cognitive restructuring, and problem-solving	5	Mean age = 22.4 (SD 2.71); 80% Female; Race/Ethnicity not reported	None	Self-report measure; Qualitative Interview	None
Furukawa et al. (2018a)	Kokoro app	Sequential modules teach skills drawn from cognitive-behavioral therapy including thought recording, behavioral activation, and cognitive restructuring	164	Mean age = 40.2 (SD 8.8); 57% Female; Race/Ethnicity not reported	Complete structured modules; Duration between sessions; Use of specific features; Average session duration	None	Clinical Improvement
Furukawa et al. (2018b)	Kokoro app	Sequential modules teach skills drawn from cognitive-behavioral therapy including thought	78	Mean age = 40.4 (SD 8.8); 56.4% Female; Race/ Ethnicity not reported	Use of specific features	None	Clinical Improvement

Table 2 (*continued*)

First author, year	mHealth programs	Key components and treatment target	Sample size	Sample demographics	Objective engagement	Subjective engagement	Assessed association between engagement and other variables
		recording, behavioral activation, and cognitive restructuring					
Hantsoo et al. (2018)	Mood Tracking and Alert app (MTA)	Uses EMA for regular assessment of activity and mood, prompts mental healthcare provider to contact participant if symptoms worsen	72	Sample 1: Mean age = 26.3 (SD 4.9); 100% Female; 96% African-American, 11% Hispanic ethnicity; Sample 2: Mean age = 26.5 (SD 6.2); 100% Female; 95% African-American, 10% Hispanic ethnicity	Use by day or week	Self-report measure	None
Hung et al. (2016)	iHOPE	Uses EMA for regular assessment of depression, anxiety, sleep quality, and cognitive functioning	54	Mean age = 37.9 (SD 13.9); 63% Female; Race/Ethnicity not reported	Use by day or week; EMA Prompts	None	Participant characteristics
Hur et al. (2018)	Todac Todac	Uses brief vignettes and quizzes to teaches cognitive behavioral strategies, promotes social engagement with other users with a "timeline" feature	34	Mean age = 23.71 (SD 3.26); 88.2% Female; Race/Ethnicity not reported	None	None	None
Inkster et al. (2018)	Wysa	Uses an AI-driven chatbot to teach strategies based on positive psychology	129	No demographics reported	Use by day or week; Use of specific features	Self-report measure	Participant characteristics; Clinical Improvement
Li et al. (2019)	Run4Love	Sequential modules teach techniques from cognitive behavioral stress management, target behavioral activation by promoting exercise	300	Mean age = 27.5; 7.7% Female; Race/Ethnicity not reported	None	None	None
Ly et al. (2014)	"BA treatment"	Uses selection and tracking of pleasurable activities to promote behavioral activation	81	Mean age = 36.1 (SD 10.8); 70% Female; Race/Ethnicity not reported	Use by day or week; Interaction with coach	Self-report measure	Compared interventions; Clinical Improvement
	"Mindfulness treatment"	Uses audio tracks to teach mindfulness skills					
Ly et al. (2015)	"Blended BA treatment"	Uses selection and tracking of pleasurable activities to promote behavioral activation, blended with in-person behavioral activation-based therapy	93	Mean age = 30.6 (SD 11.4); 69.9% Female; Race/Ethnicity not reported	None	Self-report measure	None
Mantani et al. (2017)	Kokoro app	Sequential modules teach skills drawn from cognitive-behavioral therapy including thought recording, behavioral activation, and cognitive restructuring	164	Sample 1: Mean age = 40.2 (SD 8.8); 57% Female; Race/Ethnicity not reported; Sample 2: Mean age = 41.6 (SD 8.9); 50% Female; Race/Ethnicity not reported	Complete structured modules; Duration between sessions; Use of specific features	None	None
Menezes et al. (2019)	CONEMO	Uses sequential sessions to increase pleasurable and healthy activities to promote behavioral activation	66	Age: 6% 21–40, 53% 41–60, 41% ≥ 61; 71% Female; Race/Ethnicity not reported	Complete structured modules; Duration between sessions	Self-report measure	None
Moukaddam et al. (2019)	SOLVD	Uses EMA for regular assessment of mood and anxiety, passively collects smartphone data	25	Mean age = 50.28 (SD 10.07); 76% Female; 40.9% White, 36.4% African American, 18.2% Hispanic, 4.5% Asian	EMA Prompts	None	None
Pratap et al. (2018)	Project: EVO	Uses video games designed to increase cognitive control	1040	Mean age = 34.9 (SD 10.92); 77.19% Female; 53.3% Non-Hispanic White, 30.7% Hispanic/Latino, 7.2% African-American/Black, 0.9% American Indian/Alaskan Native, 7.0% Asian, 0.9% Other	None	None	None
	iPST	Uses principles of problem-solving therapy to assist with goal-setting and action plans					
Sawyer et al. (2019)	eMums Plus	Uses sequential modules to teach strategies drawn from cognitive behavioral therapy, provides education on child development and	133	Mean age = 31.1 (SD 5.0); 100% Female; Race/Ethnicity not reported	Use by day or week; Use of specific features	Self-report measure	None

Table 2 (*continued*)

