How Does Engagement Change over Time in a Digital Eating Disorder App?

Rachael E. Flatt, Laura M. Thornton, Jenna Tregarthen, Stuart Argue, Cynthia M. Bulik

Submitted to: JMIR mHealth and uHealth
on: November 14, 2024

Table of Contents

Original Manuscript **5**
Supplementary Files **45**
Figures **46**
Figure 1 .... **47**

How Does Engagement Change over Time in a Digital Eating Disorder App?

Rachael E. Flatt[1] PhD; Laura M. Thornton[1] PhD; Jenna Tregarthen[2] PGDip; Stuart Argue[2] BS; Cynthia M. Bulik[3, 4] PhD

[1]University of North Carolina at Chapel Hill Department of Psychiatry Chapel Hill US
[2]Recovery Record (United States) Palo Alto US
[3]Karolinska Institutet Stockholm SE

Corresponding Author:
Cynthia M. Bulik PhD

Abstract

Background: Engagement with digital mental health interventions is often measured as a summary-level variable and remains under-researched despite its importance for meaningful symptom change.

Objective: The current study deepens understanding of engagement in a digital eating disorder intervention by measuring engagement with unique components of the app, on two different devices (phone and watch), and at a summary-level.

Methods: Participants with current binge-eating behavior were recruited as part of the Binge Eating Genetics Initiative (BEGIN study) to use a digital eating disorder intervention for 4 weeks. Demographic and severity of illness variables were captured in the baseline survey at enrollment, and engagement data were captured through both an iPhone and Apple Watch version of the intervention. Engagement was characterized by log type (urge, behavior, mood, or meal), device type (logs on phone or watch), and overall usage (total logs) and averaged each week for four weeks. Descriptives were tabulated for demographic and engagement variables, and multilevel growth models were conducted for each measure of engagement with baseline characteristics and time as predictors.

Results: Participants (n=893) self-reported as primarily White (n=743, 85.3%), non-Hispanic (n=92; 10.3%), females (n=772, 86.5%) with a mean age of 29.6 years (SD=7.4) and mean current BMI of 32.5 kg/m2 (SD=9.8) and used the app for a mean of 24 days. Most logs were captured on phones (96%), and mood logs were the most used app component (62% of logs). All measures of engagement declined over time, as illustrated by the visualizations, but each measure of engagement illustrated unique participant trajectories over time. Time was a significant negative predictor in every multilevel model. Sex and ethnicity were also significant predictors across several measures of engagement, with female and non-Hispanic participants demonstrating greater engagement than male and Hispanic counterparts. Other baseline characteristics (age, current BMI, and binge episodes in the past 28 days) were significant predictors of one measure of engagement each.

Conclusions: This study highlighted that engagement is far more complex and nuanced than is typically described in research, and that specific components and mode of delivery may have unique engagement profiles and predictors. Future work would benefit from developing early engagement models informed by baseline characteristics to predict intervention outcomes, thereby tailoring digital eating disorder interventions at the individual level.

(JMIR Preprints 14/11/2024:68824)
DOI: https://doi.org/10.2196/preprints.68824

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Title: How Does Engagement Change over Time in a Digital Eating Disorder App?

Author List: Rachael E. Flatt Ph.D.[1], Laura M. Thornton, Ph.D.[1], Jenna Tregarthen, PGDip[2], Stuart Argue, B.S.[2], Cynthia M. Bulik, Ph.D.[1,3,4]

Affiliations:
[1] Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[2] Recovery Record, Palo Alto, California, USA

[3] Department of Nutrition, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[4] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

Correspondence: Cynthia M. Bulik, Ph.D., FAED
University of North Carolina at Chapel Hill
CB #7160, 101 Manning Drive
Chapel Hill, NC 27599-7160

e-mail: cynthia_bulik@med.unc.edu
phone: 984-974-3233

Abstract Word Count: 372/450

Main Text Word Count: 5783/6000

Keywords: eating disorders, wearable technology, mobile application, engagement

Funding:

National Science Foundation (DGE-1650116; Flatt, PI); Foundation of Hope, Raleigh, North Carolina (Bulik, PI); National Eating Disorders Association (Bulik and Tregarthen, PIs); Brain and Behavior Research Foundation (NARSAD Distinguished Investigator Grant; Bulik, PI); National Institute of Mental Health (R01MH119084, Bulik/Butner, MPIs; U01 MH109528, Sullivan PI, Bulik Co-I). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Apple Watch devices were provided to University of North Carolina at Chapel Hill by Apple Inc. at no cost to the study.

No funding bodies were involved in the design of the study and collection, analysis, interpretation of data, or writing the manuscript.

Abstract

Background: Engagement with digital mental health interventions is often measured as a summary-level variable and remains under-researched despite its importance for meaningful symptom change. The current study deepens understanding of engagement in a digital eating disorder intervention by measuring engagement with unique components of the app, on two different devices (phone and watch), and at a summary-level.

Methods: Participants with current binge-eating behavior were recruited as part of the Binge Eating Genetics Initiative (BEGIN study) to use a digital eating disorder intervention for 4 weeks. Demographic and severity of illness variables were captured in the baseline survey at enrollment, and engagement data were captured through both an iPhone and Apple Watch version of the intervention. Engagement was characterized by log type (urge, behavior, mood, or meal), device type (logs on phone or watch), and overall usage (total logs) and averaged each week for four weeks. Descriptives were tabulated for demographic and engagement variables, and multilevel growth models were conducted for each measure of engagement with baseline characteristics and time as predictors.

Results: Participants (n=893) self-reported as primarily White (n=743, 85.3%), non-Hispanic (n=92; 10.3%), females (n=772, 86.5%) with a mean age of 29.6 years (SD=7.4) and mean current BMI of 32.5 kg/m^2 (SD=9.8) and used the app for a mean of 24 days. Most logs were captured on phones (96%), and mood logs were the most used app component (62% of logs). All measures of engagement declined over time, as illustrated by the visualizations, but each measure of engagement illustrated unique participant trajectories over time. Time was a significant negative predictor in every multilevel model. Sex and ethnicity were also significant predictors across several measures of engagement, with female and non-Hispanic participants demonstrating greater engagement than male and Hispanic counterparts. Other baseline characteristics (age, current BMI, and binge episodes in the past 28 days) were significant predictors of one measure of engagement each.

Conclusions: This study highlighted that engagement is far more complex and nuanced than is typically described in research, and that specific components and mode of delivery may have unique engagement profiles and predictors. Future work would benefit from developing early engagement models informed by baseline characteristics to predict intervention outcomes, thereby tailoring digital eating disorder interventions at the individual level.

Introduction

Digital interventions hold substantial promise for addressing mental health disorders, including eating disorders (EDs). Traditional face-to-face ED treatment is often cost-prohibitive, especially for the uninsured; difficult to access; not scalable; retrospective in nature; and unable to offer support in real time when patients need it most [1-3]. Digital ED interventions address these obstacles by offering evidence-based and in-the-moment treatment options that are affordable and accessible [4-7]. A crucial aspect of digital interventions is engagement [8,9], which parallels measures of retention and adherence in face-to-face treatments. Just like individuals attend sessions, complete self-monitoring forms for home practice, and apply learned skills outside of sessions, digital interventions for EDs can measure similar and expanded forms of engagement that may be key to achieving and maintaining treatment gains.

