Received: 3 November 2021 | Revised: 14 April 2022 | Accepted: 15 April 2022

DOI: 10.1002/eat.23726

ORIGINAL ARTICLE

International Journal of EATING DISORDERS WILEY

Retention, engagement, and binge-eating outcomes: Evaluating feasibility of the Binge-Eating Genetics Initiative study

Rachael E. Flatt MA[1,2] | Laura M. Thornton PhD[1] | Tosha Smith PhD[1] | Hannah Mitchell MA[3] | Stuart Argue BSc[4] | Brian R. W. Baucom PhD[5] | Pascal R. Deboeck PhD[5] | Colin Adamo MS[5] | Robyn E. Kilshaw MS[5] | Qinxin Shi PhD[5] | Jenna Tregarthen BA, PGDip[4] | Jonathan E. Butner PhD[5] | Cynthia M. Bulik PhD[1,6,7]

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

[2]Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[3]Department of Psychology, East Tennessee State University, Johnson City, Tennessee, USA

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

[5]Department of Psychology, University of Utah, Salt Lake City, Utah, USA

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

[7]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, USA.
Email: cynthia_bulik@med.unc.edu

Funding information
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; R01MH124871, Sullivan/Bulik, MPIs); uBiome (services grant, Bulik, PI); National. Science Foundation (DGE-1650116; Flatt, PI). 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 Watches 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

Abstract

Objective: Using preliminary data from the Binge-Eating Genetics Initiative (BEGIN), we evaluated the feasibility of delivering an eating disorder digital app, *Recovery Record*, through smartphone and wearable technology for individuals with binge-type eating disorders.

Methods: Participants ($n = 170$; 96% female) between 18 and 45 years old with lived experience of binge-eating disorder or bulimia nervosa and current binge-eating episodes were recruited through the *Recovery Record* app. They were randomized into a Watch (first-generation Apple Watch + iPhone) or iPhone group; they engaged with the app over 30 days and completed baseline and endpoint surveys. Retention, engagement, and associations between severity of illness and engagement were evaluated.

Results: Significantly more participants in the Watch group completed the study ($p = .045$); this group had greater engagement than the iPhone group (p's < .05; pseudo-$R^2_{McFadden}$ effect size = .01–.34). Overall, binge-eating episodes, reported for the previous 28 days, were significantly reduced from baseline (mean = 12.3) to endpoint (mean = 6.4): most participants in the Watch (60%) and iPhone (66%) groups reported reduced binge-eating episodes from baseline to endpoint. There were no

Rachael E. Flatt and Laura M. Thornton are co-first authors.

collection, analysis, interpretation of data, or writing the manuscript.

Action Editor: Tracey Wade

significant group differences across measures of binge eating. In the Watch group, participants with fewer episodes of binge eating at baseline were more engaged (p's < .05; pseudo-$R^2_{McFadden}$ = .01–.02). Engagement did not significantly predict binge eating at endpoint nor change in binge-eating episodes from baseline to endpoint for both the Watch and iPhone groups.

Discussion: Using wearable technology alongside iPhones to deliver an eating disorder app may improve study completion and app engagement compared with using iPhones alone.

KEYWORDS

eating disorders, genetics, microbiome, mobile application, wearable technology

1 | INTRODUCTION

Binge-eating disorder (BED) and bulimia nervosa (BN) are serious, often persist for years (mean duration of illness = 8–14 years [Hudson et al., 2007]), carry high psychiatric and somatic comorbidity (Fichter & Quadflieg, 2016; Thornton et al., 2017; Welch et al., 2016), elevated suicide risk (Crow, 2014; Forrest et al., 2016; Huas et al., 2013; Pisetsky et al., 2013) and significant impairment (Kessler et al., 2013). However, widely accessible, effective treatment for BED and BN is lacking. To increase reach of evidence-based eating disorder (ED) treatments, we tested the feasibility of digital interventions using smartphones and wearable technology.

Although cognitive-behavioral therapy (CBT) is the top evidence-based treatment for BED and BN (Brownley et al., 2016; Hay et al., 2004; Shapiro et al., 2007), problems with treatment delivery models challenge treatment potency (Kazdin et al., 2017). First, the "dominant model of treatment delivery," CBT, is confined to the walls of the clinic, delivered by a trained mental health professional, and conducted face-to-face (Kazdin et al., 2017). The *reach* of this model is poor. Second, this model of treatment delivery is underutilized (Kazdin et al., 2017), with community studies in the United States indicating that only ~43.5% of individuals with a lifetime history of BED or BN have ever sought ED treatment (Hudson et al., 2007), and poor engagement and dropout remain problematic (Beintner et al., 2014; Fassino et al., 2009; Waller, 1997). The delivery of face-to-face care has been further limited by the COVID-19 pandemic (Termorshuizen et al., 2020), highlighting the need for telehealth options. The United States also has a dire shortage of mental health providers (Health Resources & Services Administration, 2021; National Council for Mental Wellbeing, 2017; Thomas et al., 2009) with competence in CBT for EDs (Agras et al., 2017; Mussell et al., 2000), further impeding *delivery* of evidence-based treatment. Third, interventions are typically confined to ≤50 min/week, without *real-time support* in patients' daily lives (Tregarthen et al., 2015), and traditional CBT approaches rely on *retrospective* self-monitoring of "triggers" for ED behaviors (i.e., binge eating, purging), a central component of treatment (Barakat et al., 2017; Latner et al., 2002). ED symptoms and meals are frequently recalled retrospectively in therapy sessions and are subject to memory decay, social desirability, recall bias, and mood (Schoch & Raynor, 2012; Smyth et al., 2001).

A clear need exists for accessible, scalable CBT-based interventions for the millions of individuals not otherwise seeking treatment, but who may engage in a program that is discreet, affordable, and accessible. Online ED interventions offer a potential solution, given that almost all Americans have access to a cellphone and/or computer (Pew Research Center, 2021), and online and mobile application-based treatment programs for EDs have demonstrated initial efficacy (Aardoom et al., 2016; Bauer & Moessner, 2013). Support tools such as *Recovery Record* (RR) address the aforementioned problems by increasing reach, providing 24/7 support for users, and encouraging in-the-moment monitoring and interaction (Chapa et al., 2020; Tregarthen et al., 2019). However, most technology-based tools and ecological momentary assessment approaches rely on burdensome self-report data, akin to in-person treatment.

The Binge-Eating Genetics Initiative (BEGIN) fills this gap by adapting the 4-week CBT-based content for binge-type EDs delivered via a smartphone app (*RR*) (Tregarthen et al., 2019) for use on Apple Watches. Data collected will ultimately be used to design interventions to personalize CBT-based just-in-time adaptive interventions (JITAIs) to prevent unhealthy behaviors *before* they occur (i.e., an early warning system), thereby increasing the reach of evidence-based recovery tools and the scalability and utility of digital technology. This study, representing the original feasibility investigation, compares retention and engagement for participants using a first-generation Watch alongside an iPhone with participants only using an iPhone app. To date, only one study specifically evaluated *RR* engagement in 3294 users (Kim et al., 2022). This study, which did not include wearables, indicated greater engagement and more actual time spent on the app were associated with improved outcomes. This study serves as a foundation for future work with JITAIs.