First author, year	mHealth programs	Key components and treatment target	Sample size	Sample demographics	Objective engagement	Subjective engagement	Assessed association between engagement and other variables
		parenting, uses social media feature to promote social engagement with nurses and other mothers of young children					
Schlosser et al. (2017)	PRIME-D	Uses social platform to track and share goals related to health, relationships, creativity, and productivity, promotes social engagement with other users	36	Mean age = 31.33 (SD 12.4); 77.8% Female; 61.1% Caucasian, 19.5% African American, 8.3% Asian American, 11.1% Other, 83.3% Non-Hispanic ethnicity, 16.7% Hispanic ethnicity	Use by day or week; Use of specific features; Interaction with coach; Assessed active use	Self-report measure; Qualitative Interview	Participant characteristics; Clinical Improvement
Schuster et al. (2019)	MindDistrict	Uses activity scheduling to promote behavioral activation, blended with in-person ACT-based therapy	27	Mean age = 37.70 (SD 13.66); 51.9% Female; Race/Ethnicity not reported	Use of specific features	Self-report measure	None
Stiles-Shields et al. (2019)	Boost Me	Uses activity scheduling mood monitoring to promote behavioral activation	30	No demographics reported	Total number of sessions; Use of specific features; Interaction with coach	Self-report measure	Compared interventions; Clinical Improvement; Other engagement metrics
	Thought Challenger	Uses cognitive restructuring techniques					
Takahashi et al. (2019)	SPSRS	Uses videos and positive words to promote behavioral activation	22	Mean age = 20 (SD 0.62); 27.3% Female; Race/Ethnicity not reported	Total duration of use; Adherence to instructions	Self-report measure	None
Torous et al. (2015)	Mindful Moods	Uses EMA for regular assessment of mood	13	Female mean age = 35 (SD 13); Male mean age = 48 (SD 16); 77% Female; Race/Ethnicity not reported	Use by day or week; Context of use; EMA Prompts	None	None
Watts et al. (2013)	Get Happy	Uses sequential modules containing stories and homework assignments to teach cognitive behavioral strategies, interpersonal skills, and sleep hygiene	35	Mean age = 41 (SD 12.38); 80% Female; Race/Ethnicity not reported	Complete structured modules; Interaction with coach	Self-report measure	Clinical Improvement
Zhu et al. (2019)	Run4Love	Sequential modules teach techniques from cognitive behavioral stress management, target behavioral activation by promoting exercise	300	Median age = 27.5; 7.7% Female; Race/Ethnicity not reported	None	None	None

Note. EMA = Ecological Momentary Assessment.

found a statistically significant positive association between engagement and clinical improvement. Furukawa et al. (2018a) found that "beneficiaries" (i.e. participants with greater clinical improvement) logged more behavioral activation activities within the study app, completed specific behavioral activation activities at different rates, reported higher levels of mastery and pleasure during behavioral activation, and completed a higher number of cognitive restructuring exercises than "nonbeneficiaries." Using data from the same study, Furukawa et al. (2018b) found that completed behavioral activation activities with greater mastery and pleasure ratings were associated with greater clinical improvement. Inkster et al. (2018) split participants into "high use" and "low use" based on the number of times participants accessed the study app and found that "high use" participants had greater clinical improvement. Schlosser et al. (2017) found positive relationships between clinical improvement and active use of the app as well as interaction with a coach.

3.5.2. Baseline participant characteristics

Six studies examined the association between engagement and participant characteristics at baseline. Studies typically assessed associations with either demographics or baseline psychopathology. Five of the six studies found at least one statistically significant association between an engagement metric and a baseline participant characteristic.

Arean et al. (2016) found that participants with higher baseline depression and anxiety accessed their two study apps less frequently, whereas participants with higher baseline disability accessed the apps more frequently. They also found an interaction between app condition and marital status on engagement, such that married participants were less likely to open an app based on problem-solving therapy as compared to an app designed to improve cognitive control. Dahne et al. (2019a) recruited local participants from primary care clinics and remote participants using advertisements on social media. They found that remote participants demonstrated less engagement across multiple objective metrics as compared to participants who were recruited locally. Hung et al. (2016) found that participants with more restrictive smartphone data plans used the study app on more days than people with more generous or unlimited data plans. Inkster et al. (2018) conducted a thematic analysis of qualitative user feedback for their app and found more favorable feedback from participants who found it "hard to cope with daily tasks" and who reported recent relationship problems. Schlosser et al. (2017) found that female participants accessed their intervention significantly more often than men.

3.5.3. Comparison across multiple mobile interventions

Four studies examined multiple mobile interventions and assessed for differences in engagement between intervention conditions. Three of

Table 3

Engagement reporting.

Characteristic	N	%
Objective engagement		
None	7	23.3
Program use by day or week	12	40.0
Use of specific program features	10	33.3
Total number of sessions	5	16.7
Interaction with coach or therapist	5	16.7
Completion of structured modules	4	13.3
Total duration of use	4	13.3
Response to EMA prompts	4	13.3
Average duration between sessions	3	10.0
Average duration of sessions	3	10.0
Adherence to usage instructions	2	6.7
Context of use	2	6.7
Assessment of "active use"	1	3.3
Subjective engagement		
None	14	46.7
Self-report measure	15	50.0
Qualitative interview	5	16.7
Assessed association between engagement and other variables		
None	17	56.7
Clinical improvement	9	30.0
Baseline participant characteristics	6	20.0
Compared between multiple interventions	4	13.3
Engagement over time	2	6.7
Multiple engagement metrics	1	3.3

Note. Categories are not mutually exclusive except for "None."

the four studies found a statistically significant difference on at least one engagement metric between two interventions. Arean et al. (2016) tested for condition-by-baseline variable interactions and found that differences in usage between two smartphone apps were significantly associated with participant characteristics. Specifically, married participants were relatively less likely to use a problem-solving therapy app at least once, baseline depression was associated with relatively lower likelihood of using a cognitive control app at least once, and higher alcohol use was associated with relatively lower use of a cognitive control app. Dahne et al. (2019a) found that participants self-reported more frequent usage of a Spanish-language behavioral activation app as compared to a Spanish-language cognitive restructuring app. Stiles-Shields et al. (2019) found that a behavioral activation app was opened more often but rated as less usable than a cognitive restructuring app.

3.5.4. Changes in engagement over time

Two studies statistically tested for changes in engagement over time. Economides et al. (2019) found that participants used their hybrid intervention on fewer days and contacted their therapist less frequently as more time elapsed from baseline. Similarly, Cormack et al. (2019) found that participants responded to fewer EMA prompts as more time elapsed from baseline.

3.5.5. Association between engagement metrics

One study examined the association between engagement metrics. Stiles-Shields et al. (2019) tested for an association between the number and duration of coach calls and metrics of program usage. They found no significant associations.