Similar to observations in the broader digital mental health tool literature, engagement in digital ED interventions is low [6], the definitions used to describe engagement are heterogeneous and mostly limited to static or summary-level variables, and predictors of engagement are not well researched. Thus, this study analyzes engagement in a digital ED intervention, Recovery Record, over 30 days for individuals with binge-type EDs. Specifically, we model longitudinal growth curves to examine the dynamic trajectories of several measures of engagement and subsequently identify baseline predictors.

Engagement challenges in digital mental health interventions

As with the adoption of any new technology, both users and researchers have encountered several challenges with digital interventions. Common concerns include inadequate attention toward management of adverse events, inconsistent data management and privacy policies, perceived and actual logistic barriers to integration into clinical settings, and potential harm of using interventions that are unsupported or untested [10,11], all of which contribute to initial uptake and retention. In a review of digital mental health interventions, the median retention of users in the first 15 and 30 days of use is 3.9% and 3.3%, respectively [12]. Compounding poor retention is that user engagement is notoriously low and usually decreases over time [10], but research on the dynamic nature of engagement, and more specifically how engagement changes over time across delivery platforms and app features, is limited. When individuals download digital mental health apps seeking help, maintaining engagement is necessary for adequate dosing of an intervention or support; therefore, understanding engagement and strategies for improving and maintaining engagement are important research goals.

Engagement, broadly defined in this paper as the ways and extent to which an individual uses a digital intervention, is frequently cited as a key challenge [8,13,14; see Footnote 1]. From a user perspective, numerous factors may contribute to changes in and difficulties with sustained engagement over time including personal preferences, severity of illness, perceived usefulness of and evidence supporting the intervention, technological challenges, lack of personalized features within the intervention, and environmental/social factors [8]. Researchers also contend with methodological challenges when attempting to evaluate engagement in digital interventions. There are seemingly endless conceptualizations and both qualitative and quantitative measurements of engagement [14-16], and most studies evaluating engagement do not capture multiple data streams to illustrate dynamics, in part because engagement is usually a secondary aim to treatment outcomes. The types of engagement data that can be collected and analyzed also vary widely depending on the features and content delivered to users through the intervention, the data users consent to provide, and the mode of delivery (e.g., through a smartphone app, wearable technology, web-

[1] Although different terms have been used to describe how an individual uses a digital mental health app (engagement, app usage, adherence, etc.), we will use the term engagement through the remainder of this paper for consistency—even if the authors used different terminology—and will elaborate on the specific types of engagement measured in subsequent studies as needed.

based browser). In research studies, investigative teams also typically use strategies such as reminders, phone calls, and compensation to boost user engagement [6], which may inflate overall engagement results compared with naturalistic app use. As a result, the field has yet to elucidate common themes about the natural changes in engagement over time.

Engagement in digital eating disorder interventions

Across studies reporting on engagement in digital ED interventions, similar issues to the general digital mental health intervention research base arise. First, barriers to initiating engagement identified across qualitative and quantitative studies of digital ED interventions mirrored those of other digital mental health interventions and included logistical constraints such as cost, accessibility, usability/functionality [17], time [18], and privacy of personal health information [19]. Perceived barriers include treatment credibility and expectancy [20], motivation, accountability, content and feature preferences [17], severity of illness, trust of the intervention [18], satisfaction, intervention personalization, and ease of use [19.] Second, the definitions of engagement vary widely: for example, studies defined engagement as number of modules completed [21]; total number of logs completed, total number of days active with the app, and length of time using the app [22]; interactivity and usability [23]; and total app views and total number of meal logs [24]; among others.

Third, several papers also cited maintaining engagement as a significant challenge: the median percentage of users in a review of treatment studies of digital ED interventions where all prescribed modules or activities were completed was 36% and the median percentage of individuals who never accessed the interventions was 15% [6]. A few studies also reported that greater engagement was associated with better treatment outcomes [21,22], suggesting that maintaining engagement is important for users seeking meaningful change in their ED symptoms. Notably, most of these definitions only captured summary-level or endpoint measurements of engagement, so little is known about how engagement changes over time and if changes in engagement contribute to treatment gains.

The current study uses data collected through Recovery Record [3], a widely used evidence-based ED app that has been downloaded over 1 million times; however, only two studies assessed participant engagement in Recovery Record to date. The first qualitative study explored engagement in a Danish-translated version of Recovery Record in participants (n=41) with anorexia nervosa (AN) or bulimia nervosa (BN) [25], the majority of whom used the app between one and four months. Participants reported that engaging with the app helped them confront the ED or log meals more constructively leading to less concern with caloric intake, but also reported that engagement could be obstructive by increasing obsessions with logging or by giving participants ideas about other compensatory behaviors (e.g., participant sees "excessive exercise" as an option to log and begins to feel an urge to exercise). The second study conducted by Kim and colleagues [26] evaluated participants' engagement with Recovery Record in a sample of Recovery Record users as part of a larger randomized controlled trial testing efficacy of the app. The total number of meal logs and total number of days the app was used significantly and positively mediated the treatment effect on clinical response as measured by the Eating Disorders Examination Questionnaire v6 (EDE-Q) [27] eight weeks later, indicating that this greater engagement could lead to more positive treatment outcomes. However, only three measurements of engagement were sufficiently tested as mediators, and the study did not describe or evaluate changes in engagement over time.

Despite the multitude of interactive features that most digital ED apps offer, including Recovery Record, most digital ED research has focused on heterogenous summary-level measurements of engagement (e.g., number of days the app was used), and the analyses do not reflect the dynamic nature of engagement. In addition, baseline contextual factors that may influence engagement trajectories have not yet been thoroughly explored with digital ED interventions. As a result, valuable information pertaining to engagement trajectories are mostly unavailable, and we may fail to capture how different types of engagement change over time and key participant characteristics that are associated with engagement. This study addresses these issues by defining several types of engagement based on the available interactive features in Recovery Record, modeling trajectories for each measure of engagement, and identifying key baseline predictors of engagement.

Aims

The aims of the current study were to: 1) describe characteristics of the sample and how individuals engaged with the app across a variety of measures of engagement; 2) model the trajectories of engagement over the 30-day course of the study; and 3) identify baseline demographic and ED symptom predictors of engagement. Based on previously published literature, we hypothesized that individuals will generally demonstrate downward trajectories of engagement across 30 days. However, the dearth of evidence on engagement in digital ED interventions and in Recovery Record does not support more specific a priori hypotheses for baseline predictors of engagement.

Methods

Participants

Participants were recruited as part of a larger parent case-only trial, the Binge Eating Genetics Initiative (BEGIN) study. The full study protocol for the parent trial is available elsewhere [28]. Inclusion criteria for the current study included: 1) current binge eating; 2) lifetime diagnosis of either BN or binge-eating disorder (BED); 3) U.S. resident; 4) between 18-45 years old; 5) reads and speaks English; 6) current iPhone user; 7) ambulatory; 8) willing and able to participate in the study, wear an Apple Watch, and use Recovery Record. One additional criterion for analysis in the current study was completion of at least one log on Recovery Record.

Exclusion criteria included: 1) currently pregnant or breastfeeding; 2) history of bariatric surgery; 3) current use of hormone therapy; 4) inpatient treatment or hospitalization for ED in the two weeks prior to enrollment; 5) current suicidality; and 6) antibiotic or probiotic use at enrollment. Of note, some exclusion criteria were related to other aspects of the parent trial (e.g., microbiome testing).

Procedures

All study procedures were approved by the University of North Carolina Biomedical Institutional Review Board.

