

Original Paper

# Clinically Meaningful Use of Mental Health Apps and its Effects on Depression: Mixed Methods Study

Renwen Zhang<sup>1</sup>, MPhil; Jennifer Nicholas<sup>2</sup>, PhD; Ashley A Knapp<sup>2</sup>, PhD; Andrea K Graham<sup>2,3</sup>, PhD; Elizabeth Gray<sup>2</sup>, MSc; Mary J Kwasny<sup>2</sup>, ScD; Madhu Reddy<sup>1,2</sup>, PhD; David C Mohr<sup>2</sup>, PhD

<sup>1</sup>Department of Communication Studies, Northwestern University, Evanston, IL, United States

<sup>2</sup>Center for Behavioral Intervention Technologies, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

<sup>3</sup>Department of Medical Social Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

**Corresponding Author:**

Renwen Zhang, MPhil  
Department of Communication Studies  
Northwestern University  
2240 Campus Dr  
MTS Program  
Evanston, IL, 60208  
United States  
Phone: 1 8479976596  
Email: [alicezhang@u.northwestern.edu](mailto:alicezhang@u.northwestern.edu)

## Abstract

**Background:** User engagement is key to the effectiveness of digital mental health interventions. Considerable research has examined the clinical outcomes of overall engagement with mental health apps (eg, frequency and duration of app use). However, few studies have examined how specific app use behaviors can drive change in outcomes. Understanding the clinical outcomes of more nuanced app use could inform the design of mental health apps that are more clinically effective to users.

**Objective:** This study aimed to classify user behaviors in a suite of mental health apps and examine how different types of app use are related to depression and anxiety outcomes. We also compare the clinical outcomes of specific types of app use with those of generic app use (ie, intensity and duration of app use) to understand what aspects of app use may drive symptom improvement.

**Methods:** We conducted a secondary analysis of system use data from an 8-week randomized trial of a suite of 13 mental health apps. We categorized app use behaviors through a mixed methods analysis combining qualitative content analysis and principal component analysis. Regression analyses were used to assess the association between app use and levels of depression and anxiety at the end of treatment.

**Results:** A total of 3 distinct clusters of app use behaviors were identified: learning, goal setting, and self-tracking. Each specific behavior had varied effects on outcomes. Participants who engaged in self-tracking experienced reduced depression symptoms, and those who engaged with learning and goal setting at a moderate level (ie, not too much or not too little) also had an improvement in depression. Notably, the combination of these 3 types of behaviors, what we termed “clinically meaningful use,” accounted for roughly the same amount of variance as explained by the overall intensity of app use (ie, total number of app use sessions). This suggests that our categorization of app use behaviors succeeded in capturing app use associated with better outcomes. However, anxiety outcomes were neither associated with specific behaviors nor generic app use.

**Conclusions:** This study presents the first granular examination of user interactions with mental health apps and their effects on mental health outcomes. It has important implications for the design of mobile health interventions that aim to achieve greater user engagement and improved clinical efficacy.

(*J Med Internet Res* 2019;21(12):e15644) doi: [10.2196/15644](https://doi.org/10.2196/15644)

**KEYWORDS**

mHealth; mobile apps; mental health; engagement

## Introduction

Over the past decade, mobile phone apps have become portals for managing health. These digital tools help users monitor physical activity, plan healthy meals, and keep track of daily moods and other personal data. Given the accessibility and ubiquity of mobile phones, researchers and clinicians have increasingly leveraged mobile phone apps to deliver health interventions and enhance self-management of chronic conditions such as depression and anxiety [1-4]. Mental health apps can enhance skill building, deliver psychoeducation, and facilitate self-monitoring, thereby reducing symptoms of depression and anxiety [2,5,6]. These mobile technologies incorporate a wide array of system features and strategies that foster user engagement and promote behavior change, such as customization, reminders, self-monitoring, rewards, and peer support [3,7-10].

For mental health apps to be effective and successful, user engagement is critical. However, little consensus exists on how to define and measure engagement [11-13]. Engagement has been inconsistently viewed as a multidimensional construct, variably encompassing behavioral, emotional, and cognitive factors [14,15]. Consequently, engagement measures vary enormously, ranging from self-report questionnaires to system usage data or sensor data [16,17]. In this study, we define user engagement as a behavioral experience that involves people's physical interaction with mobile phone apps [18]. System usage data are the most commonly used behavioral measures of engagement in mobile health (mHealth) interventions, as they can be acquired in the course of app use and require no additional effort from the user [14]. These data quantitatively capture how users physically interact with the app in terms of intensity (eg, number of app use sessions [19]), frequency (eg, number/percentage of days the app is used [20]), duration (eg, time spent on the app [21,22]), time between first and last app use [19,20]), and types (eg, passive, active, reflective, and didactic [23,24]). These measures of behavioral engagement have been found to be related to health outcomes such as psychological well-being [20,22].

However, behavioral engagement metrics have typically employed broad use metrics that measure the *quantity* of engagement. Few studies have examined more granular user interactions with specific components of mHealth interventions, which might provide insight into *how* a person is using an app in ways that are clinically meaningful [25]. For example, a user might simply open an app in response to a prompt without any deeper engagement, whereas others might engage in more meaningful activities such as inputting or reflecting on data and reading content. These different types of activities might reflect differing levels and types of engagement, resulting in various health outcomes. Identifying the types of user behaviors could provide insight into what aspects of mHealth interventions are engaging to users and what aspects may drive behavior change and symptom improvement [15,26]. This could also inform opportunities for designing more engaging and clinically effective mHealth interventions.

This study aimed to provide a categorization of the types of user behaviors in a suite of mental health apps for depression and anxiety. We then examine how the different types of app use are related to improvements in symptoms of depression and anxiety. To provide a holistic picture of app use, we also differentiate the more nuanced app use from generic app use (ie, intensity and duration of app use) and examine how these different use metrics influence outcomes. As such, this study presents the first granular classification of user interactions with mental health apps and their impact on outcomes.

## Methods

### Participants and Procedure

This study represents a secondary analysis of data from a randomized trial examining the efficacy of coaching and app recommendations to increase engagement with IntelliCare, a suite of mental health apps (Clinicaltrials.gov NCT02801877). Full study details have been described elsewhere [27]. In brief, participants were recruited between July 2016 and May 2017 via social and print media advertising, research registries, and commercial recruitment firms. People interested in participating completed an initial Web-based questionnaire deployed through a secure Web-based data capture system (Research Electronic Data Capture; [28]). The inclusion criteria were as follows: (1) aged 18 years or older, (2) reported elevated symptoms of depression (Patient Health Questionnaire-9 [PHQ-9]  $\geq 10$ ) or anxiety (generalized anxiety disorder-7 [GAD-7]  $\geq 8$ ), (3) resided in the United States, (4) could speak and read English, and (5) had an Android phone with data and text plans. Following baseline assessment, 301 eligible participants were randomized to 1 of the 4 treatments within a 2x2 factorial design for 8 weeks. Brief descriptions of the mobile intervention and each condition are provided below. The trial was approved by the Northwestern University Institutional Review Board before participant contact and monitored by an independent data safety monitoring board throughout the study period. All participants provided informed consent.