4. Discussion

This systematic review of clinical trials of mHealth interventions for depression found that the majority of studies reported at least one objective (77%) or subjective (53%) measure of engagement, but that the specific metrics used varied widely across studies. These results are consistent with previous reviews of mHealth interventions for a variety of mental health concerns (Linardon and Fuller-Tyszkiewicz, 2020; Ng et al., 2019). This variability may prove to be a significant barrier to understanding engagement with these programs for people with

depression. Relatively few studies tested for associations between engagement and other clinically relevant variables, such as clinical improvement ($N = 9$; 30%), participant characteristics ($N = 6$; 20%) or differences in engagement between interventions ($N = 4$; 17%), changes in engagement over time ($N = 2$; 7%) or associations between engagement metrics ($N = 1$; 3%). The literature on measuring and reporting engagement with mHealth for depression is still in its infancy. What follows is a series of tentative conclusions based on a synthesis of results from the review and suggestions to improve engagement reporting in clinical trials in order to make progress toward multi-dimensional, contextualized models of engagement with mHealth for people with depression. For a list of the specific recommendations discussed below, see Fig. 3.

4.1. All objective measures of engagement are not created equal

Objective engagement was most commonly measured by reporting program use by day or week and use of specific program features. Two studies reported program use by day or week as their sole metric of engagement (Caplan et al., 2018; Hantsoo et al., 2018), which is likely to be insensitive to a substantial amount of potential variability in user activity. Conversely, use of specific program features is an excellent metric of engagement because it provides both a sensitive assessment of usage and qualitative information about the most popular features of a program. Many mHealth interventions are complex and multifaceted, so understanding which aspects of a program participants use is crucial information for program development or detailed assessments of clinical efficacy.

An innovative objective measure of engagement quantified "active" and "passive" use by comparing the amount of activity within the intervention to overall duration of use and found that active use was related to clinical improvement, but passive use was not (Schlosser et al., 2017). This is crucial because it demonstrates that longer engagement with a program may be ineffective or inefficient if a large proportion of that use is passive. In the same study, the "social initiative" of users was operationalized by the proportion of peer interactions within the program that were initiated by each user. An objective metric of engagement of social initiative could test questions about achievement of behavior change via social learning and social modeling theories within mHealth programs. For example, users that observe others initiating social contact within mHealth programs and then subsequently initiate social contact themselves support a social-cognitive model of mHealth engagement for programs that use these features. It also reflects greater motivation and social functioning, which are common deficits in depression and important potential mechanisms of improvement.

Examining the time of day that people with depression use mHealth (Cormack et al., 2019; Torous et al., 2015) is another helpful objective engagement metric because sleep disruption is a core symptom of depressive disorders (Nutt et al., 2008). It is possible that as people improve, they will use mHealth more during the day than at night. Such a metric could also be used to test whether people use mHealth during times that traditional mental health providers are typically unavailable (i.e. outside business hours), suggesting that mHealth programs overcome logistical barriers to care for people with unmet mental health needs (Su & Anderson, under review).

4.2. Subjective feedback contextualizes objective measures of engagement, but it is less widely used

Approximately half of studies (53%) measured subjective engagement, which was less commonly measured than objective engagement. This disparity has been observed previously in digital mental health research, despite findings that subjective engagement with digital interventions can sometimes be more strongly associated with clinical improvement than objective metrics (Graham et al., 2021). A small number of studies ($N = 5$; 17%) included open-ended qualitative

- Report varied metrics of objective and subjective engagement
- Collect open-ended feedback to contextualize and better understand patterns of engagement
- Engagement metrics should capture as much detailed information as possible
- Examine associations between engagement and clinical improvement
- Examine associations between engagement and participant characteristics, including specific depression symptoms, demographics, and cultural factors
- Evaluate macro-engagement, i.e. broader behavior change associated with mHealth
- Compare engagement across different types of mHealth interventions
- Develop models of "effective engagement" for specific populations and interventions
- Examine dose-response relationships for engagement with mHealth
- Experimentally manipulate engagement to examine causal relationships

Fig. 3. Recommendations for future research on engagement with mhealth interventions.

feedback from participants. This represents a significant limitation of the literature, as qualitative feedback can explain and contextualize patterns of objective engagement. For example, participants completed fewer mood assessments to "train" an ecological momentary intervention app for depression over the course of a clinical trial (Burns et al., 2011). During semi-structured interviews, participants reported that they would have completed more ratings later in the trial if the mHealth intervention had provided more prompts. This feedback points to an actionable strategy to sustain engagement that could be tested in future research. A decline in participation may reflect well-documented deficits in memory and executive functioning among people with depression (Rock et al., 2014) and may function as a specific barrier to sustained engagement with mHealth for this population.

Subjective data is key for developing culturally responsive interventions for depression. Caplan et al. (2018) assessed the experiences of depressed low-SES adults in the Dominican Republic with a Spanish-language mHealth intervention, which informed cultural adaptations to their mHealth program. For example, the researchers learned that feelings of depression were frequently expressed as anger in their sample of Dominican adults. This information was used to develop animations that depicted the relationship between depression and anger, which were well-received by participants. These examples underscore the importance of measuring subjective engagement more consistently in research on mHealth for depression.

4.3. Engagement is not consistently associated with clinical improvement

A tentative, yet important take-away is that engagement with mHealth interventions is not consistently associated with clinical improvement among people with depression (at least as measured in the studies included in this review). Only four of the nine studies examining the relationship between some form of engagement and clinical improvement found that greater engagement was associated with greater reduction in depressive symptoms. Although it is possible that there is no relation between how people with depressive symptoms engage with mHealth interventions and clinical improvement, it seems unlikely. Furthermore, it is not best practice to 'count studies' in systematic reviews in support of a conclusion. It is therefore imperative to develop and test models of engagement to maximize benefit from mHealth interventions for depression.

