



## Changes in stress, burnout, and resilience associated with an 8-week intervention with relational agent “Woebot”

Emily Durden, Maddison C. Pirner, Stephanie J. Rapoport, Andre Williams, Athena Robinson, Valerie L. Forman-Hoffman\*

Woebot Health, 535 Mission Street, 14th Floor, San Francisco, CA 94105, United States of America

### ARTICLE INFO

#### Keywords:

Conversational agents  
Mobile applications  
Cognitive behavioral therapy  
Stress  
Burnout  
Resilience

### ABSTRACT

**Background:** Research investigating the potential for digital mental health interventions with integrated relational agents to improve mental health outcomes is in its infancy. By delivering evidence-based mental health interventions through tailored, empathic conversations, relational agents have the potential to help individuals manage their stress and mood, and increase positive mental health.

**Aims:** The aims of this study were twofold: 1) to assess whether a smartphone app delivering mental health support through a relational agent, *Woebot*, is associated with changes in stress, burnout, and resilience over 8 weeks, and 2) to identify demographic and clinical factors associated with changes in these outcomes.

**Method:** This exploratory, non-randomized, single-armed, open-labeled trial was conducted from May to July 2022. A total of 256 adults (mean age  $39 \pm 13.35$ ; 72 % females) recruited through social media advertising enrolled in the study. Participants completed an 8-week intervention period during which they were invited to use a smartphone app called *Woebot-LIFE* that delivers cognitive behavioral therapy through a relational agent called “*Woebot*”. Participant-reported measures of stress, burnout, and resilience were collected at Baseline, and Week 8. Changes in these outcomes during the study period were assessed. Bivariate and stepwise multiple regression modeling was used to identify sociodemographic and clinical factors associated with observed changes over the 8-week study period.

**Results:** Exposure to *Woebot-LIFE* was associated with significant reductions in perceived stress and burnout and significantly increased resilience over the 8-week study period. A greater reduction in stress was observed among those with clinically elevated mood symptoms (i.e., Patient Health Questionnaire-8 or Generalized Anxiety Disorder 7-item scores  $\geq 10$ ) at baseline compared to those without; however, the differences in the improvements in resilience scores and burnout between the two groups were not statistically significant. Although a difference in the magnitude of change in stress was observed for participants with and without clinically elevated mood symptoms at baseline, significant improvements in stress, burnout, and resilience over the 8-week study period were observed for both groups. Bivariate analyses showed that race, insurance type, and baseline level of resilience were associated with changes in each of the outcomes, though baseline resilience was the only factor that remained significantly associated with changes in the outcomes in the stepwise multiple regression analyses.

**Conclusion:** Results of this single-arm, exploratory study suggest that conversational agent-guided mental health interventions such as *Woebot-LIFE* may be associated with reduced stress and burnout and increased resilience in both clinical and non-clinical populations.

**Abbreviations:** AIC, akaike information criterion; BRS, brief resilience scale; CBT, cognitive behavioral therapy; CONSORT, consolidated standards of reporting trials; DBT, dialectical behavior therapy; DMHIs, digital mental health interventions; GAD, generalized anxiety disorder; GAD-7, generalized anxiety disorder 7-item self-report tool; IRB, institutional review board; IPT, interpersonal psychotherapy; MCID, minimal clinically important difference; PHQ-8, patient health questionnaire 8-item self-report tool used to screen for depression and measure symptom severity; PSS, perceived stress scale; PSS-10, perceived stress scale, 10-item version; VIF, variance inflation factors; WB-LIFE, a relational agent that offers CBT guided self-help.

\* Corresponding author.

E-mail address: [valerie\\_hoffman@woebothealth.com](mailto:valerie_hoffman@woebothealth.com) (V.L. Forman-Hoffman).

<https://doi.org/10.1016/j.invent.2023.100637>

Received 17 January 2023; Received in revised form 8 June 2023; Accepted 12 June 2023

Available online 14 June 2023

2214-7829/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Key constructs within mental health include stress, burnout, and resilience. An estimated one in four United States (US) adults reports being negatively affected by stress (Cooke et al., 2020), which the World Health Organization (WHO) defines as “a state of worry or mental tension caused by a difficult situation.” (World Health Organization (WHO), 2023) When prolonged or chronic, stress increases the risk of many prevalent and costly medical conditions, including depression (Colodro Conde et al., 2019), anxiety, incident coronary heart disease (Richardson et al., 2012), type 2 diabetes (Harris et al., 2017), and autoimmune disorders (Song et al., 2018). Chronic stress is also a risk factor for burnout, a state characterized by mental and emotional exhaustion, detachment from work or social demands, and reduced feelings of accomplishment (Maslach et al., 1997). Indeed, chronic stress is a leading cause of morbidity and disability in the US and elsewhere and is associated with substantial economic costs to society and to payers of healthcare (Hassard et al., 2018). Burnout, defined as a, “syndrome conceptualized as resulting from chronic workplace stress that has not been successfully managed (World Health Organization (WHO), 2019),” also has been associated with a number of the same types of physical, mental, and occupational effects as stress (e.g. cardiovascular problems, pain, fatigue, insomnia, depression, job dissatisfaction, absenteeism and presenteeism (Salvagioni et al., 2017)).