We present findings from the feasibility phase of BEGIN focusing on associations between wearable technology, engagement in the *RR* platform, and binge eating. We assessed whether (1) wearable technology (i.e., a first-generation Watch) improved study retention and completion and enhanced engagement in the *RR* platform compared to using the app exclusively in its traditional smartphone platform

(i.e., iPhone group); (2) wearable technology was associated with change in binge-eating episode frequency from baseline to endpoint; (3) associations existed between severity of illness and *RR* engagement and whether these associations differed between those accessing *RR* on the Watch and those in the iPhone group. Specifically, we assessed whether the number of binge-eating episodes in the 28 days prior to enrollment, number of binge-eating episodes during the study, and change in binge-eating episodes from baseline to endpoint were associated with engagement.

2 | METHOD

2.1 | Participants

Participants met the following enrollment criteria: (1) have lived experience of BED or BN; (2) currently experiencing binge-eating episodes; (3) 18–45 years old, (4) be ambulatory, (5) be a U.S. resident, (6) speak English, and (7) own an iPhone 5 or newer. Exclusion criteria included: (1) current hormone therapy, (2) inpatient treatment or hospitalization for EDs in the 2-weeks prior to enrollment, (3) active suicidality at screening, (4) history of bariatric surgery, (5) pregnant or breastfeeding, or (6) antibiotic or probiotic use in the past 30 days (for fecal sampling in the parent study). There were no exclusions for other types of past or current ED, nutritional, or weight loss treatment.

2.2 | Procedure

The complete study protocol is published elsewhere (Bulik et al., 2020); individuals recruited for the feasibility study are included in the parent study. Participants in the feasibility study were only recruited through the *RR* mobile app downloaded via the Apple AppStore; recruitment was not limited to new users. *RR* sent an invitation to the ED100K-v2 eligibility questionnaire (Bulik et al., 2020; Thornton et al., 2018) to all app users who logged three meals. Eligible individuals were invited to complete consent forms. For the feasibility study, participants were randomized into the Watch group (i.e., completed the study using the *RR* application adapted for an Apple Watch *in conjunction* with their iPhone), or the iPhone group (i.e., who only used the iPhone version of the *RR* application). Participants received kits containing a first-generation Watch (if in the Watch group), saliva collection kit, fecal sampling kit, and instructions for study procedures. All participants were instructed to log behaviors, urges, mood, and meals each day for 30 days; however, instructions regarding frequency of app usage beyond daily logs were not specified. Watch group participants were asked to wear the watch during waking hours and to log behaviors, urges, and mood on the Watch; meals had to be logged on the iPhone. Finally, participants completed the EDE-Q at endpoint (at 30 days); the EDE-Q was not administered at midpoint in the feasibility study. Participants could keep the Watch and continue using *RR* after completing the study, but those data were not collected.

2.3 | Recovery record and app engagement

RR delivers an innovative 4-week CBT-based program (including self-monitoring) for binge-type EDs (Tregarthen et al., 2015; Tregarthen et al., 2019). Users log daily meals, mood, urges, and disordered eating behaviors (binge eating, vomiting, laxative/diuretic misuse, excessive exercise). *RR* provides CBT-based content including goal setting, cognitive restructuring, emotion regulation, behavioral techniques, and coping strategies (e.g., mindfulness, positive activity scheduling, seeking social support; see Figure 1).

RR is typically used as a mobile application with a smartphone. Participants used the traditional app through their mobile device, and the app was modified to deliver content through a Watch. Participants could log behaviors, urges, and mood and engage with coping strategies through the Watch and/or iPhone, but could only log meals through the iPhone. Prior studies using ecological momentary assessment that explored antecedents to and consequences of binge eating (Berg et al., 2013; Berg et al., 2017; Goldschmidt et al., 2014; Goldschmidt et al., 2018; Haedt-Matt & Keel, 2011; Hilbert & Tuschen-Caffier, 2007; Smyth et al., 2007) informed our adaptation of *RR* with reference to mood monitoring: participants could select emojis that reflect negative and positive affect flanking binge or purge episodes or when desired. If participants did not engage with *RR* for 23 days, they were sent email reminders ($n = 1$) and notifications to continue using the app. Participants who experienced technical difficulties related to the app or Watch ($n = 4$) contacted the research team, who troubleshot case-by-case concerns.

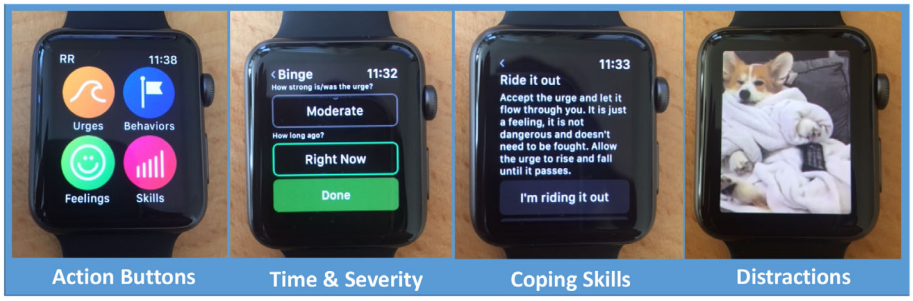

FIGURE 1 Recovery record for apple watch screen examples

For this study, we evaluated how often participants engaged with the app. For the Watch group, we counted *all* times the participant interacted with the app, whether on the iPhone or the Watch. For the iPhone group, this was the number of times the participant interacted with the app on the iPhone. We defined "engagement" four ways: (1) "$\text{DailyMean}_{\text{Meals}}$" is the *daily mean* of the number of times participants opened the app per day over the month, including the times they opened the app specifically to log *meals*; and (2) "$\text{DaysUsedRR}_{\text{Meals}}$" is the number of *days* participants *used* the *RR* app during the month, including logging *meals*. Because meal logs could only be entered on the phone, we also assessed (3) "$\text{DailyMean}_{\text{NoMeals}}$," defined as the *daily mean* number of times participants opened the app per day over the month, *not* including *meal* logs; and (4) "$\text{DaysUsedRR}_{\text{NoMeals}}$," which is the number of *days* participants *used* the *RR* app during the month, *not* including *meal* logs.

The research was reviewed and approved by an institutional review board.

Trial registration: The ClinicalTrials.gov identifier is NCT04162574.

2.4 | Measures

2.4.1 | Sample characteristics

Demographic data (age, sex, race, ethnicity) were collected at baseline. Treatment history was assessed with three questions: (1) *Have you ever received any of the following hospital-based treatments for binge eating* with response options: inpatient, residential, emergency room, I have never received any hospital-based treatment, or do not know/refuse. (2) *Have you ever received any of the following outpatient psychotherapy treatments for binge eating?* Outpatient treatment was scored as 1 if the participant endorsed any psychotherapy and scored as 0 if they endorsed not receiving any outpatient treatment or did not endorse either. (3) *Have you ever taken any of the following medications for binge eating?* If participants selected any listed medications, they were scored as 1 for medications and 0 otherwise.