Participants were recruited primarily through Recovery Record, social media posts, and email listservs. After completing three logs on Recovery Record and demonstrating initial engagement with the app, interested individuals were invited to complete online consent forms following by a screener for lifetime ED diagnosis using the ED100Kv2 [29,30]. Those who met all inclusion criteria were offered a second consent and the option to participate in the 30-day study. They were subsequently asked to complete a baseline questionnaire consisting of several demographic questions and measures assessing ED and general psychopathology (depression, anxiety, attention-deficit/hyperactivity disorder screeners). Packages containing an Apple Watch (if they did not already have one) and sampling kits (for genetic and microbiome sampling) were then sent to participants within the first few days of enrollment.

Participants were asked to use the Recovery Record, a cognitive behavioral therapy-based application designed to support individuals with an ED, for 30 days through an iPhone and a Apple Watch (1st generation) with a version of the Recovery Record app designed specifically for the parent trial. Participants were asked to log ED urges and behaviors including binge eating and compensatory behaviors (vomiting, diuretics/laxative misuse, excessive exercise, and fasting) and their mood through the Recovery Record app on the Apple Watch, although these logs could also be completed on the iPhone app. In addition, participants logged their meals on the iPhone app rather than the Apple Watch given the larger screen. Skills (e.g., distraction, mindful breathing, emotion regulation, challenging negative thoughts) were also available for participants to use. Individuals who were already working with a clinician outside of the study could connect with them through the Recovery Record app; however, this was not included as a part of engagement data collection.

Midpoint and endpoint questionnaires assessing ED, mood, and anxiety symptoms were administered 15- and 30-days, respectively, after enrollment. Individuals did not receive compensation for their participation, but those who were sent an Apple Watch were able to keep the devices at the end of the study.

Measures

Demographic information was collected via a questionnaire administered at baseline on age, gender, race, and ethnicity. Biological sex at birth was determined via saliva sample for the genetic testing component of the parent study [31]. Current body mass index (BMI) was calculated at baseline with self-reported height and weight. Lifetime ED diagnosis was determined by algorithm using items from the ED100Kv2 [29,30].

Information on current ED symptomatology was collected through the EDE-Q [27] administered in the baseline, midpoint, and endpoint questionnaires. The EDE-Q is a widely used self-report ED questionnaire and has demonstrated good validity and reliability in community samples [32]. Twenty-eight items cover various aspects of ED pathology including weight/shape/eating concerns, current BMI, and ED behaviors including binge eating and compensatory behavior frequency (vomiting, fasting, excessive exercise, diuretic/laxative misuse) over the past 28 days. The EDE-Q global score [33,34] was calculated from the 22 Likert-scale items (0=no days, 6=every day for items assessing frequency OR 0=not at all, 6=markedly for items assessing distress/impairment).

Engagement

To extend the focus of existing literature on summary-level variables and broaden the types of engagement measured, we assessed the interactive features Recovery Record offers. In addition, since a central component of evidence-based treatment for binge-type EDs is self-monitoring of meals, mood, urges, and ED behaviors to help patients identify triggers and maintaining factors of the ED [27], participants were explicitly instructed to focus on logging these aspects through the app. All data used to describe types of engagement were collected over 30 days through the Recovery Record app on both the iPhone and the Apple Watch. As such, the types and definitions of engagement used for this study, presented in Table 1, focus on use of meal/mood/behavioral logs, overall usage, and through which mode of delivery (i.e., iPhone or Apple Watch). All measures of engagement defined in Table 1 take an approach of mean usage (i.e., the average number of times per day in one week that an individual used part of the app) to help capture change in engagement over time. To characterize how participants engaged with the app, each type of engagement was tabulated over week-long periods, thus participants can have up to four repeated measures of engagement for weeks 1-4. However, since participants received their Apple Watch devices approximately one week into the study, we only include data from weeks 2-4; therefore, participants have up to three repeated measures of engagement in the watch log models only. Week 1 data collection began the day after enrollment.

Table 1: Engagement terms and descriptions of engagement definitions, each used as the

dependent variable in separate multilevel models

Engagement Measure	Definition of Engagement
Log Type	
Mean urge logs	Number of times a participant logged an urge during a day, **averaged** over a 7-day timespan
Mean eating behavior logs	Number of times a participant logged an eating behavior during a day, **averaged** over a 7-day timespan
Mean mood logs	Number of times a participant logged a mood during a day, **averaged** over a 7-day timespan
Mean meal logs	Number of times a participant logged a meal during a day, **averaged** over a 7-day timespan
Device Type	
Mean phone logs	Number of times the app was used on the iPhone in any capacity during a day, **averaged** over a 7-day timespan
Mean watch logs*	Number of times the app was used on the Apple Watch in any capacity during a day, **averaged** over a 7-day timespan
Mean Use	The number of times the app was used in any capacity during a day, **averaged** over a 7-day timespan

*Note: due to participants receiving their Apple Watch devices ~1 week into the study, we only report data from weeks 2-4.

Data analysis

All data preparation and analyses were conducted using SAS 9.4 [35]. To prepare data for analysis, we first screened for unrealistic values of binge eating and compensatory behavior episodes reported at baseline (i.e., 500 episodes reported in the past 28 days), and Winsorized those values. For model estimation and interpretation purposes and to protect the privacy of participants with demographic characteristics with cell sizes <5, we included male and female participants who reported data for all baseline predictor variables (age, sex, ethnicity, current BMI, and baseline binge-eating episodes). For the engagement data, we screened for and removed duplicates and impossible or improbable measures of engagement (i.e., future-dated timepoints, 1,000 logs of the same event in one day); no imputation methods were used since lack of engagement data at a given time point were not necessarily indicative of missingness.

To address aim 1, we first characterized the sample at baseline. Descriptive statistics (n's, percentages, means and SDs as appropriate depending on variable type) were provided on demographic variables (age, gender, sex, race, ethnicity, and current BMI), ED diagnosis, baseline ED psychopathology (EDE-Q global scores and number of binge-eating and compensatory behavior episodes in the past 28 days as measured by the EDE-Q), and for each measure of engagement listed in Table 1. Demographic characteristics with cell sizes <5 were not reported to protect the privacy of participants. We also performed a Poisson regression of the total number of days participants used Recovery Record in any capacity. Baseline predictors (age, sex, ethnicity, current BMI, and number of binge-eating episodes at baseline) were used as independent variables. Race was not used due to small cell sizes.

For aim 2, we analyzed the engagement data using multilevel growth models due to the nested structure of the data (up to four repeated measures within individuals for all measures of engagement except for mean watch logs, which is up to three repeated measures of engagement due to participants receiving their devices by the end of week 1), using the types of engagement included in Table 1 as the dependent variables. First, we visualized the data by plotting each measure of engagement over time (measured in weeks) using spaghetti plots to identify what type of functional form should be used (i.e., linear, piecewise) for each model. For each measure of engagement, we began with unconditional multilevel models using the multilevel model PROC MIXED function. If the spaghetti plots were unclear as to what functional form should be used, we compared the intraclass correlation (ICC), Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC) values in the initial unconditional models to determine which functional form was most appropriate. Findings from the spaghetti plots and unconditional models were qualitatively summarized in the text to highlight relevant themes.

For aim 3, we expanded analyses to conditional models of engagement by including time (level 1 predictor measured in weeks) and the following baseline demographic predictors as time-invariant covariates (level 2 predictors): age, sex, ethnicity, and current BMI. Given that the parent sample recruited individuals with current binge eating, we then added baseline number of binge-eating episodes in the past 28 days from the EDE-Q as a measure of severity of illness to use as a level 2 time-invariant predictor of engagement. We mean centered age, current BMI, and baseline binge-eating episodes for ease of interpretation. For the categorical predictors, the reference variables were set to week 1 for time, participants who were categorized as male based on genotype for sex, and participants who self-identified as Hispanic for ethnicity. Predictors with significant fixed effects estimates were reported for each model, and themes are summarized in the text.