Resilience is the capacity to adapt to stress and adversity while maintaining normal psychological and physical functioning (Southwick and Charney, 2012). As such, resilience is negatively correlated with psychological distress, including depression and anxiety, and positively correlated with indicators of psychological well-being including life satisfaction and positive emotions (Hu et al., 2015). Resilience can be understood as the outcome of the interaction of the person with their environment (Rutten et al., 2013), modifiable and potentially improved by intervention (Connor and Zhang, 2006; Southwick et al., 2011).

Cognitive behavioral therapy (CBT) is a manualized, time-limited psychotherapy that has demonstrated effectiveness for reducing stress (Hofmann et al., 2012) and burnout (Abola et al., 2017) and increasing resilience (Joyce et al., 2018). However, access to CBT is limited as the demand for mental health services far exceeds the supply of qualified mental health professionals. Indeed, over 150 million Americans reside in areas experiencing mental health professional shortages (Bureau of Health Workforce, Health Resources and Services Administration (HRSA), 2022). Additional barriers to professional mental health support include stigma, lack of time, and lack of affordability.

The translation of effective psychotherapeutic interventions such as CBT into digital formats offers opportunities to extend these interventions to a large scale, potentially reaching those who otherwise would receive no support. Indeed, the past two decades have seen a rapid proliferation of CBT-informed digital solutions for mental health concerns. In particular, smartphone applications (apps) consisting of self-guided interventions focused on stress management and the promotion of positive mental health have become especially popular (Lau et al., 2020). Meta-analytic evidence demonstrates that mental health interventions delivered through mobile apps are efficacious in managing stress (Linardon et al., 2019) and reducing symptoms of depression (Firth et al., 2017a) and anxiety (Firth et al., 2017b). Despite their accessibility and demonstrated efficacy, the impact of digital health solutions for mental health concerns has been limited by low uptake and engagement (Waller and Gilbody, 2009; Baumel et al., 2019).

Recent years have seen the emergence of conversational agents, otherwise known as “chatbots”, integrated into DMHIs. Mental health conversational agents perform a variety of functions, most commonly screening and diagnosis, content delivery, and symptom management (Boucher et al., 2021). Mental health conversational agents designed to build a rapport and therapeutic bond with users are referred to as relational agents. Conversational and relational agents delivering content are primarily CBT-based (Abd-alrazaq et al., 2019), with some

incorporating additional approaches including mindfulness, acceptance and commitment therapy, and dialectical behavior therapy.

While empirical research investigating these interventions is limited, available evidence suggests the potential for mental health conversational agents to improve a variety of mental health outcomes spanning both clinical and non-clinical issues. A 2020 meta-analysis of 12 studies found that conversational agents were effective for improving depression, psychological distress, stress, and acrophobia, though the evidence was considered weak due to the paucity of studies, small samples, low quality of evidence, and a high estimated risk of bias in the included studies (Abd-Alrazaq et al., 2020). A separate meta-analysis of 11 trials in adults with depression and anxiety symptoms found that conversational agents significantly improved depression symptoms, with larger effect sizes observed in samples with clinically diagnosed anxiety or depression and for those interventions with embodied (i.e., relational) conversational agents (Lim et al., 2022).

Research investigating the use of mental health conversational agents to reduce stress appears to be limited to one small pilot randomized controlled trial (RCT). In this study of 28 participants from a non-clinical population, researchers found improvements in psychological well-being and perceived stress after a two-week positive psychology- and CBT-informed conversational agent intervention, relative to a waitlist control (Ly et al., 2017). To our knowledge, no studies have investigated the effectiveness of mental health conversational agents to reduce burnout or increase resilience.

Woebot is a relational agent, embedded in a smartphone application called Woebot LIFE (WB-LIFE), that offers CBT-based guided self-help through text-based conversation, and has demonstrated feasibility, acceptability and preliminary efficacy in improving clinical outcomes among various populations including anxiety and depression symptoms among young adults (Fitzpatrick et al., 2017), mood among postpartum women (Ramachandran et al., n.d.; Suharwardy et al., n.d.), and substance use among adults (Prochaska et al., 2021a; Prochaska et al., 2021b). WB-LIFE's potential to impact stress, burnout, and resilience, while implied by research demonstrating the efficacy of CBT to improve these constructs, has not been investigated to date.