Furukawa et al. (2018a) measured engagement extensively and found a number of interesting differences between "responders" and "non-responders" to their behavioral activation intervention. For example, they found that responders logged a greater number of behavioral activation activities, reported greater levels of mastery and pleasure, and tended to select activities with longer durations. This information is highly valuable because it allows for inferences about "macro-engagement" (Yardley et al., 2016), i.e. broader behavior change associated with using a mHealth intervention. Macro-engagement is particularly important for behavioral activation,

because success in this intervention is contingent on completing activities that provide positive reinforcement (Cuijpers et al., 2007). Many of the studies in this review targeted behavioral activation and reported the number of activities that participants logged during the study, but these studies typically did not examine the relationship between frequency or type of activities and clinical improvement. Future mHealth studies, particularly those examining behavioral activation apps, can be improved by consistently examining the effects of macro-engagement on clinical improvement and measuring macro-engagement directly, as opposed to relying on self-reported data. Additionally, researchers may improve their precision by analyzing engagement as a continuous variable, as opposed to arbitrary groupings of "high" and "low" users.

4.4. Engagement is associated with demographic characteristics and other individual differences

In contrast to the relatively small number of studies that found associations between engagement and clinical improvement, each study examining engagement and baseline participant characteristics (with one exception) found significant associations. Studies that examine these questions are valuable to inform selection and tailoring of mHealth interventions to account for personal characteristics and sociocultural context. For example, Schlosser et al.'s (2017) finding that women accessed their app more often than men could reflect masculine cultural norms in the U.S. that stigmatize help-seeking (Vogel et al., 2011), a barrier that could be addressed to improve initiation and engagement with mHealth among men with depression. Two studies in this review provided useful information about the interaction between participants' social context and the types of mHealth they may find most engaging. Arean et al. (2016) found that married participants were less likely to open a problem-solving therapy app as compared to a cognitive training app. This could be because married participants receive more social support and assistance with problem-solving than single participants, making a problem-solving intervention less appealing. Inkster et al. (2018) found that participants who endorsed relationship problems provided more positive feedback for a conversation agent-based app, which could reflect that interventions which simulate social interactions are more engaging for individuals with social isolation and impairment, which are common in depression. These findings demonstrate the value of measuring and examining specific symptoms of depression, such as social impairment, and the ways that they are associated with engagement. As another example, Hung et al. (2016) found that participants with limited data cell phone plans used their app more frequently, which they attributed to the fact that their app could be used offline. This feature could be easily incorporated into mHealth apps to improve mental health equity and increase access across socioeconomic lines. Continued attention to these questions will be critical in future research, which should thoroughly evaluate the impact of individual differences on engagement across diverse participants. Researchers should also collect detailed qualitative data whenever possible to aid interpretation

of engagement patterns and minimize the need for speculation. This will be particularly important for understanding the needs of marginalized minority groups, who are underrepresented in research and stand to benefit the most from mHealth because of lower access to mental health services.

4.5. Engagement can vary across types of mHealth interventions

For example, Stiles-Shields et al.'s (2019) comparison between a behavioral activation and cognitive restructuring app is particularly interesting, because it demonstrates the potential for divergent, distinct profiles of engagement between different interventions. They found that participants launched a behavioral activation app more frequently, but rated a cognitive restructuring app as more usable at mid-treatment. Participants using the cognitive restructuring app also demonstrated clinically significant improvement as compared to a waitlist control, which was not observed for the behavioral activation app despite significantly greater use for this app. This profile of objective engagement, subjective engagement, and clinical improvement between multiple interventions provides many directions for future research, due to a study design that directly compared engagement across interventions. Stiles-Shields et al. also directly examined associations between several of their engagement metrics. This line of research could inform strategies to increase engagement, because a strategy that targets one specific type of engagement may lead to greater clinical benefits if it also affects other types of engagement that are interrelated.

4.6. Developing a model of 'effective engagement' for mHealth interventions among people who are depressed

"Effective engagement" refers to the functional importance of various types of engagement with digital health interventions among specific populations to achieve specific outcomes (Yardley et al., 2016). To develop models of "effective engagement" for digital health interventions for specific populations, researchers should measure both objective and subjective engagement within these populations and examine relationships between engagement and users' personal characteristics and sociocultural context to deepen understanding of engagement over the course of treatment. This can inform strategies to increase the most effective forms of engagement with specific interventions, while ensuring that these programs are effective and engaging for the population of interest and for minority groups that face well-documented barriers to healthcare and perhaps stand to benefit the most from mHealth. Using the construct of 'effective engagement' could help researchers of mHealth for depression choose objective and subjective measures of engagement for specific populations, examine associations with specific outcomes (e.g., clinical improvement), test theoretical models of engagement, and personalize mHealth for depression. Researchers have begun to develop theoretical models that include engagement as a mechanism of improvement for mental health interventions, which is a promising step toward developing interventions that effectively engage users to maximize symptom reduction (Graham et al., 2019).

4.7. Identifying minimal and optimal doses of mHealth interventions for people who are depressed

The dose-response relationship is a widespread concept in medical research, including mHealth (Perski et al., 2017). Understanding the association between the "dose," or level of engagement, and reduction of symptoms should be a major goal of mHealth research. Many of the studies in this review have demonstrated that mHealth programs can effectively treat depression using a range of different strategies, including behavioral activation, cognitive techniques, mindfulness, and facilitating social engagement. However, little is known about which specific types of engagement have the strongest relationships with

clinical success. The relationship between engagement (i.e. dose) and clinical response may also vary between interventions and populations. Measuring and reporting how engagement interacts with personal characteristics and context across various populations will be important for defining 'effective engagement', allowing for personalized evidence-based recommendations for users and mental health professionals.

4.8. Strengths and limitations

This is the first systematic review of engagement with mHealth for depression, which advances the literature because it focuses on a specific population that, by nature of the disorder, would be expected to have difficulty engaging with these interventions. The review is a step toward understanding 'effective engagement' with mHealth interventions, which will help these interventions fulfill their promise of improving access to, utilization of, and benefit from science-based interventions, as well as their potential to improve mental health equity. Strengths of this review include a systematic approach and comprehensive set of search terms. The review also included a range of different types of mHealth programs, including EMA programs, which capture a broad picture of the mHealth literature.