All multilevel models were fit using restricted maximum likelihood, the alpha level for fixed effects was set to .05, and we used Satterthwaite degrees of freedom approximations to reduce Type I error rates [36]. We limited the models to fixed effects to help with estimation that was consistent across models, especially given that the predictors of interest were primarily level 2 time-invariant predictors within the context of the current study, and plan to further explore random effects in future analyses.

Results

Aim 1: Demographics and sample description

A total of 893 participants engaged with Recovery Record at least once during the 30-day study period, had complete data for all baseline predictors, and were included in the current study. Of those, 86.5% (n=772) were assigned as female based on genotype and 13.5% (n=121) as male; 84.1% (n=680) self-identified as women, 14.8% (n=120) as men, and 1.1% (n=9) as non-binary/third gender (n=84 did not report gender). The sample mostly identified as White (85.3%, n=743), followed by more than one race (6.2%; n=54), African American (4.0%, n=35), Asian (3.7%; n=33), and Native American/American Indian (0.7%, n=6); n=22 did not report race. In addition, 10.3% (n=92) identified as Hispanic. The mean age of the sample was 29.6 years (SD=7.4), and the mean current BMI was 32.5 kg/m^2 (SD=9.8).

Across eating disorder characteristics, 78.3% (n=699) met ED100Kv2 criteria for lifetime BED, 26.9% (n=240) for lifetime BN, and 18.7% (n=167) for lifetime AN see Footnote [2]. The mean EDE-Q global score was 3.93 (SD=1.01). The mean EDE-Q subscale scores were 3.65 (SD=1.25) for eating concern, 2.90 (SD=1.58) for restraint, 4.71 (SD=1.12) for shape concern, and 4.44 (SD=1.12) for weight concern. At baseline, participants reported a mean of 12.92 (SD=9.57) binge episodes, 3.00 (SD=9.46) vomiting episodes, 0.89 (SD=3.43) laxative/diuretic misuse episodes, and 4.16 (6.71) compulsive exercise episodes in the past 28 days as captured by the EDE-Q.

For measures of engagement, the sample used the Recovery Record app for an average of 24.09 days (SD=7.18) out of 30, with 23.40% (n=209) using the app all 30 days. Across all 893 participants included in the current study, a total of 225,927 Recovery Record logs were captured over 4 weeks. See Table 2 for descriptive data on the total sum of logs for each engagement variable over 4 weeks as well as the mean number of logs per day for each engagement variable over 4 weeks and by week. Notably, the majority of logs were completed on the iPhone (96.11%) rather than on the Apple Watch. After centering age, current BMI, and number of binge episodes at baseline, age (ß=.00, χ^2 (1)=5.74, FDR-adjusted p-value =.03 and sex (ß=.06, χ^2 (1)=8.18, FDR-adjusted p-value <.01) were significant predictors of the total number of days participants used the Recovery Record app. Specifically, participants who were older and female used the app a greater number of days.

Table 2: Total sum and mean number of logs per day across four weeks and by week for each engagement variable.

[2] Participants could meet criteria for more than one lifetime ED diagnosis.

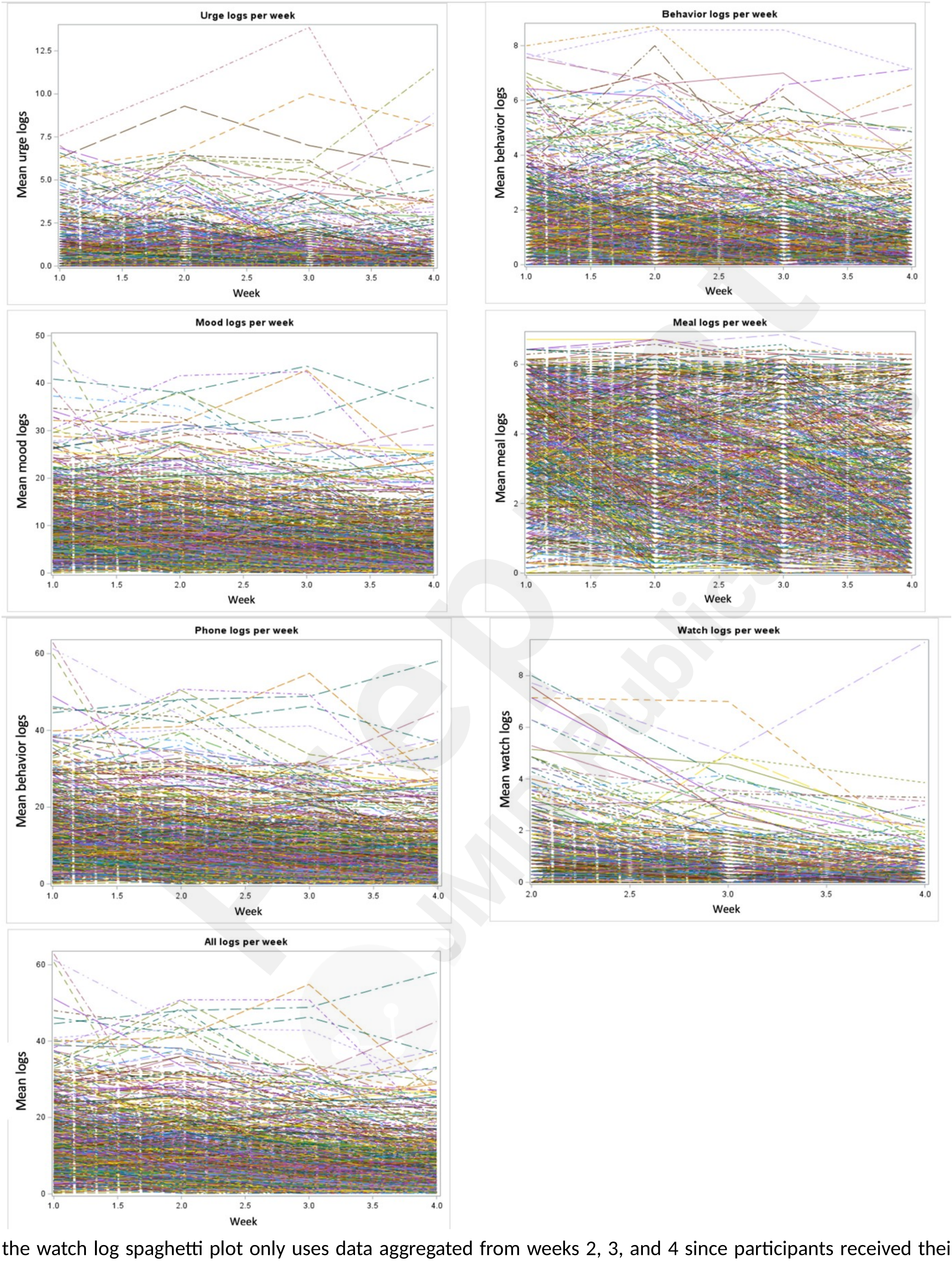
Engagement Variable	Total sum of logs over all 4 weeks	Mean logs per day (SD) over all 4 weeks	Mean logs per day (SD) during week 1 *n=893*	Mean logs per day (SD) during week 2 *n=835*	Mean logs per day (SD) during week 3 *n=784*	Mean logs per day (SD) during week 4 *n=681*	Percent change from week 1 to week 4 (%)
Log Type							
Urge logs	14,855	0.66 (1.13)	0.80 (1.17)	0.81 (1.21)	0.58 (1.09)	0.42 (0.99)	-47.50
Behavior logs	25,420	1.14 (1.29)	1.48 (1.40)	1.26 (1.36)	0.96 (1.17)	0.74 (1.01)	-50.00
Mood logs	174,818	7.83 (6.40)	9.77 (6.55)	8.31 (6.54)	6.96 (6.09)	5.71 (5.50)	-41.56
Meal logs	67,043	3.00 (1.79)	3.58 (1.64)	3.13 (1.75)	2.77 (1.79)	2.35 (1.75)	-34.36
Device Type							
Phone logs	217,143	9.73 (8.17)	12.42 (8.59)	10.24 (8.27)	8.51 (7.58)	6.99 (6.92)	-43.72
Watch logs*	6,977	0.43 (.87)	-	0.65 (1.10)	0.39 (0.76)	0.22 (0.58)	-67.69
All Logs	225,927	10.12 (8.41)	12.71 (8.75)	10.90 (8.64)	8.90 (7.82)	7.21 (7.06)	-43.27