### Treatments

#### IntelliCare

The IntelliCare platform consists of 12 clinical apps, each targeting a specific behavioral or psychological treatment strategy (eg, cognitive restructuring, behavioral activation, social support, and relaxation) to improve symptoms of depression and anxiety. The specific apps have been described in more detail elsewhere [2,19]. In addition, a Hub app consolidates notifications from the downloaded clinical apps and is able to recommend other apps in the suite. IntelliCare apps prioritize interactive skills training over psychoeducation and are designed for frequent, short interactions.

#### Coaching

Participants assigned to the coach condition received 8 weeks of coaching aimed to support engagement. Coaching was based on a low-intensity coaching model [29] that relied primarily on brief SMS text messaging (2-3 messages per week) to promote engagement. Participants assigned to the self-guided condition had no sustained contact with coaches. Participants in both

conditions received an orientation call at the beginning of the trial to ensure they had appropriately installed the Hub app and understood how to use the IntelliCare platform.

**Recommendations**

Participants randomized to the recommendation condition received recommendations for new apps weekly through the Hub app. The recommendation system leveraged app use data from approximately 80,000 users who had downloaded the IntelliCare apps to identify apps that the individual was more likely to use based on their app use profile. Participants not assigned to the recommendation condition did not receive recommendations and were encouraged to explore the apps by themselves.

**Measures**

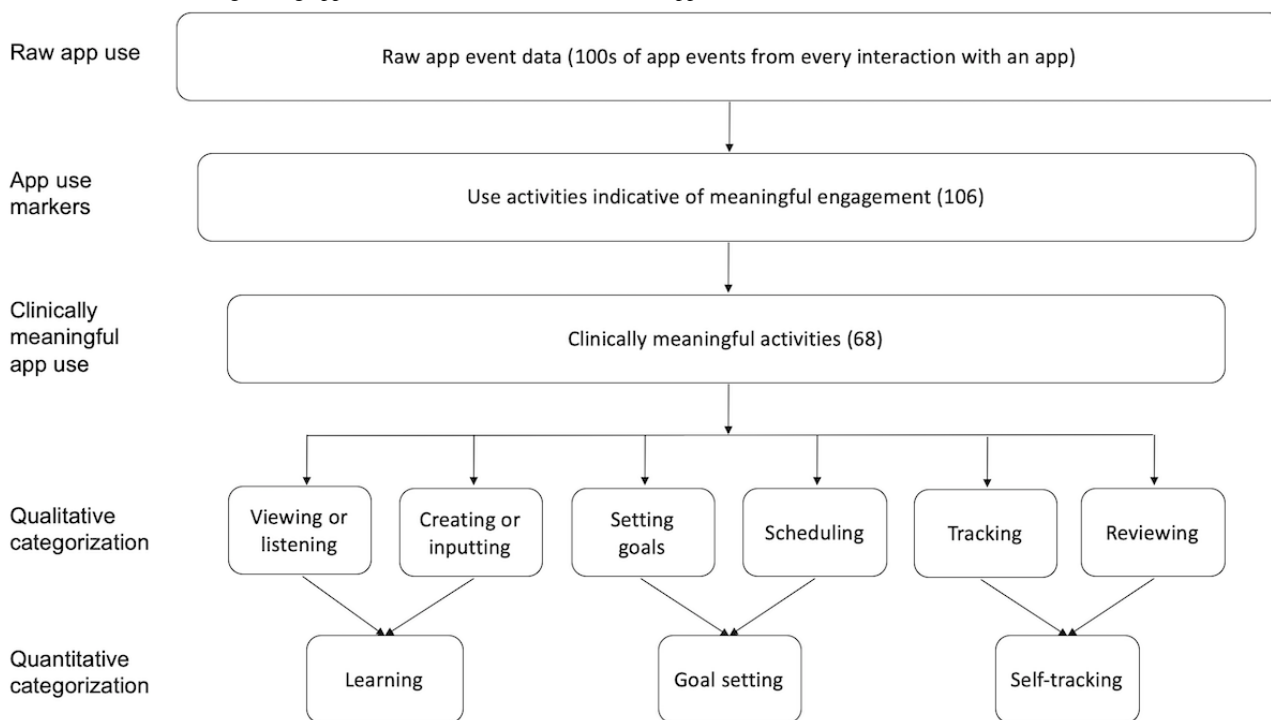
**App Use Metrics**

Usage logs for each app were recorded locally on the user’s mobile phone, which were then obtained and analyzed to extract app use metrics. In this study, we categorized 2 types of app use: clinically meaningful app use and generic app use.

Clinically meaningful app use metrics were developed in a 2-step process. First, a group of 5 psychologists who had been involved in the design of the apps created granular app use

markers from raw event data. These app use markers were each defined by a small number of app event data considered to be indicative of meaningful engagement. Examples of such markers are app event data that defined completion of a skill-building exercise, reading psychoeducational text, or logging daily activities. Second, 2 authors (AZ and JN) developed a coding scheme to categorize these app use markers, drawing on existing literature and qualitative interviews with 32 IntelliCare users. Specifically, these 2 authors first collectively coded 18.8% (20/106) of the app use markers to develop the coding scheme and then used this coding scheme to separately categorize the remaining markers. Any disagreements in categorization were sent to the third author (AK), who served as a tiebreaker, and were resolved through rigorous discussion. During this process, we removed a number of app use markers that rarely occurred (eg, texting a friend from within the app) or were not considered clinically meaningful (eg, launching the app and viewing app use tips). Through this process, we identified 67 aggregated activities considered to be clinically meaningful, which we labeled “clinically meaningful use activities.” These activities were further grouped into 6 types (see [Multimedia Appendix 1](#)). The categorization procedure is presented in the bottom 4 steps of [Figure 1](#). The 6 types of clinically meaningful activities identified were as follows:

**Figure 1.** Procedure for categorizing app use activities across 13 IntelliCare apps.



1. Viewing/listening: reading/watching/listening to content from the app (eg, playing an exercise video, viewing a coping card, and listening to a relaxing audio)
2. Creating/inputting: creating and editing content for the purpose of learning and cultivating a skill (eg, identifying a coping activity and creating a positive or self-affirming statement)
3. Setting goals: selecting, editing, or adding self-identified or assigned goals (eg, adding or deleting a checklist item and selecting a weekly goal)
4. Scheduling: scheduling activities or changing reminders to fit one’s schedule (eg, scheduling an upcoming exercise and changing the reminder time)
5. Tracking: keeping track of one’s own performance or status through checking off, rating, or logging personal activities and moods, including facts and reasons (eg, checking a

completed activity, rating a level of stress, and creating a sleep log)

6. Reviewing: reviewing one's own content and progress (eg, reviewing past activities and lessons).

*Generic app use* was measured by 2 metrics: intensity of use and duration of use [22]. Intensity of use was defined as a user's total number of app use sessions over the 8-week treatment period. An app use session was specified as a sequence of user-initiated actions or events separated by less than 5 min. Duration of use was defined as the total time an individual spent using the apps over the treatment period. It was calculated by summing the mean duration (in hours) of daily app sessions across all days in treatment.