This review also has several limitations. Although both authors participated in full-text review and selection of included articles, the first author independently conducted title and abstract review as well as data extraction. Because there was substantial heterogeneity in engagement reporting across included studies, the categories used for data extraction did not capture some important distinctions, e.g. the specific constructs assessed by self-report measures of subjective engagement. Internet-delivered programs were excluded from the review, but may be accessed via mobile devices and thus have similar patterns of engagement. Further research should address the potential similarities between engagement for Internet-based treatments and mHealth-only interventions. Most studies in this review reported data from samples that were disproportionately female, and a number of studies did not report the race or ethnicity of their samples. This raises questions about the generalizability of these studies to men who experience depression and racial and ethnic minorities. Additionally, the authors did not conduct a meta-analysis of the associations between engagement metrics and other variables due to an insufficient number of studies that examined these associations and high heterogeneity of reported engagement metrics. Accordingly, firm conclusions should not be drawn about statistically significant associations in individual studies.

Importantly, all of the studies that examined relationships between engagement and other factors in the current review did so observationally, which does not allow for inferences about causal relationships. Finding ways to experimentally manipulate engagement with mHealth for individuals with depression will be a valuable next step for clinical trials. Researchers and clinicians could draw from theories of learning and persuasion to experimentally test strategies that improve engagement with programs (Molloy et al., 2021), then examine whether these types of engagement significantly mediate clinical outcomes. For example, interventions could be designed to encourage adherence to recommendations using prompts, "gamification," and other persuasive design features (Kelders et al., 2012). This is critical not only for testing strategies that promote engagement, but also to address probable confounding variables in correlations between engagement and clinical outcomes, such as motivation and executive functioning.

4.9. Conclusion

The potential for mHealth interventions to reduce depression is limited by the fact that people who could benefit from them often do not engage with them. Research on engagement with mHealth interventions for depression is beginning; the majority of studies included in this review were published within the last two to three years. The review shows there is high heterogeneity among studies in reporting

engagement, which represents an opportunity for researchers to carefully consider and use the types of engagement metrics that will lead to a better understanding of effective engagement with mHealth interventions for people who are depressed. The authors recommend that future researchers measure and report a combination of objective and subjective engagement metrics and test for associations between these metrics and variables that are functionally important, such as clinical improvement and participant characteristics, which will assist in testing models of effective engagement in developing mHealth interventions for depression for diverse populations.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Dr. Amanda Draheim, Langting Su, Donovan Ellis, and other members of the Anxiety Research and Treatment Lab and for their helpful feedback over the course of this project.

Appendix A. Search terms

A.1. PsycINFO

(smartphone OR "smart phone" OR "cell phone" OR "cellular phone" OR "mobile device" OR "mobile phone" OR "personal digital assistant" OR "iPhone" OR "mobile app*" OR "phone app*" OR mHealth OR "m-health" OR "mobile health" OR eHealth OR "e-health" OR "eMental health" OR "eTherap*" OR "digital behavior change intervention" OR "Information and communications technology" OR "Behavioral intervention technology" OR "Digital intervention" OR "Digital health intervention")
        AND
(Depress* OR "affective disorder" OR "mood disorder" or MDD OR "affective symptoms")

A.2. Web of science

TS=(smartphone OR "smart phone" OR "cell phone" OR "cellular phone" OR "mobile device" OR "mobile phone" OR "personal digital assistant" OR "iPhone" OR "mobile app*" OR "phone app*" OR mHealth OR "m-health" OR "mobile health" OR eHealth OR "e-health" OR "eMental health" OR "eTherap*" OR "digital behavior change intervention" OR "Information and communications technology" OR "Behavioral intervention technology" OR "Digital intervention" OR "Digital health intervention")
        AND
TS=(Depress* OR "affective disorder" OR "mood disorder" OR "affective symptoms")

A.3. PubMed

("Cell Phone"[Mesh] OR "Cell Phone"[Text Word] OR "Cellular Phone"[Text Word] OR "Cell Phone Use"[Mesh] OR "Computers, Handheld"[Mesh] OR "Smartphone"[Text Word] OR "Smart Phone"[Text Word] OR "iPhone"[Text Word] OR "Mobile App*"[Text Word] OR "Phone App*"[Text Word] OR "Telemedicine"[Mesh] OR "Information Technology"[Mesh] OR "mHealth"[Text Word] OR "mobile health"[Text Word] OR "eHealth"[Text Word] OR "e-Health"[Text Word] OR "eMental health"[Text Word] OR "eTherap*"[Text Word] OR "digital behavior change intervention"[Text Word] OR "Behavioral intervention technology"[Text Word] OR "Digital intervention"[Text Word] OR "Digital health intervention"[Text Word])

AND

("Depression"[Mesh] OR "Depress*"[Text Word] OR "Depressive Disorder"[Mesh] OR "Affective Symptoms"[Mesh] OR "affective symptoms"[Text Word])

References

Arean, P.A., Hallgren, K.A., Jordan, J.T., Gazzaley, A., Atkins, D.C., Heagerty, P.J., Anguera, J.A., 2016. The use and effectiveness of mobile apps for depression: results from a fully remote clinical trial. J. Med. Internet Res. 18 (12), e330 https://doi.org/10.2196/jmir.6482.

Beck, A.T., 2008. The evolution of the cognitive model of depression and its neurobiological correlates. Am. J. Psychiatr. 165 (8), 969–977. https://doi.org/10.1176/appi.ajp.2008.08050721.

Brooke, J., 1996. SUS: a "quick and dirty" usability scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, I.L. (Eds.), Usability Evaluation in Industry. Taylor and Francis, London, pp. 189–194.

Burns, M.N., Begale, M., Duffecy, J., Gergle, D., Karr, C.J., Giangrande, E., Mohr, D.C., 2011. Harnessing context sensing to develop a mobile intervention for depression. J. Med. Internet Res. 13 (3), e55 https://doi.org/10.2196/jmir.1838.

Caplan, S., Sosa Lovera, A., Reyna Liberato, P., 2018. A feasibility study of a mental health mobile app in the Dominican Republic: the untold story. Int. J. Ment. Health 47 (4), 311–345. https://doi.org/10.1080/00207411.2018.1553486.