### 1.1. Aims

In this exploratory study, we assess changes in stress, burnout, and resilience observed over an 8-week trial with WB-LIFE as well as sociodemographic and clinical factors associated with change in these outcomes. We give special attention to baseline clinical factors, specifically depression and anxiety symptomatology, as meta-analytic evidence indicates larger effect sizes associated with conversational agent-delivered psychotherapy for clinical versus non-clinical samples (Lim et al., 2022). Identifying key associations between user demographic and clinical characteristics and changes in outcomes may help elucidate whether the potential benefits of this and other agent-guided CBT interventions may be limited to certain groups or extend to all. Indeed, such research is particularly relevant given the novelty of mental health relational agents, and how little is known about their impact on wellness outcomes. Given the burgeoning mental health crisis, and the well-recognized need for early intervention, understanding the usefulness and applicability of evidence-based, scalable programs among broad populations is now a major public health priority.

## 2. Materials and methods

### 2.1. Study design

This exploratory, non-randomized, single-armed, open-labeled trial was approved by the WIRB-Copernicus IRB Group (WCG) institutional review board on January 20, 2022 and conducted from May through July 2022. The trial was retrospectively registered on ClinicalTrials.gov (#NCT05672745) on January 4, 2023, and was designed to explore the

associations between user characteristics and clinical outcomes in the WB-LIFE program over the course of an 8-week intervention period.

## 2.2. Participants

In May 2002, a convenience sample of adult participants was recruited through paid social media advertisements for participants in a mental wellness study testing a digital tool used for emotional support and mood management. To be included, individuals had to be US residents able to read and write in English with regular access to a smartphone and available and willing to complete the study's survey assessments. Individuals were excluded from the study if they were younger than 18 years of age, disclosed current suicidal ideation with a plan or intent or a suicide attempt within the past 12 months, self-reported a lifetime diagnosis of a psychotic disorder or bipolar disorder, or reported previous use of WB-LIFE.

## 2.3. Study procedures

Individuals interested in participating in the study were directed to the study's landing page via embedded links in online advertisements. After reviewing the landing page, interested individuals were directed to the informed consent form. Upon signing the informed consent, individuals received an email with a link to the screener/baseline survey, which was to be completed within 5 days of receipt. Immediately upon completion of the screener/baseline survey, eligible individuals were emailed a link to access and download the WB-LIFE app and were instructed to register using the study's referral code and their credentials within 3 days to enroll in the study. Ineligible individuals were thanked for their interest and provided with a list of mental health-related supportive resources. Enrolled participants entered an 8-week intervention period, during which they were invited to engage with WB-LIFE as much as preferred (i.e., no expectations for use were recommended) and complete a series of online assessments at study day 3, week 4, and week 8. Access to the WB-LIFE app was terminated upon study completion.

## 2.4. Intervention

### 2.4.1. Woebot LIFE (WB-LIFE) intervention

WB-LIFE is primarily grounded in CBT, yet also offers elements of other empirically supported treatments including interpersonal psychotherapy (IPT) and dialectical behavior therapy (DBT). Such psychotherapeutic content is delivered conversationally via text-based messages with the program's relational agent, *Woebot*. The guided self-help program invites users to engage in a wide variety of ways (some guided by interactions with *Woebot* and others outside of the interactions with *Woebot*), including but not limited to mood tracking, in-vivo application of emotion regulation skills, reflection of progress, gratitude journaling, mindfulness practice as well as learning and chatting about techniques for mood management. Each conversation is tailored to the user's self-reported needs and area of desired focus in the moment. Previous research on the various programs featuring *Woebot* have demonstrated feasibility, acceptability and efficacy (Fitzpatrick et al., 2017; Suharwardy et al., n.d.; Prochaska et al., 2021a; Prochaska et al., 2021b). Moreover, a retrospective analysis with over 36,000 users of products that include *Woebot* demonstrated that *Woebot* establishes a working alliance with users (Darcy et al., 2021), a construct thought to be central to therapeutic outcomes (Martin et al., 2000). A sample screenshot from WB-LIFE is shown in Fig. 1. For the purpose of this study, WB-LIFE was tested as an 8-week program.

### 2.4.2. Safety assessment

No adverse events (AE) were reported to the Principal Investigator during the conduct of the study.