*Note: due to participants receiving their Apple Watch devices ~1 week into the study, we only report watch log data from weeks 2-4.

Aim 2: Characterization of engagement

To illustrate participant engagement over the course of four weeks (three weeks for the watch logs) and to evaluate the functional forms to be used for subsequent conditional multilevel models, spaghetti plots for each measure of engagement were created using the mean number of logs over each week, seen in Figure 1. Across each measure, there was a general downward and linear trajectory of engagement across time. The mean meal logs plot had the most variability in individual trajectories and was initially difficult to discern a clear functional form. However, after visualizing the data via spaghetti plots in smaller groups of participants (i.e., n=100) in combination with the data from Table 2, both linear and quadratic functional forms were tested in unconditional models. After comparing the ICC, AIC, and BIC, a linear functional form was determined to be the best fit for each measure of engagement and was used in the subsequent multilevel models detailed in the next section.

Figure 1: Spaghetti plots of each engagement measure across four weeks.



Note: the watch log spaghetti plot only uses data aggregated from weeks 2, 3, and 4 since participants received their Apple Watch devices by the end of week 1.

Aim 3: Predictors of engagement

Table 3 presents the results from the conditional models evaluating which baseline characteristics were significant predictors of engagement. Time, measured in weeks, was a significant negative predictor in nearly every instance, indicating that the number of weeks into the study consistently predicted a decline in engagement, regardless of how it was measured. Across demographic characteristics, sex and ethnicity were significant, positive predictors of mean mood, meal, and phone logs, as well as mean use. Mean-centered age was a significant positive predictor of mean meal log engagement such that those who were older logged more meals, and mean-centered current BMI was a significant negative predictor of mean urge logs such that those with higher BMIs at baseline logged fewer urges.

In the second set of models, the number of binge episodes at baseline was a significant, positive predictor of mean behavior logs. However, number of binge episodes at baseline was not a significant predictor of any other measure of engagement, and there were no changes to significance for any other predictors when this variable was added to each engagement model. Notably, the model fit statistics and parameter estimates demonstrated negligible changes between models 1 and 2 for all measures of engagement.

Table 3: Multilevel model summaries for baseline demographic and severity of illness variables predicting engagement measures.

Engagement Variable	Model	Week ß (SE) 2	Week ß (SE) 3	Week ß (SE) 4	Age ß (SE)	Sex ß (SE) Female	Ethnicity ß (SE) Non-Hispanic	Current BMI ß (SE)	Binge Episodes ß (SE)	ICC	AIC	BIC
Mean urge logs	Null									0.58	8545.3	8554.9
	Model 1	-.01 (.03)	-.25 (.03)***	-.47 (.04)***	.00 (.00)	.12 (.09)	-.03 (.10)	-.01 (.00)**		0.61	8356.3	8365.9
	Model 2	-.01 (.03)	-.26 (.03)***	-.47 (.04)***	.00 (.00)	.12 (.09)	-.03 (.10)	-.01 (.00)**	.00 (.00)	0.61	8364.5	8374.1
Mean behavior logs	Null									.58	9416.4	9426.0
	Model 1	-.25 (.03)***	-.59 (.04)***	-.87 (.03)***	.01 (.01)	.00 (.11)	.15 (.12)	.00 (.00)		.66	8923.0	8932.6
	Model 2	-.26 (.04)***	-.59 (.04)***	-.87 (.04)***	.01 (.01)	-.02 (.11)	.15 (.12)	.00 (.00)	.01 (.00)*	.65	8937.3	8946.9
Mean mood logs	Null									.63	19441.0	19450.6
	Model 1	-1.73 (.16)***	-3.34 (.16)***	-5.10 (.17)***	.05 (.03)	2.39 (.55)***	1.97 (.62)***	-.03 (.02)		.73	18604.7	18614.3
	Model 2	-1.73 (.16)***	-3.34 (.16)***	-5.10 (.17)***	.05 (.03)	2.37 (.55)***	1.97 (.62)***	-.03 (.02)	.01 (.02)	.73	18610.2	18619.8
Mean meal logs	Null									.57	11654.1	11663.7
	Model 1	-.55 (.05)***	-1.01 (.05)***	-1.59 (.05)***	.02 (.01)**	.61 (.15)***	.60 (.17)**	.00 (.01)		.70	10790.7	10800.3
	Model 2	-.55 (.05)***	-1.01 (.05)***	-1.59 (.05)***	.02 (.01)**	.63 (.15)***	.60 (.17)**	.00 (.01)	.00 (.01)	.70	10795.4	10805.0
Mean phone logs	Null									.61	21068.0	21077.6
	Model 1	-2.52 (.20)***	-4.57 (.21)***	-6.70 (.22)***	.05 (.03)	2.84 (.70)***	2.22 (.79)**	-.05 (.03)		.72	20205.6	20215.2
	Model 2	-2.52 (.20)***	-4.57 (.21)***	-6.70 (.22)***	.05 (.03)	2.81 (.70)***	2.22 (.79)**	-.05 (.03)	.03 (.02)	.72	20209.9	20219.4
Mean watch logs	Null									.48	5412.4	5421.8
	Model 1	n/a	-.28 (.03)***	-.48 (.03)***	.01 (.00)	-.09 (.07)	-.12 (.08)	-.00 (.00)		.55	5209.8	5219.3
	Model 2	n/a	-.28	-.48	.00	-.10	-.12	-.00	.00	.55	5217.0	5226.4

			(.03)***	(.03)***	(.00)	(.07)	(.08)	(.00)	(.00)			
	Null									.61	21246.8	21256.4
Mean use	Model 1	-2.17	-4.50	-6.83	.06	2.81	2.19	-.05		.72	20392.7	20402.3
		(.21)***	(.21)***	(.22)***	(.03)	(.72)***	(.81)**	(.03)				
	Model 2	-2.17	-4.50	-6.83	.06	2.76	2.19	-.05	.03	.72	20392.7	20402.3
		(.21)***	(.21)***	(.22)***	(.03)	(.72)***	(.81)**	(.03)	(.03)			

FDR-corrected p-values: *<.05, **<.01, ***<.001
AIC: Akaike Information Criterion
BIC: Bayesian information criterion
ICC: Intraclass correlation
Notes: The reference values for categorical variables were set to week 1, males, and individuals who identified as Hispanic. The watch log models only use data aggregated from weeks 2, 3, and 4 since most participants received their Apple Watch devices by the end of week 1.