The *ED100K-v2* is a self-report assessment based on the Structured Clinical Interview for DSM-5, Eating Disorders Module, administered prior to enrollment. Items assess DSM-5 criteria for BED, BN, and anorexia nervosa; algorithms determined lifetime diagnosis of all three EDs. Participant self-reported height and weight was used to calculate current, lifetime highest, and adult lifetime lowest BMI. ED100K-v1 is a valid measure of eating symptoms and behaviors (Thornton et al., 2018).

Participants completed the following questionnaires at baseline. If participants did not complete these questionnaires within 4 days, they received reminder emails.

The *Eating Disorders Examination Questionnaire-V6.0 (EDE-Q)* (Fairburn & Beglin, 1994) captures EDs pathology, including the frequency and severity of binge episodes over the past 28 days, and demonstrates moderate-to-good validity and reliability (Berg et al., 2012). The EDE-Q was administered at enrollment and endpoint 30 days later. The number of binge-eating episodes was evaluated by a single item: On how many of these times (where you ate what other people would regard as an unusually large amount of food) did you have a sense of having lost control over your eating (at the time you were eating)? Discrete change in binge-eating episodes was calculated as (endpoint episodes – baseline episodes).

"Retention" in the study was defined as completing the EDE-Q at baseline and endpoint. Study "completion" was defined as completing the EDE-Q at baseline and endpoint and returning saliva (DNA) and fecal samples. Analyses of saliva and fecal samples are the focus of future reports.

The *Patient Health Questionnaire* (*PHQ-9*) (Kroenke et al., 2001) is a 9-item, self-administered version of the PRIME-MD with good reliability and validity (Kroenke et al., 2001). Items are based on DSM-IV criteria for major depressive disorder and are scored as "0" (not at all) to "3" (nearly every day). Sum scores indicate severity: 5–9 = mild, 10–14 = moderate, 15+ = severe symptoms.

The *Generalized Anxiety Disorder 7* (*GAD-7*) (Lowe et al., 2008) is a 7-item, self-report questionnaire for generalized anxiety disorder with good reliability (Spitzer et al., 2006) and validity (Lowe et al., 2008) in the general population. Each symptom is scored on a 3-point scale: "not at all" (0), "several days" (1), or "more than half the days" (2). Items are summed to create a "GAD-7 total score." Scores indicate severity: 5–9 = mild, 10–14 = moderate, 15+ = severe symptoms.

2.5 | Data analysis

Statistical analyses were conducted using SAS® version 9.4 (SAS Institute Inc., 2013). Prior to analyses, descriptive statistics and graphics were used to screen data for implausible values, errors, and check distributional assumptions. To evaluate missing data, participants who met retention criteria were compared to those who did not across demographic variables using χ^2 tests for independence and independent samples *t*-tests. Retention was defined by a single item, so missing data were not imputed.

Descriptive statistics were generated for sex, ethnicity, race, lifetime ED diagnosis (i.e., BED, BN, or both determined by algorithm from the ED100K-v2), and treatment variables (*n* and %), and for age, BMI variables, PHQ-9, and GAD-7 scores (mean and std) by group. Groups were compared on these variables using χ^2 tests for independence and independent samples *t*-tests.

For Aim 1, χ^2 tests were applied to evaluate whether the Watch and iPhone groups differed in retention and completion. To evaluate engagement, only participants who met retention criteria (i.e., completed the endpoint EDE-Q) were included because we cannot accurately determine when a participant decided to end participation based solely on app usage. To evaluate differences in engagement, analysis of variance models were applied using PROC GENMOD, predicting engagement from study group. Current BMI was entered as a covariate. For these analyses and the generalized linear models described below, distribution and link functions were set to normal and identity for $\text{DailyMean}_{\text{Meals}}$, $\text{DailyMean}_{\text{NoMeals}}$, and

discrete change in binge-eating episodes, and to negative binomial (overdispersion was observed) and log for $DaysUsedRR_{Meals}$, $DaysUsedRR_{NoMeals}$, and number of binge-eating episodes.

For Aim 2, we evaluated differences between groups on two binge eating measurements: (1) number of binge-eating episodes at endpoint, and (2) discrete change in binge-eating episodes from baseline to endpoint (endpoint episodes—baseline episodes) using analysis of variance models, applied using PROC GENMOD. Current BMI and baseline binge-eating episodes were entered as covariates; effect sizes were calculated using McFadden's pseudo-R^2; [footnote 1]. The number of participants in each group who entered or sustained remission, defined as zero episodes of binge eating over 30 days, is also reported.

For Aim 3, linear regression models (using PROC GENMOD) were used to evaluate associations between severity of illness and *RR* engagement. Specifically, we predicted engagement from the number of binge-eating episodes reported at baseline; current BMI was entered into models as a covariate. This indicates whether baseline severity of illness influences engagement. We then predicted number of binge-eating episodes at endpoint from engagement, accounting for number of baseline episodes and current BMI. This indicates how engagement influences severity at end of study. Lastly, to determine if engagement was associated with a change in severity, discrete change in binge-eating episodes from baseline to endpoint was predicted from engagement, accounting for current BMI and baseline binge episodes. Group (iPhone vs Watch) was entered into all models as a covariate. If group was significant, the model was then stratified by group. Results from the initial model, stratified models, and McFadden's pseudo-R^2 effect sizes are presented.

3 | RESULTS

3.1 | Sample characteristics

Across the United States, 170 participants (95.9% female, 87.1% non-Hispanic, 89.7% White) met inclusion criteria. Most participants (70.6%) reported a lifetime history of both BED and BN, whereas 11.8% reported a lifetime diagnosis of BED but not BN, and 17.6% reported a history of BN but not BED. Almost 30% of participants had never been treated for their ED. For the missing data analysis, those who met retention criteria ($n = 107$) were more likely to be older ($p = .018$) and female ($p = .01$) compared to those who did not meet retention criteria ($n = 63$).

Participants were randomly assigned to the Watch group ($n = 86$) or the iPhone group ($n = 84$). These sample sizes provide post-hoc power ($\beta = .80$, $\alpha = .05$) to detect moderate effect sizes, Cohen's $D = .40$ and odds ratios ≥ 2.4.

Table 1 presents descriptive statistics and questionnaire responses by group. Significant differences in ED diagnosis were observed: the Watch group had more participants with BN only and fewer with both BN and BED. On average, participants reported being mild-to-moderately depressed and anxious. In the 28 days prior to enrollment, the number of binge-eating episodes ranged from 0 to 56 and the average number of episodes was 11.7. Current BMI was significantly higher in the iPhone group than in the Watch group. No other significant differences were noted. Participants in the Watch group wore Watches for an average of 8 hours/day. No adverse events were reported.