Cormack, F., McCue, M., Taptiklis, N., Skirrow, C., Glazer, E., Panagopoulos, E., van Schaik, T.A., Fehnert, B., King, J., Barnett, J.H., 2019. Wearable technology for high-frequency cognitive and mood assessment in major depressive disorder: longitudinal observational study. JMIR Mental Health 6 (11), e12814. https://doi.org/10.2196/12814.

Cuijpers, P., van Straten, A., Warmerdam, L., 2007. Behavioral activation treatments of depression: a meta-analysis. Clin. Psychol. Rev. 27 (3), 318–326. https://doi.org/10.1016/j.cpr.2006.11.001.

Dahne, J., Kustanowitz, J., Lejuez, C.W., Dahne, J., Kustanowitz, J., Lejuez, C.W., 2018. Development and preliminary feasibility study of a brief behavioral activation mobile application (behavioral apptivation) to be used in conjunction with ongoing therapy. Cogn. Behav. Pract. 25 (1), 44–56. https://doi.org/10.1016/j.cbpra.2017.05.004.

Dahne, J., Collado, A., Lejuez, C.W., Risco, C.M., Diaz, V.A., Coles, L., Kustanowitz, J., Zvolensky, M.J., Carpenter, M.J., 2019a. Pilot randomized controlled trial of a Spanish-language behavioral activation mobile app (¡Aptívate!) for the treatment of depressive symptoms among United States Latinx adults with limited English proficiency. J. Affect. Disord. 250 (March), 210–217. https://doi.org/10.1016/j.jad.2019.03.009.

Dahne, J., Lejuez, C.W., Diaz, V.A., Player, M.S., Kustanowitz, J., Felton, J.W., Carpenter, M.J., 2019b. Pilot randomized trial of a self-help behavioral activation mobile app for utilization in primary care. Behav. Ther. 50 (4), 817–827. https://doi.org/10.1016/j.beth.2018.12.003.

Devilly, G.J., Borkovec, T.D., 2000. Psychometric properties of the credibility/expectancy questionnaire. J. Behav. Ther. Exp. Psychiatry 31 (2), 73–86. https://doi.org/10.1016/S0005-7916(00)00012-4.

Donker, T., Petrie, K., Proudfoot, J., Clarke, J., Birch, M.R., Christensen, H., 2013. Smartphones for smarter delivery of mental health programs: a systematic review. J. Med. Internet Res. 15 (11), e247 https://doi.org/10.2196/jmir.2791.

Dubad, M., Winsper, C., Meyer, C., Livanou, M., Marwaha, S., 2018. A systematic review of the psychometric properties, usability and clinical impacts of mobile mood-monitoring applications in young people. Psychol. Med. 48 (2), 208–228. https://doi.org/10.1017/S0033291717001659.

Economides, M., Ranta, K., Nazander, A., Hilgert, O., Goldin, P.R., Raevuori, A., Forman-Hoffman, V., 2019. Long-term outcomes of a therapist-supported, smartphone-based intervention for elevated symptoms of depression and anxiety: quasiexperimental, pre-postintervention study. JMIR MHealth and UHealth 7 (8), e14284. https://doi.org/10.2196/14284.

Firth, J., Torous, J., Nicholas, J., Carney, R., Pratap, A., Rosenbaum, S., Sarris, J., 2017. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 16 (3), 287–298. https://doi.org/10.1002/wps.20472.

Fuller-Tyszkiewicz, M., Richardson, B., Klein, B., Skouteris, H., Christensen, H., Austin, D., Castle, D., Mihalopoulos, C., O'Donnell, R., Arulkadacham, L., Shatte, A., Ware, A., 2018. A mobile app-based intervention for depression: end-user and expert usability testing study. JMIR Mental Health 5 (3), e54. https://doi.org/10.2196/mental.9445.

Furukawa, T.A., Horikoshi, M., Fujita, H., Tsujino, N., Jinnin, R., Kako, Y., Ogawa, S., Sato, H., Kitagawa, N., Shinagawa, Y., Ikeda, Y., Imai, H., Tajika, A., Ogawa, Y., Akechi, T., Yamada, M., Shimodera, S., Watanabe, N., Inagaki, M., Hasegawa, A., 2018a. Cognitive and behavioral skills exercises completed by patients with major depression during smartphone cognitive behavioral therapy: secondary analysis of a randomized controlled trial. JMIR Mental Health 5 (1), e4. https://doi.org/10.2196/mental.9092.

Furukawa, T.A., Imai, H., Horikoshi, M., Shimodera, S., Hiroe, T., Funayama, T., Akechi, T., 2018b. Behavioral activation: is it the expectation or achievement of mastery or pleasure that contributes to improvement in depression? J. Affect. Disord. 238, 336–341. https://doi.org/10.1016/j.jad.2018.05.067.

Graham, A.K., Lattie, E.G., Mohr, D.C., 2019. Experimental therapeutics for digital mental health. JAMA Psychiatry 76 (12), 1223–1224. https://doi.org/10.1001/jamapsychiatry.2019.2075.

Graham, A.K., Kwasny, M.J., Lattie, E.G., Greene, C.J., Gupta, N.V., Reddy, M., Mohr, D. C., 2021. Targeting subjective engagement in experimental therapeutics for digital mental health interventions. Internet Interv. 25, 100403. https://doi.org/10.1016/j.invent.2021.100403.

Hantsoo, L., Criniti, S., Khan, A., Moseley, M., Kincler, N., Faherty, L.J., Epperson, C.M., Bennett, I.M., 2018. A mobile application for monitoring and management of depressed mood in a vulnerable pregnant population. Psychiatr. Serv. 69 (1), 104–107. https://doi.org/10.1176/appi.ps.201600582.

Hawton, K., i Comabella, C.C., Haw, C., Saunders, K., 2013. Risk factors for suicide in individuals with depression: a systematic review. J. Affect. Disord. 147 (1–3), 17–28. https://doi.org/10.1016/j.jad.2013.01.004.

van Heerden, A., Tomlinson, M., Swartz, L., 2012. Point of care in your pocket: a research agenda for the field of m-health. Bulletin of the World Health Organisation 90, 393–394.