Discussion

This study described various measures of engagement with a digital ED app, Recovery Record, and deepened our understanding of how individuals with binge eating use different components and delivery methods of the app. All measures of engagement declined over the course of the study, consistent with trends observed in other digital ED and mental health interventions; however, participants engaged with the app for an average of ~3.5 weeks, which was greater than expected given that most digital mental health interventions observe engagement rates that are much lower [12,37,38]. Several baseline variables emerged as significant predictors of unique measures of engagement, highlighting the importance of more nuanced assessments of engagement with digital ED interventions. Findings for each aim are discussed in turn below, followed by notable discussion points for measures of engagement and study limitations.

Principal results

Although every measure of engagement declined, the percentage of participants in this study that were still using Recovery Record at 30 days (23%) was substantially greater than that observed in other self-monitoring apps (6%) [12]. Across log types, mean behavior and urge logs had the largest percent reduction over four weeks, which could reflect both the overall decline in engagement in combination with decreased ED symptomatology as observed in the BEGIN feasibility study [39]. In addition, participants logged an average of ~3 meals/day, and mean meal logs had the smallest percent reduction (34%) of any measure of engagement over the course of four weeks, illustrating that most participants were on track with a regular eating treatment target in cognitive behavioral ED treatments [27].

Mood logs were the most frequently utilized logs, accounting for 62% of all logs, which is in part explained by participants having the ability to create a unique mood log for each emotion as part of meal or behavior log, or as a separate mood log altogether. The greater number of opportunities to log moods in addition to the more limited nature of when meals and ED behaviors or urges occur may in large part explain the vast difference in sample sizes between log types. When measuring engagement across device type, participants completed substantially more phone logs than watch logs, which could partially be explained by meal logs only being offered on the phone. However, the percent reduction in watch logs was 35% greater than the percent reduction of phone logs over the last 3 weeks, demonstrating that the decline in watch engagement was much steeper than the phone engagement. Although the watch app was designed to improve discreetness of completing logs while simultaneously enhancing the usability of and engagement with the app, qualitative research may be necessary to understand what maintained participants' engagement on their phones more than on the watch.

Second, the engagement trajectory visualizations illustrated that most of the engagement measures had similar overall downward trajectories. In conjunction with weekly summary data presented in Table 2, the spaghetti plots exposed the individual variability in engagement trajectories and underscored that not all components of the app were used the same over the course of a month. Although each measure of engagement used a linear functional form for subsequent multilevel models, the most variation in individual trajectories was observed in mean meal logs, which could vary tremendously on an individual's availability to complete a more time-intensive log. As an extension of this aim in future work, engagement visualizations may serve useful in characterizing subgroups of engagement profiles through a repeated measures latent profile analysis and can subsequently be used to identify unique engagement and symptom profiles associated with positive intervention outcomes (see [40] for an example in treatment-resistant depression). Subsequent studies may also seek to visualize digital intervention usage at more granular levels in terms of time and within individuals, which would be essential to tailor interventions to individual users and to provide just-in-time adaptive interventions that are responsive to engagement.

Third, the multilevel models yielded results that primarily illustrated that time was the best predictor of engagement. Time, measured in weeks, was a significant predictor in every model, and the addition of other predictors to the models typically did not improve model fit. This result was unsurprising because a consistent theme in digital intervention literature is that engagement drops as time progresses [12,41]. Beyond time, sex and ethnicity were the most common significant baseline predictors of engagement. Those who were assigned as female based on genotyping or self-identified as not Hispanic were more engaged than their male and Hispanic counterparts, respectively, for four measures of engagement: mean meal logs, mean mood logs, mean phone logs, and mean use. Although there is little to no research evaluating baseline predictors of engagement in digital ED interventions, there is some evidence to suggest sex and ethnicity are consistently significant predictors of engagement in other digital mental health interventions (e.g., [42,43]). A key point to consider is how the design of the app may have biased engagement toward demographic subgroups (e.g., the intervention in this study was not offered in Spanish nor was tailored to be culturally specific, which could have biased against

Hispanic users). Even though acceptance-facilitating interventions can increase participants' acceptance, motivations, and positive attitudes toward digital ED interventions regardless of demographic characteristics, initial work in this area highlights that engagement does not improve [44], underscoring the importance of employing user-centered design principles for improving engagement for target demographics [45].

A second set of baseline characteristics included significant predictors of only one measure of engagement each. Current BMI was a significant negative predictor of mean urge logs, and this result could be interpreted as those with higher BMIs may not have experienced as many urges or did not log their urges as often as those with lower BMIs. Similarly, the number of baseline binge episodes was a significant predictor of the mean behaviors logged, possibly reflecting that those who had greater mean behavior logs over time were more engaged in the app. However, an alternative explanation is that participants with a higher number of baseline binge episodes had more opportunities to log behaviors. To disentangle these results in future studies, it will be essential to assess and compare retrospective self-reported urges and behaviors over a given interval to the behaviors logged in the moment via digital intervention. Notably, the addition of binge episodes at baseline to the second set of engagement models did not improve fit statistics, illustrating little to no contribution to improvement in engagement prediction.

Finally, age was only a significant predictor of mean meal logs, with older individuals logging more meals than younger counterparts, which was somewhat surprising given the technological literacy of younger generations. Although age, current BMI, and binge episodes at baseline were only significant predictors in one engagement model each, these results support findings from machine learning algorithms recently identifying age and measures of ED symptom severity as significant predictors of engagement with and drop out from another digital ED intervention [46]. An important caveat to these three results is that the parameter estimates for age, current BMI, and number of binge episodes at baseline hovered around 0, warranting future research as to their clinical and practical utility.

Three additional points are worth noting. First, participants used the phone to engage with Recovery Record far more than the watch. Despite being confounded by the recruitment of prior Recovery Record users who only had access to the phone app, the percent change and more consistent engagement with the phone illustrates that participants were more inclined to use this method of delivery. The models of mean phone and watch logs also demonstrated that two baseline characteristics (sex and ethnicity) may be used to help differentiate subsequent engagement depending on the delivery platform, suggesting that design is important for delivering ED interventions through wearables and smartphones. Second, meal logs were unique compared with the other measures of engagement for a few reasons: this measure had the smallest percent reduction over four weeks, demonstrated the most variability of participant trajectories as evidenced by the spaghetti plots, and was significantly predicted by three baseline demographic predictors (age, sex, and ethnicity). Although this study is an exploratory investigation into expanding how we define engagement, this collection of findings warrants future replication and qualitative studies in other samples with ED psychopathology to evaluate if and why meal logs are consistently and proportionally more utilized than other components. Finally, an important point about mean use is that several multilevel models had the same significant baseline predictors, suggesting that overall engagement results may be reflecting groups of individuals with similar characteristics engaging with specific digital intervention functions in similar capacities. This may also mean that using overall measures of engagement obscures other results that are more sensitive or have less power. Taken together, our findings underscore that engagement with a digital ED intervention is more nuanced and complex than current research often presents when describing engagement through a single measure and is worth deeper exploration to optimize what individuals gain from using a digital ED intervention.