### Outcome Assessment

The primary outcomes of the study were depression and anxiety symptom severity, measured with the PHQ-9 [30] and GAD-7 [31] at baseline and end of treatment (week 8). Higher scores reflect higher levels of depression or anxiety.

### Data Analysis

Principal component analysis was performed on the 6 identified types of clinically meaningful activities, standardizing by type, to explore any underlying patterns of these activity types. Medians and IQRs of app use metrics were reported. Then, the relationship between app use metrics and treatment outcomes was analyzed using linear regression analyses, adjusting for baseline PHQ-9 or GAD-7 and randomization strata. We first plotted the bivariate relations between all use metrics and end-of-treatment outcomes, which revealed nonlinear patterns. In response, we categorized each app use metric into 4 quartiles. We considered the first quartile minimal intensity of use, the second quartile low intensity of use, the third quartile moderate intensity of use, and the fourth quartile high intensity of use. Regression models were fit to examine the relationship between the quartiles of app use metrics and outcomes, using the lowest quartile as the reference group. Regression coefficients (beta) with their 95% CIs and significance levels were reported for

both unadjusted and adjusted models. In addition, the  $R^2$  values were reported for the unadjusted models to assess the magnitude of the effect. All analyses were performed using R version 3.5.1.

### Ethical Standards

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

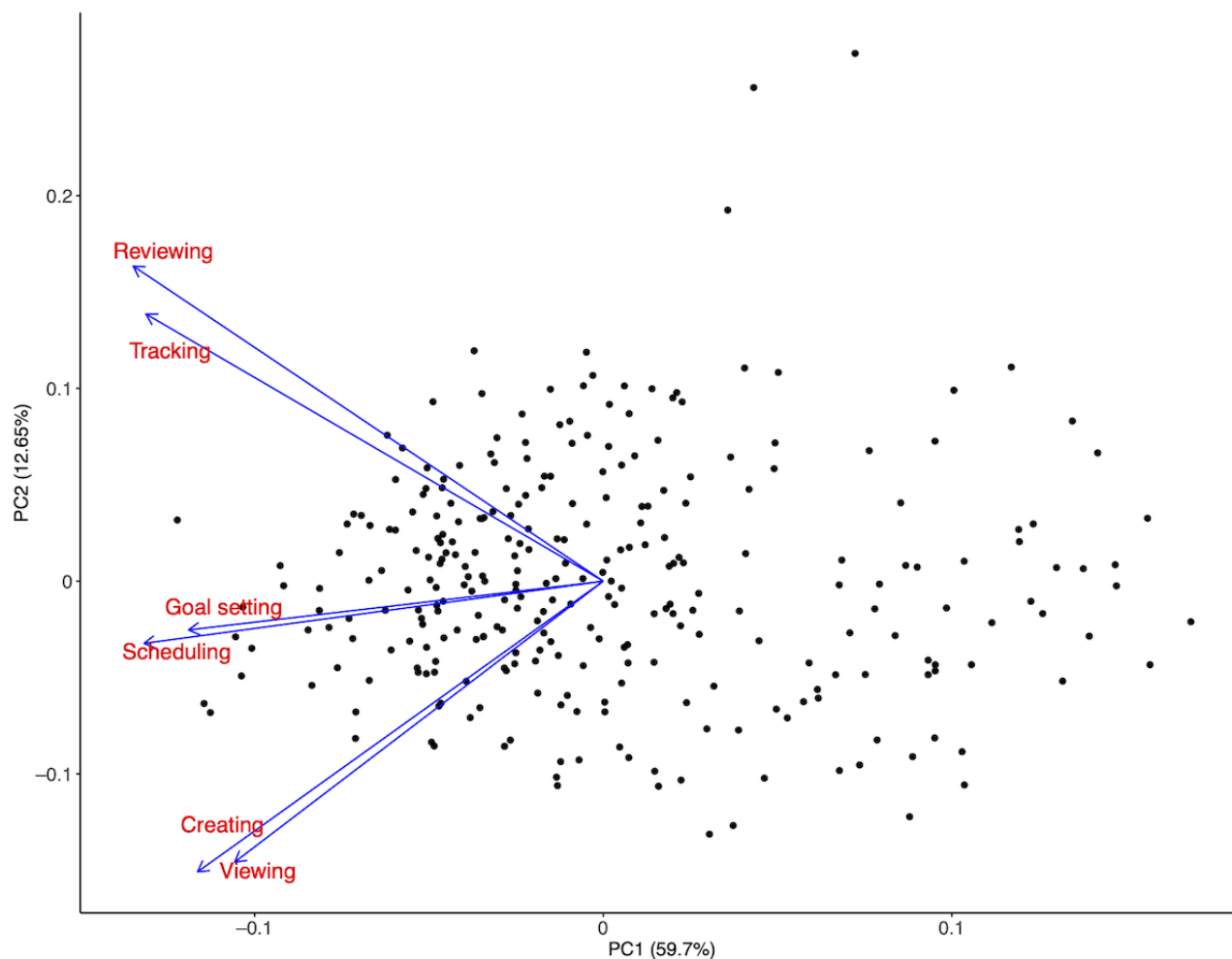
## Results

### Participants

A total of 301 eligible participants were enrolled in the randomized trial. The majority of participants were female (228/301, 75.7%), and the mean age was 37 (SD 11.84) years, ranging from 18 to 69 years. Most (237/301, 78.7%) of the participants identified themselves as white, 29 (9.6%) as African American, 10 (3.3%) as Asian, and 25 (8.3%) as "other." The mean baseline level of depression (PHQ-9) was 13.21 (SD 4.63), and the mean baseline level of anxiety (GAD-7) was 11.98 (SD 4.02). A total of 10 participants discontinued treatment and were lost to follow-up. Further details of the sample and participant flow through the study are reported in the study by Mohr et al [27].