3.2 | Aim 1: Retention, completion, and engagement

Although more participants in the Watch group (66.3%) met criteria for retention (completing the EDE-Q at baseline and endpoint) than those in the iPhone (59.5%) group, the difference was not significant [$\chi^2(1,n = 170) = 0.83$; $p = .37$]. Significantly more participants in the Watch group (54.6%) met criteria for study completion (completing the EDE-Q at baseline and endpoint and submitting saliva and microbiome samples) than in the iPhone group [39.3%; $\chi^2(1,n = 170) = 4.03$; $p = .045$].

To be included in engagement analyses, participants had to meet retention criteria: 57 participants in the Watch group and 50 participants in the iPhone group met retention criteria. However, one participant in the Watch group and two in the iPhone group did not have any logs; thus, 56 in the Watch group and 48 in the iPhone group were included in the engagement analyses. Notably, of the total observations for participants in the Watch group that were not meal logs, 15.7% of observations (e.g., behaviors, urges) were logged on phones.

Table 2 presents mean engagement by group and results. The groups differed significantly in the mean number of times the app was opened daily *including* meal logs ($DailyMean_{Meals}$) and *excluding* meal logs ($DailyMean_{NoMeals}$), and the number of days the app was used over 30 days *excluding* meal logs ($DaysUsedRR_{NoMeals}$). In all cases, the Watch group had greater engagement. No differences were found between groups for the number of days used in 30 days *including* meal logs ($DaysUsedRR_{Meals}$).

3.3 | Aim 2: Binge eating

To assess group differences in binge eating at endpoint and change in binge eating across the study, only participants with data from the EDE-Q at baseline and endpoint were included (Watch group = 57; iPhone group = 50). In this reduced sample, the mean (SD) number of binge-eating episodes in the last 30 days recorded at baseline was 12.3 (11.6). At endpoint, the average number of episodes [mean (SD) = 6.4 (6.5)] was significantly reduced (paired *t*-test *t*-value [106] = 5.43, $p < .0001$). Most participants in each group (Watch group = 59.7%; iPhone group = 66.0%) reduced the number of binge-eating episodes from baseline to endpoint. However, no significant differences between the groups were observed in number of binge-eating

[1]McFadden's pseudo R^2 was calculated as the following: $1\text{-}LL_{mod}/LL_0$, where LL_{mod} was the log likelihood of the value for the fitted model, and LL_0 was the log likelihood of the value for the null model.

TABLE 1 Descriptive statistics for demographic, treatment, and clinically related variables measured at baseline, by group, and results evaluating differences between groups on each variable

Variable	Categories	Watch $n = 86$ % (n)	iPhone $n = 84$ % (n)	$\chi^2_{(df,n)}$ statistic (p-value)
Sex	Female	NR	NR	$3.60_{(1,170)}$ (.12)[a]
	Male	NR	NR	
Ethnicity	Hispanic	9.3 (8)	16.7 (14)	$2.05_{(1,170)}$ (.16)
	Non-Hispanic	90.7 (78)	83.3 (70)	
Race	African American	NR	NR	$1.60_{(5,170)}$ (.90)
	Asian	NR	NR	
	White	87.2 (75)	86.9 (73)	
	More than one race	4.6 (4)	7.1 (6)	
	Native American	NR	NR	
	Not Reported	NR	NR	
Diagnosis	BED	12.8 (11)	10.7 (9)	$6.18_{(2,170)}$ (.046)
	BN	24.4 (21)	10.7 (9)	
	BED and BN	62.8 (54)	78.6 (66)	
History of inpatient treatment	Yes	22.1 (19)	20.7 (17)	$0.05_{(1,168)}$ (.83)
	No	77.9 (67)	79.3 (65)	
History of outpatient treatment	Yes	67.4 (58)	75.6 (62)	$1.37_{(1,168)}$ (.25)
	No	32.6 (28)	24.4 (20)	
History of any treatment	Yes	67.4 (58)	75.6 (63)	$1.49_{(1,169)}$ (.23)
	No	32.6 (28)	24.1 (20)	
History of medication for binge eating	Yes	57.0 (49)	61.0 (50)	$0.28_{(1,168)}$ (.60)
	No	43.0 (37)	39.0 (32)	
		Mean (*SD*) [range]	Mean (*SD*) [range]	t-value$_{(df)}$ (p-value)
Age		28.0 (6.6) [18, 45]	29.5 (6.3) [20, 45]	$-1.54_{(168)}$ (.13)
Current BMI		29.8 (9.3) [17.7, 56.9]	33.6 (11.0) [16.6, 63.8]	$-2.41_{(168)}$ (.018)
Highest BMI		33.7 (9.9) [17.7, 63.5]	36.4 (11.4) [20.9, 66.7]	$-1.64_{(168)}$ (.11)
Lowest BMI		22.4 (6.0) [12.3, 48.2]	23.4 (7.5) [12.3, 54.9]	$-0.97_{(168)}$ (.34)
PHQ-9 Score ($n_{Watch} = 76$; $n_{iPhone} = 69$)[b]		10.9 (5.3) [0, 25]	10.5 (5.3) [0, 23]	$0.47_{(143)}$ (.64)
GAD-7 Score ($n_{Watch} = 76$; $n_{iPhone} = 69$)[b]		9.3 (5.1) [0, 21]	9.3 (5.5) [0, 21]	$-0.04_{(143)}$ (.98)
Binge-eating episodes at baseline ($n_{Watch} = 75$; $n_{iPhone} = 74$)[b]		12.0 (11.2) [0, 56]	11.4 (10.0) [0, 50]	$0.32_{(147)}$ (.76)

Note: NR, Not reported here to protect the privacy of participants.
[a]Probability based on Fisher's exact test.
[b]Participants were not required to complete the PHQ-9, GAD-7, or EDE-Q (where binge episodes at baseline were reported), resulting in different Ns per group.

TABLE 2 Mean (SD) engagement with the recovery record app by group; results from analysis of variance models[a] evaluating group differences in various engagement variables with current BMI as a covariate; and effect sizes

Engagement[a,b]	Watch $n = 56$ Mean (*SD*)	iPhone $n = 48$ Mean (*SD*)	χ^2 statistic (df = 1) (p-value)	McFadden's pseudo-R^2
$DailyMean_{Meals}$	13.6 (10.0)	10.2 (7.3)	3.95 (.047)	0.01
$DaysUsedRR_{Meals}$	23.0 (6.6)	23.1 (6.9)	0.02 (.899)	0.00
$DailyMean_{NoMeals}$	3.5 (3.8)	0.2 (0.3)	33.01 (<.0001)	0.06
$DaysUsedRR_{NoMeals}$	13.8 (9.2)	1.1 (1.7)	90.46 (<.0001)	0.34

[a]For $DailyMean_{Meals}$ and $DailyMean_{NoMeals}$, analysis of variance was applied with PROC GENMOD (distribution = normal, link = identity). For $DaysUsedRR_{Meals}$ and $DaysUsedRR_{NoMeals}$, analysis of variance was applied with PROC GENMOD (distribution = negative binomial, link = log).
[b]$DailyMean_{Meals}$ = the mean number of times participants opened the app per day over the month, including meal logs; $DaysUsedRR_{Meals}$ = the number of days the participant used the app during the month including meal logs; $DailyMean_{NoMeals}$ = the mean number of times participants opened the app per day over the month not counting meals logs; $DaysUsedRR_{NoMeals}$ = the number of days the participant used the app during the month, not counting meal logs.

episodes at endpoint [$\chi^2(1,n = 107) = 1.11$, $p = .30$] or in the discrete change in the number of binge-eating episodes from baseline to endpoint [$\chi^2(1,n = 107) = 1.70$, $p = .20$].