Limitations

Overall engagement with various functions of the app was high, particularly the number of days used. A key limitation that could partially explain this observation is that many participants who entered the study had already used the app, and they were required to complete three logs in Recovery Record prior to enrolling in the study. Therefore, the baseline level of engagement may be higher than what would be observed outside of a research study. In addition, participants' greater likelihood of engaging could be due to the convenience and accessibility of using the app on both the phone and the watch coupled with the relative ease of completing simple functions compared with other digital eating disorder interventions (i.e., logging a behavior is easier and quicker than completing a guided self-help session in an app). Future studies may consider comparing the same granularity of engagement with new and existing users of digital interventions. In addition, the measures of engagement had significant overlap in the data used to test the multilevel models, so it was unsurprising to observe patterns across the significance of baseline predictors. To address this in future work, it may be useful to conduct split-half studies where the discovery sample would identify significant predictors and these models would be tested with the replication sample. Finally, another limitation is that the four measures of engagement where sex and ethnicity were significant predictors of engagement had the largest sample size of logs, which could indicate that these models were overpowered. An alternative explanation is that enough data were acquired to consistently detect sex and ethnicity as significant predictors, and the amount of data collected in the other three models was insufficient or lacked sufficient variability to detect differences. Future studies may seek to replicate the current findings prior to evaluating thresholds to determine what baseline characteristics are clinically useful in identifying meaningful changes in engagement.

Conclusion

This study provided a novel view of engagement that characterized participants' usage of different functions, method of delivery, and overall usage of a digital ED intervention. Key predictors, time and two demographic predictors (sex, ethnicity), were consistently significant despite unique measurements of engagement across log type, device type, and overall engagement. Other baseline demographic and severity of illness predictors were significant in only one measure of engagement each, highlighting opportunities to tease out more complex and nuanced understanding of the utility of and engagement with different functions of a digital ED app. Future work may consider identifying unique engagement profiles that, in combination with baseline characteristics, can be used to predict intervention outcomes, thereby allowing researchers and clinicians to intervene earlier on engagement and harmful eating behaviors.

Considering the importance of consistent engagement in traditional psychotherapy, which is required for meaningful symptom change, this study sets the stage for understanding what types of engagement with digital ED interventions may be most helpful and can reliably predict change in symptoms.

Abbreviations

AIC - Akaike Information Criterion

AN - anorexia nervosa

BIC - Bayesian Information Criterion

BMI - body mass index

BN - bulimia nervosa

BED - binge-eating disorder

BEGIN - Binge Eating Genetics Initiative

ED - eating disorder

EDE-Q - Eating Disorder Examination Questionnaire v6

ICC - intraclass correlation

Conflicts of Interest:

C.M. Bulik reports: Pearson (author, royalty recipient). J. Tregarthen reports: Recovery Record (shareholder, employee).

Acknowledgements

We are deeply grateful to all the individuals who, through their participation, made BEGIN a success.

References

1. Kazdin AE, Fitzsimmons-Craft EE, Wilfley DE. Addressing critical gaps in the treatment of eating disorders. *Int J Eat Disord*. Mar 2017;50(3):170-189. doi:10.1002/eat.22670
2. Hart LM, Granillo MT, Jorm AF, Paxton SJ. Unmet need for treatment in the eating disorders: a systematic review of eating disorder specific treatment seeking among community cases. *Clin Psychol Rev*. Jul 2011;31(5):727-35. doi:10.1016/j.cpr.2011.03.004
3. Tregarthen JP, Lock J, Darcy AM. Development of a smartphone application for eating disorder self-monitoring. *Int J Eat Disord*. Nov 2015;48(7):972-82. doi:10.1002/eat.22386
4. Kass AE, Balantekin KN, Fitzsimmons-Craft EE, Jacobi C, Wilfley DE, Taylor CB. The economic case for digital interventions for eating disorders among United States college students. *Int J Eat Disord*. Mar 2017;50(3):250-258. doi:10.1002/eat.22680
5. Smith KE, Juarascio A. From Ecological Momentary Assessment (EMA) to Ecological Momentary Intervention (EMI): Past and Future Directions for Ambulatory Assessment and Interventions in Eating Disorders. *Curr Psychiatry Rep*. Jun 4 2019;21(7):53. doi:10.1007/s11920-019-1046-8
6. Linardon J, Shatte A, Messer M, Firth J, Fuller-Tyszkiewicz M. E-mental health interventions for the treatment and prevention of eating disorders: An updated systematic review and meta-analysis. *J Consult Clin Psychol*. Nov 2020;88(11):994-1007. doi:10.1037/ccp0000575
7. Taylor CB, Graham AK, Flatt RE, Waldherr K, Fitzsimmons-Craft EE. Current state of scientific evidence on Internet-based interventions for the treatment of depression,

anxiety, eating disorders and substance abuse: an overview of systematic reviews and meta-analyses. *Eur J Public Health*. Jul 7 2021;31(31 Suppl 1):i3-i10. doi:10.1093/eurpub/ckz208

8. Borghouts J, Eikey E, Mark G, et al. Barriers to and Facilitators of User Engagement With Digital Mental Health Interventions: Systematic Review. *J Med Internet Res*. Mar 24 2021;23(3):e24387. doi:10.2196/24387

9. Kaveladze BT, Wasil AR, Bunyi JB, Ramirez V, Schueller SM. User Experience, Engagement, and Popularity in Mental Health Apps: Secondary Analysis of App Analytics and Expert App Reviews. *JMIR Hum Factors*. Jan 31 2022;9(1):e30766. doi:10.2196/30766

10. Huckvale K, Nicholas J, Torous J, Larsen ME. Smartphone apps for the treatment of mental health conditions: status and considerations. *Current Opinion in Psychology*. 2020/12/01/ 2020;36:65-70. doi:https://doi.org/10.1016/j.copsyc.2020.04.008

11. Palmer KM, Burrows V. Ethical and Safety Concerns Regarding the Use of Mental Health-Related Apps in Counseling: Considerations for Counselors. *J Technol Behav Sci*. 2021;6(1):137-150. doi:10.1007/s41347-020-00160-9

12. Baumel A, Muench F, Edan S, Kane JM. Objective User Engagement With Mental Health Apps: Systematic Search and Panel-Based Usage Analysis. *J Med Internet Res*. Sep 25 2019;21(9):e14567. doi:10.2196/14567

13. Yeager CM, Benight CC. If we build it, will they come? Issues of engagement with digital health interventions for trauma recovery. *Mhealth*. 2018;4:37. doi:10.21037/mhealth.2018.08.04

14. 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*. Aug 2018;21(3):116-119. doi:10.1136/eb-2018-102891

15. 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*. Jun 2017;7(2):254-267. doi:10.1007/s13142-016-0453-1

16. Ng MM, Firth J, Minen M, Torous J. User Engagement in Mental Health Apps: A Review of Measurement, Reporting, and Validity. *Psychiatr Serv*. Jul 1 2019;70(7):538-544. doi:10.1176/appi.ps.201800519

17. Venkatesh A, Chang A, Green EA, et al. Perceived Facilitators and Barriers to Engaging with a Digital Intervention among Those with Food Insecurity, Binge Eating, and Obesity. *Nutrients*. Jul 19 2021;13(7)doi:10.3390/nu13072458

18. Nitsch M, Dimopoulos CN, Flaschberger E, et al. A Guided Online and Mobile Self-Help Program for Individuals With EatingDisorders: An Iterative Engagement and Usability Study. *J Med Internet Res*. 2016/01/11 2016;18(1):e7. doi:10.2196/jmir.4972

19. Anastasiadou D, Folkvord F, Serrano-Troncoso E, Lupiañez-Villanueva F. Mobile Health Adoption in Mental Health: User Experience of a Mobile Health App for Patients With an Eating Disorder. *JMIR Mhealth Uhealth*. May 31 2019;7(6):e12920. doi:10.2196/12920