### Clinically Meaningful Use

Correlation analysis showed that the 6 identified types of clinically meaningful activities were highly correlated; accordingly, we conducted a principal component analysis to further group these activity types. The analysis identified 3 clusters of meaningful activities that could best be described as: (1) "learning," encompassing "viewing" and "creating;" (2) "goal setting," including "setting goals" and "scheduling;" and (3) "self-tracking," consisting of "reviewing" and "tracking." The first 2 principal components explained 72.4% of the variability in the data (see Figure 2). Figure 1 presents a visual illustration of the categorization of clinically meaningful activities across the apps.

**Figure 2.** Principal component analysis of the types of clinically meaningful activities.

Self-tracking was performed most often, with a median frequency of 152 activities (IQR 61-300). Learning was performed less often, with a median frequency of 110 activities (IQR 52-191). Goal setting was used the least, with a median frequency of 59 activities (IQR 15-141). We also examined the frequency of overall clinically meaningful app use by combining all the 67 identified clinically meaningful use activities. The median frequency of clinically meaningful app use was 400 (IQR 200-608).

### Generic App Use

Over the 8-week treatment period, the median number of app use sessions was 184 (IQR 116-306), and the median duration of app use over the 8-week treatment period was 3.0 hours (IQR 1.7-5.0).

### App Use and Depression Outcome

#### *Clinically Meaningful Use*

We first examined how each of the 3 clusters of clinically meaningful activities predicted individuals' level of depression at the end of treatment, compared with the lowest quartile (minimal use) of each cluster (Table 1). End of treatment PHQ-9, after controlling for treatment condition and baseline PHQ-9, was significantly related to moderate intensity of learning (beta=-2.17;  $P=.002$ ); moderate intensity of goal setting (beta=-2.08;  $P=.007$ ); and low, moderate, and high intensity of self-tracking (beta=-2.46;  $P=.002$ ; beta=-1.94;  $P=.01$ ; and beta=-1.92;  $P=.009$ , respectively), compared with minimal intensity use. Moreover, learning and control variables accounted for 29.2% variance of the model (adjusted  $R^2=0.292$ ), goal setting and control variables accounted for 26.5% variance of the model (adjusted  $R^2=0.265$ ), and self-tracking and control variables accounted for 27.4% variance of the model (adjusted  $R^2=0.274$ ).

**Table 1.** Regression models of 3 clusters of clinically meaningful activities predicting depression outcome.

Covariate	Model 1 <sup>a</sup>		Model 2 <sup>b</sup>		Model 3 <sup>c</sup>	
	Estimate (SE)	P value	Estimate (SE)	P value	Estimate (SE)	P value
Intercept	1.86 (0.93)	.047	1.57 (0.93)	.09	2.16 (0.91)	.02
Coached	-0.78 (0.52)	.13	-0.42 (0.56)	.46	-0.12 (0.56)	.83
Full Hub	-0.21 (0.50)	.67	-0.28 (0.51)	.59	-0.05 (0.51)	.93
PHQ9_baseline	0.54 (0.05)	<.001	0.56 (0.06)	<.001	0.55 (0.06)	<.001
<b>Learning_minimal intensity<sup>d</sup></b>						
Learning_low intensity	0.64 (0.73)	.39	— <sup>e</sup>	—	—	—
Learning_moderate intensity	-2.17 (0.71)	.002	—	—	—	—
Learning_high intensity	-1.22 (0.73)	.09	—	—	—	—
<b>Goal setting_minimal intensity<sup>d</sup></b>						
Goal setting_low intensity	—	—	-0.62 (0.76)	.41	—	—
Goal setting_moderate intensity	—	—	-2.08 (0.76)	.007	—	—
Goal setting_high intensity	—	—	-0.76 (0.76)	.32	—	—
<b>Self-tracking_minimal intensity<sup>d</sup></b>						
Self-tracking_low intensity	—	—	—	—	-2.46 (0.78)	.002
Self-tracking_moderate intensity	—	—	—	—	-1.94 (0.76)	.01
Self-tracking_high intensity	—	—	—	—	-1.92 (0.73)	.009

<sup>a</sup> $R^2=0.307$ ; Adjusted  $R^2=0.292$ .

<sup>b</sup> $R^2=0.281$ ; Adjusted  $R^2=0.265$ .

<sup>c</sup> $R^2=0.289$ ; Adjusted  $R^2=0.274$ .

<sup>d</sup>Values of reference group.

<sup>e</sup>Not applicable.

In addition to examining the 3 identified clusters of clinically meaningful activities, we also explored how outcomes were related to overall clinically meaningful app use. As shown in [Table 2](#) (model 1), after controlling for treatment condition and baseline level of depression, PHQ-9 at the end of treatment was significantly and negatively associated with low intensity of meaningful use (beta=-2.00;  $P=.007$ ), moderate intensity of meaningful use (beta=-2.07;  $P=.006$ ), and high intensity of meaningful use (beta=-2.05;  $P=.006$ ), compared with minimal intensity of clinically meaningful app use. Clinically meaningful app use and the control variables accounted for 27.3% variance of the model (adjusted  $R^2=0.273$ ).

### Generic App Use

PHQ-9 at the end of treatment was significantly and negatively associated with low intensity of generic app use (beta=-1.44;  $P=.047$ ), moderate intensity of generic app use (beta=-2.38;  $P=.001$ ), and high intensity of generic app use (beta=-2.45;  $P=.001$ ), compared with minimal intensity of generic app use ([Table 2](#), model 2). In general, as the number of generic app use sessions increased, symptoms of depression at the end of treatment decreased. The intensity metrics and control variables accounted for 27.9% variance of the model (adjusted  $R^2=0.279$ ). Only moderate duration of use was significantly associated with PHQ-9 at week 8 (beta=-1.52;  $P=.045$ ) when compared with minimal duration of use ([Table 2](#), model 3). The duration metrics and control variables accounted for 25.8% variance of the model (adjusted  $R^2=0.258$ ).

**Table 2.** Regression models of total meaningful app use, generic app use, and duration of app use predicting depression outcome.