Hung, S., Li, M.S., Chen, Y.L., Chiang, J.H., Chen, Y.Y., Hung, G.C.L., 2016. Smartphone-based ecological momentary assessment for Chinese patients with depression: an exploratory study in Taiwan. Asian J. Psychiatr. 23, 131–136. https://doi.org/10.1016/j.ajp.2016.08.003.

Hur, J.W., Kim, B., Park, D., Choi, S.W., 2018. A scenario-based cognitive behavioral therapy mobile app to reduce dysfunctional beliefs in individuals with depression: a randomized controlled trial. Telemedicine and E-Health 24 (9), 710–716. https://doi.org/10.1089/tmj.2017.0214.

Inkster, B., Sarda, S., Subramanian, V., 2018. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. JMIR MHealth and UHealth 6 (11), e12106. https://doi.org/10.2196/12106.

James, S.L., Abate, D., Abate, K.H., Abay, S.M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A., Abdollahpour, I., Abdulkader, R.S., Abebe, Z., Abera, S.F., Abil, O.Z., Abraha, H.N., Abu-Raddad, L.J., Abu-Rmeile, N.M. E., Accrombessi, M.M.K., Murray, C.J.L., 2018. Global, regional, and national incidence, prevalence, and years lived with disability for 354 Diseases and Injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 392 (10159), 1789–1858. https://doi.org/10.1016/S0140-6736(18)32279-7.

Kelders, S.M., Kok, R.N., Ossebaard, H.C., Van Gemert-Pijnen, J.E.W.C., 2012. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J. Med. Internet Res. 14 (6), e152 https://doi.org/10.2196/jmir.2104.

Lattie, E.G., Schueller, S.M., Sargent, E., Stiles-Shields, C., Tomasino, K.N., Corden, M.E., Begale, M., Karr, C.J., Mohr, D.C., 2016. Uptake and usage of IntelliCare: a publicly available suite of mental health and well-being apps. Internet Interv. 4, 152–158. https://doi.org/10.1016/j.invent.2016.06.003.

Li, Y., Guo, Y., Hong, Y.A., Zhu, M., Zeng, C., Qiao, J., Xu, Z., Zhang, H., Zeng, Y., Cai, W., Li, L., Liu, C., 2019. Mechanisms and effects of a wechat-based intervention on suicide among people living with HIV and depression: path model analysis of a randomized controlled trial. J. Med. Internet Res. 21 (11), e14729 https://doi.org/10.2196/14729.

Linardon, J., Fuller-Tyszkiewicz, M., 2020. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. J. Consult. Clin. Psychol. 88 (1), 1–13. https://doi.org/10.1037/ccp0000459.

Liu, J.X., Goryakin, Y., Maeda, A., Bruckner, T., Scheffler, R., 2017. Global health workforce labor market projections for 2030. Hum. Resour. Health 15, 11. https://doi.org/10.1186/s12960-017-0187-2.

Ly, K.H., Trüschel, A., Jarl, L., Magnusson, S., Windahl, T., Johansson, R., Carlbring, P., Andersson, G., 2014. Behavioural activation versus mindfulness-based guided self-help treatment administered through a smartphone application: a randomised controlled trial. BMJ Open 4, e003440. https://doi.org/10.1136/bmjopen-2013-003440.

Ly, K.H., Topooco, N., Cederlund, H., Wallin, A., Bergstrom, J., Molander, O., Carlbring, P., Andersson, G., 2015. Smartphone-supported versus full behavioural activation for depression: a randomised controlled trial. PLoS One 10 (5), e0126559. https://doi.org/10.1371/journal.pone.0126559.

Mantani, A., Kato, T., Furukawa, T.A., Horikoshi, M., Imai, H., Hiroe, T., Chino, B., Funayama, T., Yonemoto, N., Zhou, Q., Kawanishi, N., 2017. Smartphone cognitive behavioral therapy as an adjunct to pharmacotherapy for refractory depression: randomized controlled trial. J. Med. Internet Res. 19 (11), e373 https://doi.org/10.2196/jmir.8602.

McKnight, P.E., Kashdan, T.B., 2009. The importance of functional impairment to mental health outcomes: a case for reassessing our goals in depression treatment research. Clin. Psychol. Rev. 29 (3), 243–259. https://doi.org/10.1016/j.cpr.2009.01.005.

Menezes, P., Quayle, J., Claro, H.G., Da Silva, S., Brandt, L.R., Diez-Canseco, F., Miranda, J.J., Price, L.N., Mohr, D.C., Araya, R., 2019. Use of a mobile phone app to treat depression comorbid with hypertension or diabetes: a pilot study in Brazil and Peru. JMIR Mental Health 6 (4), e11698. https://doi.org/10.2196/11698.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann. Intern. Med. 151 (4), 264–269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135.

Molloy, A., Ellis, D.M., Su, L., Anderson, P.L., 2021. Improving acceptability and uptake behavior for internet-based cognitive-behavioral therapy. Frontiers in Digital Health 3, 653686. https://doi.org/10.3389/fdgth.2021.653686.

Moukaddam, N., Truong, A., Cao, J., Shah, A., Sabharwal, A., 2019. Findings from a trial of the smartphone and online usage-based evaluation for depression (SOLVD) application: what do apps really tell us about patients with depression? Concordance

between app-generated data and standard psychiatric questionnaires for depression and anxiety. J. Psychiatr. Pract. 25 (5), 365–373. https://doi.org/10.1097/PRA.0000000000000420.

Ng, M.M., Firth, J., Minen, M., Torous, J., 2019. User engagement in mental health apps: a review of measurement, reporting, and validity. Psychiatr. Serv. 70 (7), 538–544. https://doi.org/10.1176/appi.ps.201800519.