Despite endorsing current binge eating on the ED100K-v2, three participants in the Watch group and four in the iPhone group reported no episodes of binge eating at baseline and at endpoint in the EDE-Q, indicating they may have been in remission. One participant in each group was in remission at baseline (reported no episodes of binge eating on the EDE-Q at baseline) but reported binge-eating episodes at endpoint, indicating possible relapse. Three additional participants in the Watch group and seven in the iPhone reported entering remission at endpoint.

3.4 | Aim 3: Associations between severity of illness and engagement

Table 3 presents results from models evaluating associations between severity of illness and engagement. Baseline binge-eating episodes significantly predicted $DailyMean_{NoMeals}$; group was significant in this model. Stratified analyses revealed that, in the Watch group, fewer binge-eating episodes at baseline predicted greater engagement as measured by $DailyMean_{NoMeals}$. This association was not significant in the iPhone group. Although $DaysUsedRR_{NoMeals}$ was not significantly associated with baseline binge-eating episodes in the initial model, group was significant. The stratified analyses indicate that $DaysUsedRR_{NoMeals}$ was significantly associated with baseline binge-eating episodes in the Watch group but not the iPhone group.

In contrast, no engagement variables were associated with the number of binge-eating episodes at endpoint or with discrete change in number of binge-eating episodes from baseline to endpoint. The most consistently significant predictor of both endpoint episodes and change in number of episodes was number of baseline binge-eating episodes.

4 | DISCUSSION

This study explored the feasibility of adapting the widely used CBT-based ED app *RR* for use on the Apple Watch.

4.1 | Aim 1: Retention, completion, and engagement

Although there was no difference in retention, significantly more participants in the Watch group met criteria for study completion than

TABLE 3 Results from generalized linear models evaluating associations between severity of illness and engagement, accounting for current BMI and group. For models where group was significant, results from analyses stratified by group and effect sizes are presented

	Results from unstratified analyses		Results from stratified analyses					
			Watch			iPhone		
Model[a]	B (*SE*)	χ^2 statistic (df = 1) (*p*-value)	B (*SE*)	χ^2 statistic (df = 1) (*p*-value)	McFadden's pseudo-R^2	B (*SE*)	χ^2 statistic (df = 1) (*p*-value)	McFadden's pseudo-R^2
Baseline binge episodes predicting engagement								
$DailyMean_{Meals}$	−0.015 (0.075)	0.04 (.84)	0.003 (0.110)	0.00 (.98)	0.01	0.011 (0.095)	0.01 (.92)	0.01
$DaysUsedRR_{Meals}$	−0.002 (0.003)	0.46 (.50)						
$DailyMean_{NoMeals}$	−0.052 (0.023)	5.02 (.025)	−0.081 (0.040)	3.85 (.050)	0.02	−0.001 (0.004)	0.07 (.80)	0.05
$DaysUsedRR_{NoMeals}$	−0.011 (0.008)	1.79 (.19)	−0.020 (0.008)	4.98 (.026)	0.01	0.016 (0.019)	0.69 (.41)	0.01
Engagement predicting endpoint binge episodes[b]								
$DailyMean_{Meals}$	0.004 (0.011)	0.11 (.74)						
$DaysUsedRR_{Meals}$	−0.003 (0.016)	0.03 (.86)						
$DailyMean_{NoMeals}$	−0.031 (0.039)	0.60 (.45)						
$DaysUsedRR_{NoMeals}$	−0.009 (0.017)	0.31 (.59)						
Engagement predicting discrete change in binge episodes[b]								
$DailyMean_{Meals}$	0.020 (0.067)	0.09 (.77)						
$DaysUsedRR_{Meals}$	0.003 (0.089)	0.00 (.98)						
$DailyMean_{NoMeals}$	−0.152 (0.218)	0.48 (.49)						
$DaysUsedRR_{NoMeals}$	−0.040 (0.096)	0.18 (.68)						

[a]$DailyMean_{Meals}$ = the mean number of times participants opened the app per day over the month, including meal logs; $DaysUsedRR_{Meals}$ = the number of days the participant used the app during the month including meal logs; $DailyMean_{NoMeals}$ = the mean number of times participants opened the app per day over the month after meals logs were removed; $DaysUsedRR_{NoMeals}$ = the number of days the participant used the app during the month, calculated after meal logs were removed.
[b]Baseline binge-eating episodes were as a covariate in models.

those in the iPhone group, indicating that the Watch group was more involved in the study. Similarly, the Watch group had greater engagement than the iPhone group across three measures of engagement ($DailyMean_{Meals}$, $DailyMean_{NoMeals}$, $DaysUsedRR_{NoMeals}$). Using wearable technology alongside a smartphone may promote overall engagement in apps like *RR*, which should be replicated with other digital ED interventions in future research.

4.2 | Aim 2: Binge eating

Participants reporting an average of 11.7 binge episodes per month at enrollment and the majority reporting prior ED treatment or pharmacotherapy for binge eating suggests that symptoms had not been fully alleviated or they had relapsed at time of enrollment, underscoring the need for effective and accessible treatments. Mean PHQ-9 and GAD-7 total scores indicated the sample was also experiencing moderate depressive symptoms and mild–moderate anxiety symptoms, which is commonly observed in ED samples (Grilo et al., 2009; Hudson et al., 2007).

By end of study, most individuals who met retention criteria in both groups reported substantial reductions in binge-eating episodes: on average, the sample halved their binge episode frequency over 30 days. This result suggests that using *RR* as an intervention over 1 month could help individuals reduce the frequency of binge-eating episodes. No significant differences emerged between the Watch and iPhone groups in binge-eating episodes at endpoint and change in binge-eating episodes from baseline to endpoint, suggesting that both forms of the app may be suitable options for future users. Our results support the utility and acceptability of a digital ED intervention app delivered through wearable technology.

4.3 | Aim 3: Associations between severity of illness and engagement

For those who met retention criteria, severity of illness, as indicated by the frequency of binge episodes at baseline, predicted engagement outside of meal logs ($DailyMean_{NoMeals}$ and $DaysUsedRR_{NoMeals}$) in the app in the Watch but not the iPhone group: participants with less frequent binge episodes engaged with the app more by logging behaviors, urges, and mood. This observation has several possible explanations. Participants using the app on the Watch who had fewer binge episodes at baseline may have been more willing and/or able to engage with the apps other features. Those with more baseline eating pathology may have prioritized logging meals and binges over the other available features. Alternatively, those with less severe ED pathology and less impairment at baseline may have been more motivated to engage or were already further along in recovery, and this study provided an added boost.