Covariate	Model 1 <sup>a</sup>		Model 2 <sup>b</sup>		Model 3 <sup>b</sup>	
	Estimate (SE)	<i>P</i> value	Estimate (SE)	<i>P</i> value	Estimate (SE)	<i>P</i> value
Intercept	2.25 (0.92)	.02	2.24 (0.92)	.02	1.71 (0.93)	.07
Coached	−0.34 (0.53)	.52	−0.58 (0.51)	.26	−0.57 (0.54)	.23
Full Hub	−0.11 (0.51)	.82	0.26 (0.54)	.63	0.02 (0.53)	.98
PHQ9_baseline	0.55 (0.06)	<.001	0.55 (0.05)	<.001	0.54 (0.06)	<.001
<b>Meaningful use_minimal intensity</b>						
Meaningful use_low intensity	−2.00 (0.74)	.007	— <sup>e</sup>	—	—	—
Meaningful use_moderate intensity	−2.07 (0.74)	.006	—	—	—	—
Meaningful use_high intensity	−2.05 (0.74)	.006	—	—	—	—
<b>Generic app use_minimal intensity<sup>d</sup></b>						
Generic app use_low intensity	—	—	−1.44 (0.72)	.047	—	—
Generic app use_moderate intensity	—	—	−2.38 (0.73)	.001	—	—
Generic app use_high intensity	—	—	−2.45 (0.76)	.001	—	—
<b>Generic app use_minimal duration<sup>d</sup></b>						
Generic app use_low duration	—	—	—	—	−0.32 (0.75)	.68
Generic app use_moderate duration	—	—	—	—	−1.52 (0.76)	.045
Generic app use_high duration	—	—	—	—	−1.24 (0.78)	.12

<sup>a</sup> $R^2=0.288$ ; Adjusted  $R^2=0.273$ .

<sup>b</sup> $R^2=0.295$ ; Adjusted  $R^2=0.279$ .

<sup>c</sup> $R^2=0.274$ ; Adjusted  $R^2=0.258$ .

<sup>d</sup>Values of reference group.

<sup>e</sup>Not applicable.

## App Use and Anxiety Outcome

Anxiety (GAD-7) at the end of the treatment was neither significantly associated with the 3 clusters of clinically meaningful activities (all  $P>.11$ ) nor the overall clinically meaningful app use (all  $P>.13$ ). Therefore, no further analyses were conducted regarding the association between anxiety and additional generic app use metrics.

## Discussion

### Principal Findings

This study provided a categorization of user behaviors in a suite of mental health apps and investigated how different types of app use were related to improvements in depression and anxiety symptoms following an 8-week intervention. The results showed that different types of clinically meaningful activities (ie, learning, goal setting, and self-tracking) had varied effects on outcomes. Self-tracking at varied levels of intensity was related to improvement in depression symptoms, whereas only moderate intensity of learning and goal setting predicted improvement in depression symptoms. Thus, this study provides insight into how different types of app use might be conducive to improved intervention outcomes.

Drawing on a mixed methods approach, we identified 6 types of clinically meaningful activities across multiple apps, which

were further grouped into 3 clusters—learning, goal setting, and self-tracking. This categorization was achieved through a combination of qualitative content analysis and quantitative statistical analysis. The results show that users engaged in self-tracking most frequently, followed by learning and goal setting. These 3 types of use activities have been well documented in mHealth and human-computer interaction (HCI) research as approaches to drive engagement and promote behavior change [8,32]. However, little is known about how these activities are related to health outcomes, as previous studies have primarily focused on the clinical outcomes of generic app use [21,33]. By shifting attention from generic app use to a more granular examination of meaningful app use, this study provides a more nuanced understanding of user engagement with mental health apps.

Notably, overall clinically meaningful app use (combination of all 67 identified clinically meaningful use activities) accounted for roughly the same amount of variance in depression severity as explained by the intensity of overall app use (ie., total number of app use sessions). Therefore, our identification of clinically meaningful app use was successful at capturing the activities associated with better mental health outcomes. This suggests that we accurately identified the clinically meaningful intervention components within this suite of apps. As such, we believe that the association between app use and outcome can be largely explained by these clinically meaningful use activities,

which clustered into 3 types of activities, reinforcing the importance of self-tracking, goal setting, and psychoeducation elements within mHealth interventions for depression.

More specifically, these 3 clusters of clinically meaningful activities were associated with reductions in depression symptoms at the end of treatment. In particular, self-tracking was found to be beneficial at all levels of intensity compared with minimal intensity of use. This is in accordance with HCI research suggesting that self-tracking, or personal informatics, can lead to behavior change [34], chronic disease management [35], and self-knowledge and self-reflexivity [36,37]. Our study extends this line of research by demonstrating the clinical benefits of self-tracking in the context of mental health. It is important to note that self-tracking in our study consisted of both data collection (ie, tracking) and data reflection (ie, review), as delineated in the stage-based model of personal informatics systems [34]. This suggests that mHealth could better support users by facilitating self-tracking through the incorporation of design features that promote data collection and data reflection.

It is important to note that greater amounts of engagement did not necessarily lead to greater reductions in depression. Although self-tracking was generally beneficial, only a moderate level of engagement with learning and goal setting was associated with reduced depressive symptoms. Neither high nor low intensity of app use could predict better outcomes compared with minimal intensity of use. This result suggests that mHealth interventions might follow the Goldilocks principle—“Not too much. Not too little. Just right” [38]. Just as in many digital technologies, mental health apps do not promise that “the more engagement, the better outcomes;” rather, we argue that the benefits of app use may only be seen when the doses of various classes of intervention features “just right.” Too frequent engagement with goal setting and learning may lead to fatigue, whereas too scarce engagement may lead to ineffectiveness. Thus, mental health technologies should be designed to promote use at the right amount, possibly by sending users reminders based on their app use data. An alternate explanation is that people who engage in learning and goal setting more frequently may be less responsive to treatment. That is, perhaps higher engagement in these activities is associated with a more treatment resistant course of depression, or is exhibited by individuals with more complex comorbidities, thereby indicating that the intervention components do not fit the needs of certain individuals. Thus, high engagement in these activities could be an indicator of risk of lower responsiveness to treatment and could be used to guide the implementation of alternative treatment strategies for individuals who are likely to benefit from additional support.

The overall intensity of generic app use also predicted reductions in depression symptoms. Generally, it appears that people who engaged in higher intensity of app use had lower levels of depression at the end of treatment. However, the duration of app use minimally contributed to better outcomes. This finding corresponds to prior work suggesting that people tend to use mobile apps in very short bursts of time, given their habit of using smartphones in spare moments [39,40]. Although duration of engagement plays a critical role in Web-based interventions [21], the current ways in which people interact with smartphones

suggest that mHealth apps should be designed to be quick to use, have simple interactions, and support a single or limited set of related tasks [2,40]. IntelliCare aligns with these endeavors to facilitate frequent but short interactions.

However, our investigation of meaningful app use was not associated with reduced anxiety symptoms. This is consistent with the findings in the main trial, where significant reductions in anxiety symptom were not related to number of app sessions or time between first and last app use but were only associated with the number of app downloads [27]. The discrepancy in effects of app use on depression and anxiety is a novel finding and suggests that different types of use may be more effective for some psychological states and not others. We speculate that motivation may be an important factor. Users in low motivational states may require self-tracking, goal setting, and learning features in specific doses such that they receive enough to be beneficial but not so much that it overwhelms them. On the other hand, anxiety may be less sensitive to dose responses because it is a more activating condition. To understand such nuance, we need more research specifically designed to examine the relationship between clinically meaningful activities and symptoms across various disorders or symptomatology.