20. Watson HJ, Levine MD, Zerwas SC, et al. Predictors of dropout in face-to-face and internet-based cognitive-behavioral therapy for bulimia nervosa in a randomized controlled trial. *Int J Eat Disord*. May 2017;50(5):569-577. doi:10.1002/eat.22644

21. Fitzsimmons-Craft EE, Taylor CB, Graham AK, et al. Effectiveness of a Digital Cognitive

Behavior Therapy-Guided Self-Help Intervention for Eating Disorders in College Women: A Cluster Randomized Clinical Trial. *JAMA Netw Open*. Aug 3 2020;3(8):e2015633. doi:10.1001/jamanetworkopen.2020.15633

22. Kim JP, Sadeh-Sharvit S, Welch HA, Neri E, Tregarthen J, Lock J. Eating disorders early app use mediates treatment effect on clinical improvement. *Int J Eat Disord*. Mar 2022;55(3):382-387. doi:10.1002/eat.23652

23. Linardon J, King T, Shatte A, Fuller-Tyszkiewicz M. Usability Evaluation of a Cognitive-Behavioral App-Based Intervention for Binge Eating and Related Psychopathology: A Qualitative Study. *Behav Modif*. Jun 2 2021:1454455211021764. doi:10.1177/01454455211021764

24. Linardon J, Messer M, Shatte A, et al. Does the Method of Content Delivery Matter? Randomized Controlled Comparison of an Internet-Based Intervention for Eating Disorder Symptoms With and Without Interactive Functionality. *Behavior Therapy*. 2022/05/01/ 2022;53(3):508-520. doi:https://doi.org/10.1016/j.beth.2021.12.001

25. Lindgreen P, Lomborg K, Clausen L. Patient Experiences Using a Self-Monitoring App in Eating Disorder Treatment: Qualitative Study. *JMIR Mhealth Uhealth*. Jun 22 2018;6(6):e10253. doi:10.2196/10253

26. Tregarthen J, Paik Kim J, Sadeh-Sharvit S, Neri E, Welch H, Lock J. Comparing a Tailored Self-Help Mobile App With a Standard Self-Monitoring App for the Treatment of Eating Disorder Symptoms: Randomized Controlled Trial. *JMIR Ment Health*. Nov 21 2019;6(11):e14972. doi:10.2196/14972

27. Fairburn CG. *Cognitive Behavior Therapy and Eating Disorders*. Guilford Press; 2008.

28. Bulik CM, Butner JE, Tregarthen J, et al. The Binge Eating Genetics Initiative (BEGIN): study protocol. *BMC Psychiatry*. 2020/06/16 2020;20(1):307. doi:10.1186/s12888-020-02698-7

29. Watson HJ, Yilmaz Z, Thornton LM, et al. Genome-wide association study identifies eight risk loci and implicates metabo-psychiatric origins for anorexia nervosa. *Nat Genet*. Aug 2019;51(8):1207-1214. doi:10.1038/s41588-019-0439-2

30. Thornton LM, Munn-Chernoff MA, Baker JH, et al. The Anorexia Nervosa Genetics Initiative (ANGI): Overview and methods. *Contemp Clin Trials*. Nov 2018;74:61-69. doi:10.1016/j.cct.2018.09.015

31. Igudesman D, Abbaspour A, Reed K, et al. Laxative abuse is associated with a depleted gut microbial community structure among females and males with binge-eating disorder or bulimia nervosa: The Binge Eating Genetics Initiative (BEGIN). *Psychosom Med*. Jun 19 2023;doi:10.1097/psy.0000000000001226

32. Mond JM, Hay PJ, Rodgers B, Owen C, Beumont PJV. Validity of the Eating Disorder Examination Questionnaire (EDE-Q) in screening for eating disorders in community samples. *Behaviour Research and Therapy*. 2004/05/01/ 2004;42(5):551-567. doi:https://doi.org/10.1016/S0005-7967(03)00161-X

33. Carey M, Kupeli N, Knight R, Troop NA, Jenkinson PM, Preston C. Eating Disorder Examination Questionnaire (EDE-Q): Norms and psychometric properties in U.K. females and males. *Psychol Assess*. Jul 2019;31(7):839-850. doi:10.1037/pas0000703

34. Allen KL, Byrne SM, Lampard A, Watson H, Fursland A. Confirmatory factor analysis of the Eating Disorder Examination-Questionnaire (EDE-Q). *Eating Behaviors*. 2011/04/01/

2011;12(2):143-151. doi:https://doi.org/10.1016/j.eatbeh.2011.01.005

35. *SAS/STAT® Software: Version 9.4*. SAS Institute, Inc; 2014.

36. Luke SG. Evaluating significance in linear mixed-effects models in R. *Behav Res Methods*. Aug 2017;49(4):1494-1502. doi:10.3758/s13428-016-0809-y

37. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. *Gen Hosp Psychiatry*. Jul-Aug 2013;35(4):332-8. doi:10.1016/j.genhosppsych.2013.03.008

38. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital Mental Health Interventions for Depression, Anxiety, and Enhancement of Psychological Well-Being Among College Students: Systematic Review. *J Med Internet Res*. Jul 22 2019;21(7):e12869. doi:10.2196/12869

39. Flatt RE, Thornton LM, Smith T, et al. Retention, engagement, and binge-eating outcomes: Evaluating feasibility of the Binge-Eating Genetics Initiative study. *Int J Eat Disord*. Aug 2022;55(8):1031-1041. doi:10.1002/eat.23726

40. Peiper NC, Nelson BW, Aschbacher K, Forman-Hoffman VL. Trajectories of depression symptoms in a therapist-supported digital mental health intervention: a repeated measures latent profile analysis. *Soc Psychiatry Psychiatr Epidemiol*. Aug 2023;58(8):1237-1246. doi:10.1007/s00127-022-02402-y

41. Carlo AD, Hosseini Ghomi R, Renn BN, Areán PA. By the numbers: ratings and utilization of behavioral health mobile applications. *NPJ Digit Med*. 2019;2:54. doi:10.1038/s41746-019-0129-6

42. Marinova N, Rogers T, MacBeth A. Predictors of adolescent engagement and outcomes - A cross-sectional study using the togetherall (formerly Big White Wall) digital mental health platform. *J Affect Disord*. Aug 15 2022;311:284-293. doi:10.1016/j.jad.2022.05.058

43. Arnold C, Farhall J, Villagonzalo KA, Sharma K, Thomas N. Engagement with online psychosocial interventions for psychosis: A review and synthesis of relevant factors. *Internet Interv*. Sep 2021;25:100411. doi:10.1016/j.invent.2021.100411

44. Linardon J, Anderson C, Chapneviss T, Hants E, Shatte A, Fuller-Tyszkiewicz M. Effects of an Acceptance-Facilitating Intervention on Acceptance and Usage of Digital Interventions for Binge Eating. *Psychiatr Serv*. Oct 1 2022;73(10):1173-1176. doi:10.1176/appi.ps.202100616

45. Graham AK, Kosmas JA, Massion TA. Designing Digital Interventions for Eating Disorders. *Curr Psychiatry Rep*. Apr 2023;25(4):125-138. doi:10.1007/s11920-023-01415-x

46. Linardon J, Fuller-Tyszkiewicz M, Shatte A, Greenwood CJ. An exploratory application of machine learning methods to optimize prediction of responsiveness to digital interventions for eating disorder symptoms. *Int J Eat Disord*. Jun 2022;55(6):845-850. doi:10.1002/eat.23733

Supplementary Files

Figures

Spaghetti plots of each engagement measure across four weeks.