App engagement was not predictive of endpoint binge-eating episodes or change in binge-eating episodes after accounting for baseline binge-eating episodes. Nonetheless, binge-eating episode frequencies were nearly halved by the end of the 30-day study, suggesting that using a Watch and iPhone may not be required for change in binge episodes. When combined with the high mean engagement rates, where participants used the app 76% of the days in the study and multiple times per day, these results suggest that consistent use of *RR* through either smart phone or wearable technology may lead to symptom improvement. Future research may evaluate if such gains are sustained long term.

4.4 | Strengths and limitations

Strengths of the study include employing new wearable technology that is more discrete to deliver evidence-based ED treatment for BED and BN and that can be used alongside a smartphone. Moreover, the absence of any reported adverse events speaks to the safety and acceptability of this approach. Our use of a readily downloadable app allows for immediate translation to real world scenarios.

Limitations should be noted. Although participants were recruited from the United States, more specific geographic location is not reported to protect confidentiality. Participants were not randomized by diagnosis, which resulted in small but significant differences in diagnosis frequency between Watch and iPhone groups. The sample primarily comprised White females, which, despite not representing U.S. demographics, aided in planning recruitment strategies to ensure appropriate representation in the parent study. Differences were observed in age and sex between those who met retention criteria (thus did not have missing data and were included in analyses for Aims 2 and 3) and those who did not. However, differences in retention might be explained by other variables that were not measured.

Participants in the Watch group may have felt that the Watch was a reward for participation or were excited about the novelty of the product and were subsequently more inclined to engage with the *RR* app. Recruitment was open to all *RR* users, new or ongoing, if they were actively engaged as indicated by logging three meals. It is possible that new users may have engaged more due to the novelty of the app and/or that the participant pool skewed more towards actively engaged individuals who had experience with the app. Also, reminders may have influenced participants' engagement, especially if received on both the iPhone and Watch. Future research may address these limitations and establish evidence for acceptability and satisfaction with the app. Finally, an inclusion criterion was current usage of an iPhone; research suggests these products are more widely used in higher socioeconomic samples (Bertrand & Kamenica, 2018) and may limit the generalizability of the results. Socioeconomic status was not queried—future research should consider complementary options for lower cost platforms.

5 | CONCLUSIONS AND FUTURE DIRECTIONS

Results of the feasibility study encouraged launching the parent study to gather 30-days of *RR* data and passively collected data on actigraphy and heart rate from Apple Watches. This combination

extends beyond ecological momentary assessment technology and will dramatically reduce participant burden by collecting passive and active data discretely while gathering orders of magnitude more data over a 30-day period than possible with standard approaches. These intensive data allow for complex modeling that mirrors the dynamic nature of BED/BN. Identification of low- and high-risk data patterns will facilitate prediction of transitions to high-risk states signaling impending binge or purge episodes and will provide a strong foundation for the contextual nature of BED/BN critical to implementing JITAIs.

ACKNOWLEDGMENT

We are deeply grateful to all the individuals who, through their participation, made BEGIN a success. Apple Watches were provided to the University of North Carolina at Chapel Hill by Apple Inc. at no cost.

CONFLICT OF INTEREST

Cynthia M. Bulik reports: Shire (grant recipient, Scientific Advisory Board member); Idorsia (consultant); Pearson (author, royalty recipient); Equip Health Inc. (Clinical Advisory Board). Jenna Tregarthen reports: Recovery Record (shareholder, employee).

AUTHOR CONTRIBUTIONS

Rachael Elizabeth Flatt: Data curation; formal analysis; investigation; methodology; project administration; writing – original draft; writing – review and editing. **Laura M. Thornton:** Conceptualization; data curation; formal analysis; funding acquisition; investigation; methodology; supervision; writing – original draft; writing – review and editing. **Tosha Woods Smith:** Conceptualization; data curation; methodology; project administration; writing – review and editing. **Hannah Mitchell:** Data curation; project administration; writing – review and editing. **Stuart Argue:** Project administration; resources; software; writing – review and editing. **Brian RW Baucom:** Funding acquisition; methodology; project administration; writing – review and editing. **Pascal Deboeck:** Funding acquisition; project administration; writing – review and editing. **Colin Adamo:** Project administration; writing – review and editing. **Robyn Kilshaw:** Project administration; writing – review and editing. **Qinxin Shi:** Project administration; writing – review and editing. **Jenna P Tregarthen:** Conceptualization; funding acquisition; project administration; software; writing – review and editing. **Jonathan Butner:** Funding acquisition; investigation; methodology; project administration; supervision; writing – review and editing. **Cynthia M. Bulik:** Funding acquisition; conceptualization; investigation; methodology; project administration; supervision; writing – review and editing.

DATA AVAILABILITY STATEMENT

Data are available upon request.

ORCID

Rachael E. Flatt https://orcid.org/0000-0002-2661-7987
Cynthia M. Bulik https://orcid.org/0000-0001-7772-3264

REFERENCES

Aardoom, J. J., Dingemans, A. E., & Van Furth, E. F. (2016). E-health interventions for eating disorders: Emerging findings, issues, and opportunities. *Current Psychiatry Reports*, *18*(4), 42. https://doi.org/10.1007/s11920-016-0673-6

Agras, W. S., Fitzsimmons-Craft, E. E., & Wilfley, D. E. (2017). Evolution of cognitive-behavioral therapy for eating disorders. *Behaviour Research and Therapy*, *88*, 26–36. https://doi.org/10.1016/j.brat.2016.09.004

Barakat, S., Maguire, S., Surgenor, L., Donnelly, B., Miceska, B., Fromholtz, K., Russell, J., Hay, P., & Touyz, S. (2017). The role of regular eating and self-monitoring in the treatment of bulimia nervosa: A pilot study of an online guided self-help CBT program. *Behav Sci (Basel)*, *7*(3), 39. https://doi.org/10.3390/bs7030039

Bauer, S., & Moessner, M. (2013). Harnessing the power of technology for the treatment and prevention of eating disorders. *The International Journal of Eating Disorders*, *46*(5), 508–515. https://doi.org/10.1002/eat.22109

Beintner, I., Jacobi, C., & Schmidt, U. H. (2014). Participation and outcome in manualized self-help for bulimia nervosa and binge eating disorder—a systematic review and metaregression analysis. *Clinical Psychology Review*, *34*(2), 158–176. https://doi.org/10.1016/j.cpr.2014.01.003

Berg, K. C., Cao, L., Crosby, R. D., Engel, S. G., Peterson, C. B., Crow, S. J., le Grange, D., Mitchell, J. E., Lavender, J. M., Durkin, N., & Wonderlich, S. A. (2017). Negative affect and binge eating: Reconciling differences between two analytic approaches in ecological momentary assessment research. *INT J EAT DISORD*, *50*(10), 1222–1230. https://doi.org/10.1002/eat.22770