Overall, this study has important implications for the design of mHealth for depression, which includes the following:

- Self-tracking, goal setting, and learning are 3 components that have clinical benefits, which should be incorporated into mental health apps.
- Mental health apps could be designed according to the Goldilocks principle, incorporating the “just right” amount of intervention components and promoting use at the right amount, possibly through sending user reminders or alerts based on app use data.
- People tend to use apps in very short bursts of time, so mental health apps should be quick to use, have simple interactions, and support a single or limited set of related tasks.

However, because of the exploratory nature of the research, design considerations derived from this study focus only on app content and engagement. Within the wider context, research indicates that app design and quality assessment must also consider users’ lived experience, app usability and stability, and data privacy and security [41]. For example, the critical importance of privacy and security in relation to mental health apps was highlighted in 2 recent studies, which suggest that currently available mental health apps often misuse user data [42] and that users’ willingness to share personal sensing data varies depending on the type of data collected and with whom they are shared [43]. Indeed, the importance of these factors is evident within the ever-growing array of app quality measures and guidelines, including those from the American Psychiatric Association [41] and the US Food and Drug Administration [44].

### Limitations

Despite its contributions and implications, this study has some limitations. First, the user activities identified in this study were not exhaustive; some activities were eliminated because of their



low frequency. As a secondary analysis, this study is exploratory by nature, and future studies should continue exploring more specific types and patterns of user behaviors in using mHealth technologies and their relationships with outcomes of mental health conditions. Second, although this study demonstrated the associations of both generic and specific app use with clinical outcomes over the treatment period, it is difficult to make causal claims about the effects. The relationship between app use and symptom change is likely dynamic. For example, app use may contribute to lower subsequent symptoms, and symptom changes may in turn increase app use [45]. Experimental studies are warranted to examine how different intervention components uniquely contribute to outcomes. Third, this study only examined outcomes within the 8-week treatment period. Future research could build on this preliminary model to explore the long-term effects of different app use behaviors.

## Conclusions

Engagement with digital health interventions is a long-standing problem; however, little is known about how users interact with mental health apps in clinically meaningful ways. This study employed a novel, mixed methods methodology to derive greater understanding of users' engagement with apps that cannot be seen through generic use data. Using a combination of qualitative and quantitative methods, we uncovered 3 clusters of clinically meaningful activities—learning, goal setting, and self-tracking—with each type associated with reductions in depression symptoms. However, different activities and intensities of use produced varied effects. Although only moderate intensity of learning and goal setting led to reductions in symptoms of depression, self-tracking at all levels of intensity predicted improvement in depression. Understanding the relationship between different types of user activities and clinical outcomes could inform the design of mental health apps that are more clinically effective for users.

## Acknowledgments

Research was supported by the National Institute of Mental Health (T32 MH115882; R01 MH100482) and National Institute of Diabetes and Digestive and Kidney Diseases (K01 DK116925).

## Conflicts of Interest

DCM has an ownership interest in Adaptive Health, Inc, which has a license from Northwestern University for IntelliCare and has accepted honoraria from Apple Inc. AKG and MJK have received consulting fees from Actualize Therapy, LLC. None of the other authors have conflicts to declare.

## Multimedia Appendix 1

Codebook of the categorization of app use behaviors.

[[XLSX File \(Microsoft Excel File\), 18 KB-Multimedia Appendix 1](#)]

## References

1. Bardram JE, Frost M, Szántó K, Faurholt-Jepsen M, Vinberg M, Kessing LV. Designing Mobile Health Technology for Bipolar Disorder: A Field Trial of the Monarca System. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: CHI'13; April 27-May 2, 2013; Paris, France p. 2627-2636. [doi: [10.1145/2470654.2481364](#)]
2. Mohr DC, Tomasino KN, Lattie EG, Palac HL, Kwasny MJ, Weingardt K, et al. IntelliCare: an eclectic, skills-based app suite for the treatment of depression and anxiety. *J Med Internet Res* 2017 Jan 5;19(1):e10 [FREE Full text] [doi: [10.2196/jmir.6645](#)] [Medline: [28057609](#)]
3. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile apps for bipolar disorder: a systematic review of features and content quality. *J Med Internet Res* 2015 Aug 17;17(8):e198 [FREE Full text] [doi: [10.2196/jmir.4581](#)] [Medline: [26283290](#)]
4. Torous J, Friedman R, Keshavan M. Smartphone ownership and interest in mobile applications to monitor symptoms of mental health conditions. *JMIR Mhealth Uhealth* 2014 Jan 21;2(1):e2 [FREE Full text] [doi: [10.2196/mhealth.2994](#)] [Medline: [25098314](#)]
5. 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. *J Affect Disord* 2017 Aug 15;218:15-22 [FREE Full text] [doi: [10.1016/j.jad.2017.04.046](#)] [Medline: [28456072](#)]
6. Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR Ment Health* 2017 Jun 6;4(2):e19 [FREE Full text] [doi: [10.2196/mental.7785](#)] [Medline: [28588005](#)]
7. Doherty G, Coyle D, Sharry J. Engagement With Online Mental Health Interventions: an Exploratory Clinical Study of a Treatment for Depression. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2012 Presented at: CHI'12; May 5-10, 2012; Austin, Texas, USA p. 1421-1430. [doi: [10.1145/2207676.2208602](#)]
8. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. *Commun Assoc Inf Syst* 2009;24:485-500. [doi: [10.17705/1CAIS.02428](#)]