Nutt, D.J., Wilson, S., Paterson, L., 2008. Sleep disorders as core symptoms of depression. Dialogues in Clinical Neuroscience 10 (3), 329–336. https://doi.org/10.31887/dcns.2008.10.3/dnutt.

O'Brien, H.L., Toms, E.G., 2010. The development and evaluation of a survey to measure user engagement. J. Am. Soc. Inf. Sci. Technol. 61 (1), 50–69. https://doi.org/10.1002/asi.

Perski, O., Blandford, A., West, R., Michie, S., 2017. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl. Behav. Med. 7 (2), 254–267. https://doi.org/10.1007/s13142-016-0453-1.

Pew Research Center, 2019. Mobile phone ownership over time. Retrieved from. http://www.pewinternet.org/fact-sheet/mobile/.

Pham, Q., Graham, G., Carrion, C., Morita, P.P., Seto, E., Stinson, J.N., Cafazzo, J.A., 2019. A library of analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions: scoping review. JMIR MHealth and UHealth 7 (1), e11941. https://doi.org/10.2196/11941.

Pratap, A., Renn, B.N., Volponi, J., Mooney, S.D., Gazzaley, A., Arean, P.A., Anguera, J. A., 2018. Using mobile apps to assess and treat depression in hispanic and latino populations: fully remote randomized clinical trial. J. Med. Internet Res. 20 (8), e10130 https://doi.org/10.2196/10130.

Rock, P.L., Roiser, J.P., Riedel, W.J., Blackwell, A.D., 2014. Cognitive impairment in depression: a systematic review and meta-analysis. Psychol. Med. 44 (10), 2029–2040. https://doi.org/10.1017/S0033291713002535.

Sawyer, A., Kaim, A., Le, H.N., McDonald, D., Mittinty, M., Lynch, J., Sawyer, M., 2019. The effectiveness of an app-based nurse-moderated program for new mothers with depression and parenting problems (EMUMS Plus): pragmatic randomized controlled trial. J. Med. Internet Res. 21 (6), e13689 https://doi.org/10.2196/13689.

Schlosser, D.A., Campellone, T.R., Truong, B., Anguera, J.A., Vergani, S., Vinogradov, S., Arean, P., 2017. The feasibility, acceptability, and outcomes of PRIME-D: a novel mobile intervention treatment for depression. Depression and Anxiety 34 (6), 546–554. https://doi.org/10.1002/da.22624.

Schuster, R., Kalthoff, I., Walther, A., Köhldorfer, L., Partinger, E., Berger, T., Laireiter, A. R., 2019. Effects, adherence, and therapists' perceptions of web- and mobile-supported group therapy for depression: mixed-methods study. J. Med. Internet Res. 21 (5), e11860 https://doi.org/10.2196/11860.

Shiffman, S., Stone, A.A., Hufford, M.R., 2008. Ecological momentary assessment. Annu. Rev. Clin. Psychol. 4 (1), 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415.

Sigg, S., Lagerspetz, E., Peltonen, E., Nurmi, P., Tarkoma, S., 2016. Sovereignty of the apps: there's more to relevance than downloads. Retrieved from: https://arxiv.org/pdf/1611.10161.pdf.

Statista, 2019. Number of smartphone users worldwide 2016 to 2021 (in billions). Retrieved from: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/.

Stiles-Shields, C., Montague, E., Kwasny, M.J., Mohr, D.C., 2019. Behavioral and cognitive intervention strategies delivered via coached apps for depression: pilot trial. Psychol. Serv. 16 (2), 233–238. https://doi.org/10.1037/ser0000261.

Takahashi, K., Takada, K., Hirao, K., 2019. Feasibility and preliminary efficacy of a smartphone application intervention for subthreshold depression. Early Intervention in Psychiatry 13 (1), 133–136. https://doi.org/10.1111/eip.12540.

Torous, J., Staples, P., Shanahan, M., Lin, C., Peck, P., Keshavan, M., Onnela, J.-P., 2015. Utilizing a personal smartphone custom app to assess the patient health questionnaire-9 (PHQ-9) depressive symptoms in patients with major depressive disorder. JMIR Mental Health 2 (1), e8. https://doi.org/10.2196/mental.3889.

Vogel, D.L., Heimerdinger-Edwards, S.R., Hammer, J.H., Hubbard, A., 2011. "Boys don't cry": examination of the links between endorsement of masculine norms, self-stigma, and help-seeking attitudes for men from diverse backgrounds. J. Couns. Psychol. 58 (3), 368–382. https://doi.org/10.1037/a0023688.

Watts, S., Mackenzie, A., Thomas, C., Griskaitis, A., Mewton, L., Williams, A., Andrews, G., 2013. CBT for depression: a pilot RCT comparing mobile phone vs. computer. BMC Psychiatry 13, 49. https://doi.org/10.1186/1471-244X-13-49.

Weisel, K.K., Fuhrmann, L.M., Berking, M., Baumeister, H., Cuijpers, P., Ebert, D.D., 2019. Standalone smartphone apps for mental health—a systematic review and meta-analysis. Npj Digital Medicine 2, 118. https://doi.org/10.1038/s41746-019-0188-8.

World Health Organization, 2017. Depression and other common mental disorders: global health estimates. Retrieved from: http://apps.who.int/iris/bitstream/handle/10665/254610/WHOMSD?sequence=1.

Yardley, L., Spring, B.J., Riper, H., Morrison, L.G., Crane, D.H., Curtis, K., Merchant, G. C., Naughton, F., Blandford, A., 2016. Understanding and promoting effective engagement with digital behavior change interventions. Am. J. Prev. Med. 51 (5), 833–842. https://doi.org/10.1016/j.amepre.2016.06.015.

Zhu, M., Cai, W., Li, L., Guo, Y., Monroe-Wise, A., Li, Y., Zeng, C., Qiao, J., Xu, Z., Zhang, H., Zeng, Y., Liu, C., 2019. Mediators of intervention effects on depressive symptoms among people living with HIV: secondary analysis of a mobile health randomized controlled trial using latent growth curve modeling. JMIR MHealth and UHealth 7 (11), e15489. https://doi.org/10.2196/15489.