Berg, K. C., Crosby, R. D., Cao, L., Peterson, C. B., Engel, S. G., Mitchell, J. E., & Wonderlich, S. A. (2013). Facets of negative affect prior to and following binge-only, purge-only, and binge/purge events in women with bulimia nervosa. *Journal of Abnormal Psychology*, *122*(1), 111–118. https://doi.org/10.1037/a0029703

Berg, K. C., Peterson, C. B., Frazier, P., & Crow, S. J. (2012). Psychometric evaluation of the eating disorder examination and eating disorder examination-questionnaire: A systematic review of the literature. *The International Journal of Eating Disorders*, *45*(3), 428–438. https://doi.org/10.1002/eat.20931

Bertrand, M., & Kamenica, E. (2018). *Coming apart? Cultural distances in the United States over time. National Bureau of Economic Research Working Paper Series, No. 24771*. https://doi.org/10.3386/w24771

Brownley, K. A., Berkman, N. D., Peat, C. M., Lohr, K. N., Cullen, K. E., Bann, C. M., & Bulik, C. M. (2016). Binge-eating disorder in adults: A systematic review and meta-analysis. *Annals of Internal Medicine*, *165*(6), 409–420. https://doi.org/10.7326/M15-2455

Bulik, C. M., Butner, J. E., Tregarthen, J., Thornton, L. M., Flatt, R. E., Smith, T., Carroll, I. M., Baucom, B. R. W., & Deboeck, P. R. (2020). The binge eating genetics initiative (BEGIN): Study protocol. *BMC Psychiatry*, *20*(1), 307. https://doi.org/10.1186/s12888-020-02698-7

Chapa, D. A. N., Hagan, K. E., Forbush, K. T., Clark, K. E., Tregarten, J. P., & Argue, S. (2020). Longitudinal trajectories of behavior change in a national sample of patients seeking eating-disorder treatment. *The International Journal of Eating Disorders*, *53*(6), 917–925. https://doi.org/10.1002/eat.23272

Crow, S. (2014). Treatment of binge-eating disorder. *Curr Treat Options Psychiatry*, *1*(4), 307–314. https://doi.org/10.1007/s40501-014-0023-4

Fairburn, C. G., & Beglin, S. J. (1994). Assessment of eating disorders: Interview or self-report questionnaire? *The International Journal of Eating Disorders*, *16*(4), 363–370.

Fassino, S., Pierò, A., Tomba, E., & Abbate-Daga, G. (2009). Factors associated with dropout from treatment for eating disorders: A comprehensive literature review. *BMC Psychiatry*, *9*(1), 67. https://doi.org/10.1186/1471-244X-9-67

Fichter, M. M., & Quadflieg, N. (2016). Mortality in eating disorders - results of a large prospective clinical longitudinal study. *The International Journal of Eating Disorders*, *49*(4), 391–401. https://doi.org/10.1002/eat.22501

Forrest, L. N., Bodell, L. P., Witte, T. K., Goodwin, N., Bartlett, M. L., Siegfried, N., Eddy, K. T., Thomas, J. J., Franko, D. L., & Smith, A. R. (2016). Associations between eating disorder symptoms and suicidal ideation through thwarted belongingness and perceived burdensomeness among eating disorder patients. *J Affect Disorders*, *195*, 127–135. https://doi.org/10.1016/j.jad.2016.02.017

Goldschmidt, A. B., Crosby, R. D., Cao, L., Engel, S. G., Durkin, N., Beach, H. M., Berg, K. C., Wonderlich, S. A., Crow, S. J., & Peterson, C. B. (2014). Ecological momentary assessment of eating episodes in obese adults. *Psychosom Med*, *76*(9), 747–752. https://doi.org/10.1097/PSY.0000000000000108

Goldschmidt, A. B., Crosby, R. D., Cao, L., Wonderlich, S. A., Mitchell, J. E., Engel, S. G., & Peterson, C. B. (2018). A preliminary study of momentary, naturalistic indicators of binge-eating episodes in adults with obesity. *The International Journal of Eating Disorders*, *51*(1), 87–91. https://doi.org/10.1002/eat.22795

Grilo, C. M., White, M. A., & Masheb, R. M. (2009). DSM-IV psychiatric disorder comorbidity and its correlates in binge eating disorder. *The International Journal of Eating Disorders*, *42*(3), 228–234. https://doi.org/10.1002/eat.20599

Haedt-Matt, A. A., & Keel, P. K. (2011). Revisiting the affect regulation model of binge eating: A meta-analysis of studies using ecological momentary assessment. *Psychological Bulletin*, *137*(4), 660–681. https://doi.org/10.1037/a0023660

Hay, P., Bacaltchuk, J., & Stefano, S. (2004). Psychotherapy for bulimia nervosa and binging. *Cochrane Database of Systematic Reviews*, *3*, CD000562. https://doi.org/10.1002/14651858.CD000562.pub2

Health Resources & Services Administration. (2021, September 1). Shortage Areas. Retrieved from: https://data.hrsa.gov/topics/health-workforce/shortage-areas

Hilbert, A., & Tuschen-Caffier, B. (2007). Maintenance of binge eating through negative mood: A naturalistic comparison of binge eating disorder and bulimia nervosa. *The International Journal of Eating Disorders*, *40*(6), 521–530. https://doi.org/10.1002/eat.20401

Huas, C., Godart, N., Caille, A., Pham-Scottez, A., Foulon, C., Divac, S. M., Lavoisy, G., Guelfi, J. D., Falissard, B., & Rouillon, F. (2013). Mortality and its predictors in severe bulimia nervosa patients. *Eur Eat Disord Rev*, *21*(1), 15–19. https://doi.org/10.1002/erv.2178

Hudson, J. I., Hiripi, E., Pope, H. G., Jr., & Kessler, R. C. (2007). The prevalence and correlates of eating disorders in the National Comorbidity Survey Replication. *Biological Psychiatry*, *61*(3), 348–358. https://doi.org/10.1016/j.biopsych.2006.03.040

Kazdin, A. E., Fitzsimmons-Craft, E. E., & Wilfley, D. E. (2017). Addressing critical gaps in the treatment of eating disorders. *The International Journal of Eating Disorders*, *50*(3), 170–189. https://doi.org/10.1002/eat.22670

Kessler, R. C., Berglund, P. A., Chiu, W. T., Deitz, A. C., Hudson, J. I., Shahly, V., Aguilar-Gaxiola, S., Alonso, J., Angermeyer, M. C., Benjet, C., Bruffaerts, R., de Girolamo, G., de Graaf, R., Maria Haro, J., Kovess-Masfety, V., O'Neill, S., Posada-Villa, J., Sasu, C., Scott, K., … Xavier, M. (2013). The prevalence and correlates of binge eating disorder in the World Health Organization world mental health surveys. *Biol Psychiatry*, *73*(9), 904–914. https://doi.org/10.1016/j.biopsych.2012.11.020

Kim, J. P., Sadeh-Sharvit, S., Welch, H. A., Neri, E., Tregarthen, J., & Lock, J. (2022). Eating disorders early app use mediates treatment effect on clinical improvement. *The International Journal of Eating Disorders*, *55*(3), 382–387. https://doi.org/10.1002/eat.23652