9. Stawarz K, Preist C, Tallon D, Wiles N, Coyle D. User experience of cognitive behavioral therapy apps for depression: an analysis of app functionality and user reviews. *J Med Internet Res* 2018 Jun 6;20(6):e10120 [FREE Full text] [doi: [10.2196/10120](https://doi.org/10.2196/10120)] [Medline: [29875087](https://pubmed.ncbi.nlm.nih.gov/29875087/)]
10. Gravenhorst F, Muaremi A, Bardram J, Grünerbl A, Mayora O, Wurzer G, et al. Mobile phones as medical devices in mental disorder treatment: an overview. *Pers Ubiquit Comput* 2015;19(2):335-353. [doi: [10.1007/s00779-014-0829-5](https://doi.org/10.1007/s00779-014-0829-5)]
11. Doherty K, Doherty G. Engagement in HCI: conception, theory and measurement. *ACM Comput Surv* 2018;51(5):1-39. [doi: [10.1145/3234149](https://doi.org/10.1145/3234149)]
12. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. *Evid Based Ment Health* 2018 Aug;21(3):116-119. [doi: [10.1136/eb-2018-102891](https://doi.org/10.1136/eb-2018-102891)] [Medline: [29871870](https://pubmed.ncbi.nlm.nih.gov/29871870/)]
13. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. *Am J Prev Med* 2016 Nov;51(5):833-842. [doi: [10.1016/j.amepre.2016.06.015](https://doi.org/10.1016/j.amepre.2016.06.015)] [Medline: [27745683](https://pubmed.ncbi.nlm.nih.gov/27745683/)]
14. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. *Transl Behav Med* 2017 Jun;7(2):254-267 [FREE Full text] [doi: [10.1007/s13142-016-0453-1](https://doi.org/10.1007/s13142-016-0453-1)] [Medline: [27966189](https://pubmed.ncbi.nlm.nih.gov/27966189/)]
15. Short CE, DeSmet A, Woods C, Williams SL, Maher C, Middelweerd A, et al. Measuring engagement in eHealth and mHealth behavior change interventions: viewpoint of methodologies. *J Med Internet Res* 2018 Nov 16;20(11):e292 [FREE Full text] [doi: [10.2196/jmir.9397](https://doi.org/10.2196/jmir.9397)] [Medline: [30446482](https://pubmed.ncbi.nlm.nih.gov/30446482/)]
16. Arapakis I, Lalmas M, Cambazoglu BB, Marcos M, Jose JM. User engagement in online news: Under the scope of sentiment, interest, affect, and gaze. *J Assn Inf Sci Tec* 2014;65(10):1988-2005. [doi: [10.1002/asi.23096](https://doi.org/10.1002/asi.23096)]
17. O'Brien HL, Toms EG. The development and evaluation of a survey to measure user engagement. *J Am Soc Inf Sci* 2010;61(1):50-69. [doi: [10.1002/asi.21229](https://doi.org/10.1002/asi.21229)]
18. Pham Q, Graham G, Carrion C, Morita PP, Seto E, Stinson JN, et al. A library of analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions: scoping review. *JMIR Mhealth Uhealth* 2019 Jan 18;7(1):e11941 [FREE Full text] [doi: [10.2196/11941](https://doi.org/10.2196/11941)] [Medline: [30664463](https://pubmed.ncbi.nlm.nih.gov/30664463/)]
19. Lattie EG, Schueller SM, Sargent E, Stiles-Shields C, Tomasino KN, Corden ME, et al. Uptake and usage of IntelliCare: a publicly available suite of mental health and well-being apps. *Internet Interv* 2016 May;4(2):152-158 [FREE Full text] [doi: [10.1016/j.invent.2016.06.003](https://doi.org/10.1016/j.invent.2016.06.003)] [Medline: [27398319](https://pubmed.ncbi.nlm.nih.gov/27398319/)]
20. Kwasny MJ, Schueller SM, Lattie E, Gray EL, Mohr DC. Exploring the use of multiple mental health apps within a platform: secondary analysis of the IntelliCare field trial. *JMIR Ment Health* 2019 Mar 21;6(3):e11572 [FREE Full text] [doi: [10.2196/11572](https://doi.org/10.2196/11572)] [Medline: [30896433](https://pubmed.ncbi.nlm.nih.gov/30896433/)]
21. Couper MP, Alexander GL, Zhang N, Little RJ, Maddy N, Nowak MA, et al. Engagement and retention: measuring breadth and depth of participant use of an online intervention. *J Med Internet Res* 2010 Nov 18;12(4):e52 [FREE Full text] [doi: [10.2196/jmir.1430](https://doi.org/10.2196/jmir.1430)] [Medline: [21087922](https://pubmed.ncbi.nlm.nih.gov/21087922/)]
22. Mattila E, Lappalainen R, Välkkyinen P, Sairanen E, Lappalainen P, Karhunen L, et al. Usage and dose response of a mobile acceptance and commitment therapy app: secondary analysis of the intervention arm of a randomized controlled trial. *JMIR Mhealth Uhealth* 2016 Jul 28;4(3):e90 [FREE Full text] [doi: [10.2196/mhealth.5241](https://doi.org/10.2196/mhealth.5241)] [Medline: [27468653](https://pubmed.ncbi.nlm.nih.gov/27468653/)]
23. Cussler EC, Teixeira PJ, Going SB, Houtkooper LB, Metcalfe LL, Blew RM, et al. Maintenance of weight loss in overweight middle-aged women through the internet. *Obesity (Silver Spring)* 2008 May;16(5):1052-1060 [FREE Full text] [doi: [10.1038/oby.2008.19](https://doi.org/10.1038/oby.2008.19)] [Medline: [18309301](https://pubmed.ncbi.nlm.nih.gov/18309301/)]
24. Glasgow RE, Christiansen SM, Kurz D, King DK, Woolley T, Faber AJ, et al. Engagement in a diabetes self-management website: usage patterns and generalizability of program use. *J Med Internet Res* 2011 Jan 25;13(1):e9 [FREE Full text] [doi: [10.2196/jmir.1391](https://doi.org/10.2196/jmir.1391)] [Medline: [21371992](https://pubmed.ncbi.nlm.nih.gov/21371992/)]
25. Graham ML, Strawderman MS, Demment M, Olson CM. Does usage of an eHealth intervention reduce the risk of excessive gestational weight gain? Secondary analysis from a randomized controlled trial. *J Med Internet Res* 2017 Jan 9;19(1):e6 [FREE Full text] [doi: [10.2196/jmir.6644](https://doi.org/10.2196/jmir.6644)] [Medline: [28069560](https://pubmed.ncbi.nlm.nih.gov/28069560/)]
26. Davies C, Corry K, van Itallie A, Vandelanotte C, Caperchione C, Mummery WK. Prospective associations between intervention components and website engagement in a publicly available physical activity website: the case of 10,000 Steps Australia. *J Med Internet Res* 2012 Jan 11;14(1):e4 [FREE Full text] [doi: [10.2196/jmir.1792](https://doi.org/10.2196/jmir.1792)] [Medline: [22260810](https://pubmed.ncbi.nlm.nih.gov/22260810/)]
27. Mohr DC, Schueller SM, Tomasino KN, Kaiser SM, Alam N, Karr C, et al. Comparison of the effects of coaching and receipt of app recommendations on depression, anxiety, and engagement in the IntelliCare platform: factorial randomized controlled trial. *J Med Internet Res* 2019 Aug 28;21(8):e13609 [FREE Full text] [doi: [10.2196/13609](https://doi.org/10.2196/13609)] [Medline: [31464192](https://pubmed.ncbi.nlm.nih.gov/31464192/)]
28. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009 Apr;42(2):377-381 [FREE Full text] [doi: [10.1016/j.jbi.2008.08.010](https://doi.org/10.1016/j.jbi.2008.08.010)] [Medline: [18929686](https://pubmed.ncbi.nlm.nih.gov/18929686/)]
29. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. *J Med Internet Res* 2011 Mar 10;13(1):e30 [FREE Full text] [doi: [10.2196/jmir.1602](https://doi.org/10.2196/jmir.1602)] [Medline: [21393123](https://pubmed.ncbi.nlm.nih.gov/21393123/)]

30. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001 Sep;16(9):606-613 [FREE Full text] [doi: [10.1046/j.1525-1497.2001.016009606.x](https://doi.org/10.1046/j.1525-1497.2001.016009606.x)] [Medline: [11556941](https://pubmed.ncbi.nlm.nih.gov/11556941/)]
31. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med* 2006 May 22;166(10):1092-1097. [doi: [10.1001/archinte.166.10.1092](https://doi.org/10.1001/archinte.166.10.1092)] [Medline: [16717171](https://pubmed.ncbi.nlm.nih.gov/16717171/)]
32. Munson S, Consolvo S. Exploring Goal-setting, Rewards, Self-monitoring, and Sharing to Motivate Physical Activity. In: *Proceedings of the 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*. 2012 Presented at: PervasiveHealth'12; May 21-24, 2012; San Diego, CA, USA. [doi: [10.4108/icst.pervasivehealth.2012.248691](https://doi.org/10.4108/icst.pervasivehealth.2012.248691)]
33. Cheung K, Ling W, Karr CJ, Weingardt K, Schueller SM, Mohr DC. Evaluation of a recommender app for apps for the treatment of depression and anxiety: an analysis of longitudinal user engagement. *J Am Med Inform Assoc* 2018 Aug 1;25(8):955-962 [FREE Full text] [doi: [10.1093/jamia/ocy023](https://doi.org/10.1093/jamia/ocy023)] [Medline: [29659857](https://pubmed.ncbi.nlm.nih.gov/29659857/)]
34. Li I, Dey A, Forlizzi J. A Stage-Based Model of Personal Informatics Systems. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2010 Presented at: CHI'10; April 10-15, 2010; Atlanta, Georgia, USA p. 557-566. [doi: [10.1145/1753326.1753409](https://doi.org/10.1145/1753326.1753409)]
35. MacLeod H, Tang A, Carpendale S. Personal Informatics in Chronic Illness Management. In: *Proceedings of Graphics Interface 2013*. 2013 Presented at: GI'13; May 29-31, 2013; Regina, Saskatchewan, Canada.
36. Lupton D. *The Quantified Self*. Malden, MA, USA: Polity Press; 2016.
37. Neff G, Nafus D. *Self-Tracking*. Cambridge, MA: MIT Press; 2016.
38. Turkle S. *Alone Together: Why We Expect More From Technology And Less From Each Other*. New York, NY: Basic Books; 2017.
39. Oulasvirta A, Tamminen S, Roto V, Kuorelahti J. Interaction in 4-second Bursts: the Fragmented Nature of Attentional Resources in Mobile HCI. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2005 Presented at: CHI'05; April 2-7, 2005; Portland, Oregon, USA. [doi: [10.1145/1054972.1055101](https://doi.org/10.1145/1054972.1055101)]
40. Vaish R, Wyngarden K, Chen J, Cheung B, Bernstein M. Twitch Crowdsourcing: Crowd Contributions in Short Bursts of Time. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2014 Presented at: CHI'14; April 26 - May 1, 2014; Toronto, Ontario, Canada p. 3645-3654. [doi: [10.1145/2556288.2556996](https://doi.org/10.1145/2556288.2556996)]
41. Torous JB, Chan SR, Gipson SY, Kim JW, Nguyen T, Luo J, et al. A hierarchical framework for evaluation and informed decision making regarding smartphone apps for clinical care. *Psychiatr Serv* 2018 May 1;69(5):498-500. [doi: [10.1176/appi.ps.201700423](https://doi.org/10.1176/appi.ps.201700423)] [Medline: [29446337](https://pubmed.ncbi.nlm.nih.gov/29446337/)]
42. Huckvale K, Torous J, Larsen ME. Assessment of the data sharing and privacy practices of smartphone apps for depression and smoking cessation. *JAMA Netw Open* 2019 Apr 5;2(4):e192542 [FREE Full text] [doi: [10.1001/jamanetworkopen.2019.2542](https://doi.org/10.1001/jamanetworkopen.2019.2542)] [Medline: [31002321](https://pubmed.ncbi.nlm.nih.gov/31002321/)]
43. Nicholas J, Shilton K, Schueller SM, Gray EL, Kwasny MJ, Mohr DC. The role of data type and recipient in individuals' perspectives on sharing passively collected smartphone data for mental health: cross-sectional questionnaire study. *JMIR Mhealth Uhealth* 2019 Apr 5;7(4):e12578 [FREE Full text] [doi: [10.2196/12578](https://doi.org/10.2196/12578)] [Medline: [30950799](https://pubmed.ncbi.nlm.nih.gov/30950799/)]
44. US Food and Drug Administration. 2019. Digital Health Software Precertification (Pre-Cert) Program. URL: <https://www.fda.gov/medical-devices/digital-health/digital-health-software-precertification-pre-cert-program> [accessed 2019-11-08]
45. Mohr DC, Duffecy J, Jin L, Ludman EJ, Lewis A, Begale M, et al. Multimodal e-mental health treatment for depression: a feasibility trial. *J Med Internet Res* 2010 Dec 19;12(5):e48 [FREE Full text] [doi: [10.2196/jmir.1370](https://doi.org/10.2196/jmir.1370)] [Medline: [21169164](https://pubmed.ncbi.nlm.nih.gov/21169164/)]

## Abbreviations

**GAD-7:** generalized anxiety disorder-7

**HCI:** human-computer interaction

**mHealth:** mobile health

**PHQ-9:** Patient Health Questionnaire-9

*Edited by G Eysenbach; submitted 25.07.19; peer-reviewed by K Stawarz, C Striley, V Strotbaum, C Pasarelu; comments to author 21.09.19; revised version received 04.10.19; accepted 20.10.19; published 20.12.19*

*Please cite as:*

Zhang R, Nicholas J, Knapp AA, Graham AK, Gray E, Kwasny MJ, Reddy M, Mohr DC

*Clinically Meaningful Use of Mental Health Apps and its Effects on Depression: Mixed Methods Study*

*J Med Internet Res* 2019;21(12):e15644

URL: <http://www.jmir.org/2019/12/e15644/>

doi: [10.2196/15644](https://doi.org/10.2196/15644)

PMID: [31859682](https://pubmed.ncbi.nlm.nih.gov/31859682/)

©Renwen Zhang, Jennifer Nicholas, Ashley A Knapp, Andrea K Graham, Elizabeth Gray, Mary J Kwasny, Madhu Reddy, David C Mohr. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 20.12.2019. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.