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*(9), 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Latner, J. D., Wilson, G. T., Stunkard, A. J., & Jackson, M. L. (2002). Self-help and long-term behavior therapy for obesity. *Behaviour Research and Therapy*, *40*(7), 805–812. https://doi.org/10.1016/s0005-7967(01)00102-4

Lowe, B., Decker, O., Muller, S., Brahler, E., Schellberg, D., Herzog, W., & Herzberg, P. Y. (2008). Validation and standardization of the generalized anxiety disorder screener (GAD-7) in the general population. *Medical Care*, *46*(3), 266–274. https://doi.org/10.1097/MLR.0b013e318160d093

Mussell, M. P., Crosby, R. D., Crow, S. J., Knopke, A. J., Peterson, C. B., Wonderlich, S. A., & Mitchell, J. E. (2000). Utilization of empirically supported psychotherapy treatments for individuals with eating disorders: A survey of psychologists. *The International Journal of Eating Disorders*, *27*(2), 230–237. https://doi.org/10.1002/(sici)1098-108x(200003)27:2<230::aid-eat11>3.0.co;2-0

National Council for Mental Wellbeing. (2017). The psychiatric shortage causes and solutions. National Council for Mental Wellbeing Retrieved from https://www.thenationalcouncil.org/wp-content/uploads/2017/03/Psychiatric-Shortage_National-Council-.pdf?daf=375ateTbd56

Pew Research Center. (2021). *Mobile fact sheet*. Pew Research Center Retrieved from https://www.pewresearch.org/internet/fact-sheet/mobile/

Pisetsky, E. M., Thornton, L. M., Lichtenstein, P., Pedersen, N. L., & Bulik, C. M. (2013). Suicide attempts in women with eating disorders. *Journal of Abnormal Psychology*, *122*(4), 1042–1056. https://doi.org/10.1037/a0034902

SAS Institute Inc. (2013). *Base SAS® 9.4 procedures guide: Statistical procedures*. SAS Institute Inc.

Schoch, A. H., & Raynor, H. A. (2012). Social desirability, not dietary restraint, is related to accuracy of reported dietary intake of a laboratory meal in females during a 24-hour recall. *Eating Behaviors*, *13*(1), 78–81. https://doi.org/10.1016/j.eatbeh.2011.11.010

Shapiro, J. R., Berkman, N. D., Brownley, K. A., Sedway, J. A., Lohr, K. N., & Bulik, C. M. (2007). Bulimia nervosa treatment: A systematic review of randomized controlled trials. *The International Journal of Eating Disorders*, *40*(4), 321–336. https://doi.org/10.1002/eat.20372

Smyth, J., Wonderlich, S., Crosby, R., Miltenberger, R., Mitchell, J., & Rorty, M. (2001). The use of ecological momentary assessment approaches in eating disorder research. *The International Journal of Eating Disorders*, *30*(1), 83–95. https://doi.org/10.1002/eat.1057

Smyth, J. M., Wonderlich, S. A., Heron, K. E., Sliwinski, M. J., Crosby, R. D., Mitchell, J. E., & Engel, S. G. (2007). Daily and momentary mood and stress are associated with binge eating and vomiting in bulimia nervosa patients in the natural environment. *Journal of Consulting and Clinical Psychology*, *75*(4), 629–638. https://doi.org/10.1037/0022-006X.75.4.629

Spitzer, R. L., Kroenke, K., Williams, J. B., & Lowe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Termorshuizen, J. D., Watson, H. J., Thornton, L. M., Borg, S., Flatt, R. E., MacDermod, C. M., Harper, L. E., Furth, E. F., Peat, C. M., & Bulik, C. M. (2020). Early impact of COVID-19 on individuals with self-reported eating disorders: A survey of ~1,000 individuals in the United States and The Netherlands. *Int J Eat Disord*, *53*(11), 1780–1790. https://doi.org/10.1002/eat.23353

Thomas, K. C., Ellis, A. R., Konrad, T. R., Holzer, C. E., & Morrissey, J. P. (2009). County-level estimates of mental health professional shortage in the United States. *Psychiatric Services*, *60*(10), 1323–1328. https://doi.org/10.1176/ps.2009.60.10.1323

Thornton, L. M., Munn-Chernoff, M. A., Baker, J. H., Juréus, A., Parker, R., Henders, A. K., Larsen, J. T., Petersen, L., Watson, H. J., Yilmaz, Z., Kirk, K. M., Gordon, S., Leppä, V. M., Martin, F. C., Whiteman, D. C., Olsen, C. M., Werge, T. M., Pedersen, N. L., Kaye, W., … Bulik, C. M.

(2018). The anorexia nervosa genetics initiative (ANGI): Overview and methods. *Contemp Clin Trials*, *74*, 61–69. https://doi.org/10.1016/j.cct.2018.09.015

Thornton, L. M., Watson, H. J., Jangmo, A., Welch, E., Wiklund, C., von Hausswolff-Juhlin, Y., Norring, C., Herman, B. K., Larsson, H., & Bulik, C. M. (2017). Binge-eating disorder in the Swedish national registers: Somatic comorbidity. *Int J Eat Disord*, *50*(1), 58–65. https://doi.org/10.1002/eat.22624

Tregarthen, J., Paik Kim, J., Sadeh-Sharvit, S., Neri, E., Welch, H., & Lock, J. (2019). 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*, *6*(11), e14972. https://doi.org/10.2196/14972

Tregarthen, J. P., Lock, J., & Darcy, A. M. (2015). Development of a smartphone application for eating disorder self-monitoring. *The International Journal of Eating Disorders*, *48*(7), 972–982. https://doi.org/10.1002/eat.22386

Waller, G. (1997). Drop-out and failure to engage in individual outpatient cognitive behavior therapy for bulimic disorders. *The International Journal of Eating Disorders*, *22*(1), 35–41. https://doi.org/10.1002/(sici)1098-108x(199707)22:1<35::aid-eat4>3.0.co;2-3

Welch, E., Jangmo, A., Thornton, L. M., Norring, C., von Hausswolff-Juhlin, Y., Herman, B. K., Pawaskar, M., Larsson, H., & Bulik, C. M. (2016). Treatment-seeking patients with binge-eating disorder in the Swedish national registers: Clinical course and psychiatric comorbidity. *BMC Psychiatry*, *16*, 163. https://doi.org/10.1186/s12888-016-0840-7

How to cite this article: Flatt, R. E., Thornton, L. M., Smith, T., Mitchell, H., Argue, S., Baucom, B. R. W., Deboeck, P. R., Adamo, C., Kilshaw, R. E., Shi, Q., Tregarthen, J., Butner, J. E., & Bulik, C. M. (2022). Retention, engagement, and binge-eating outcomes: Evaluating feasibility of the Binge-Eating Genetics Initiative study. *International Journal of Eating Disorders*, *55*(8), 1031–1041. https://doi.org/10.1002/eat.23726