While not designed to be a crisis service, WB-LIFE's natural language processing algorithm detects utterances that might indicate a crisis. WB-

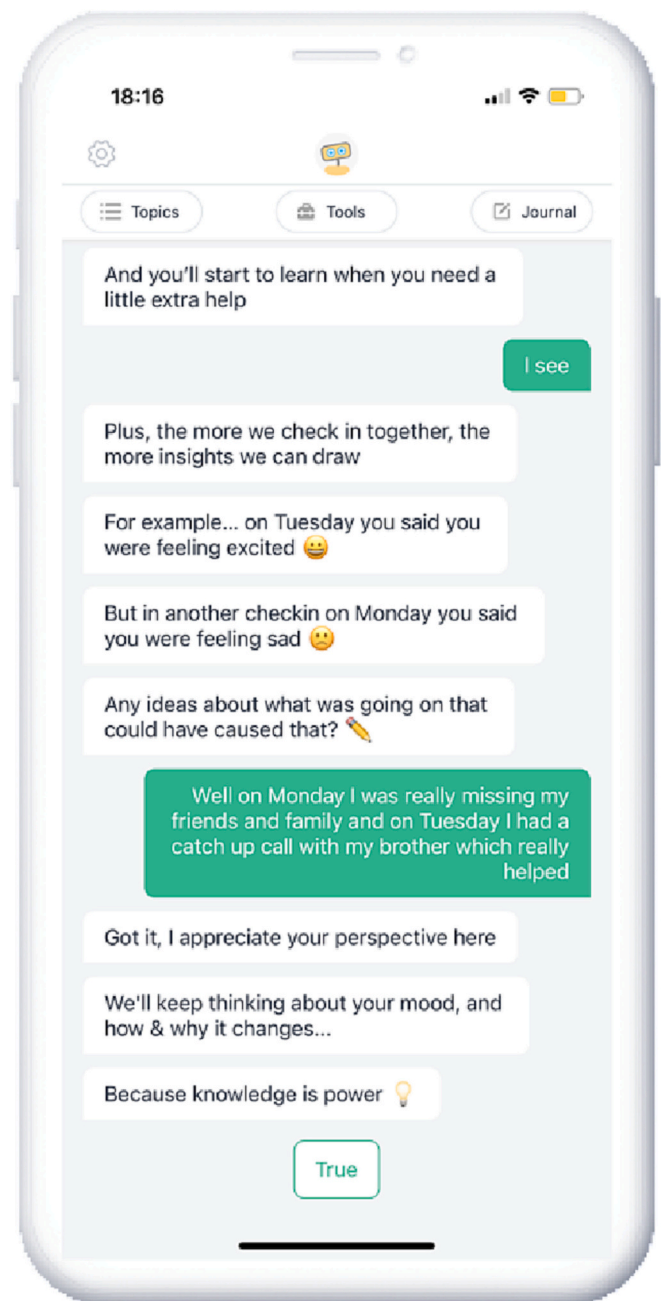


Fig. 1. Mood monitoring and tracking within the WB-LIFE product.

LIFE responds to such language detection by asking users to confirm whether or not they may be in crisis and then immediately provides users with a list of resources, including emergency contact phone numbers and crisis hotline information. This list of resources is also always available to the user via their settings and if a user enters "SOS" into free-text. After providing the list of resources, *Woebot* offers tools to assist with addressing distressing thoughts and emotions in the moment.

## 2.5. Measurement of stress, burnout, and resilience

Perceived stress was measured using the Perceived Stress Scale (PSS), a 10-item instrument that measures how often an individual perceives life as unpredictable, uncontrollable, and overwhelming in the past month (Cohen et al., 1983). Participants respond on a 5-point Likert scale ranging from 0 (never) to 4 (very often). Four items are reverse coded and added to the sum of the responses to the remaining six items

to obtain a final score ranging from 0 to 40, with higher scores indicating more perceived stress. Scores of 0–13 are generally considered low stress, 14–26 moderate stress, and 27–40 are high perceived stress. Studies of the internal consistency of the PSS report alphas ranging from 0.74 to 0.91 (Lee, 2012).

Burnout was assessed using a non-proprietary single item validated measure (Dolan et al., 2015; West et al., 2009) that instructs respondents to define burnout for themselves: “Overall, based on your definition of burnout, how would you rate your level of burnout?” Responses range from 1 (“I enjoy my work. I have no symptoms of burnout”) to 5 (“I feel completely burned out and often wonder if I can go on. I am at the point where I may need some changes or may need to seek some sort of help.”) (Dolan et al., 2015). The item is dichotomized as  $\leq 2$  (no symptoms of burnout) vs.  $\geq 3$  (1 or more symptoms).

Resilience was measured using the Brief Resilience Scale (BRS), a 6-item tool used to assess an individual's perceived ability to bounce back or recover from stress or a setback (Smith et al., 2008). The six items contain statements about responses to stressful events and respondents indicate the extent to which they agree with the statements using a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Three statements of the six items are negatively framed; after reverse-coding these items, all items are summed and divided by 6 to obtain an overall resilience score. Internal consistency of the BRS ranges from 0.80 to 0.91 (Smith et al., 2008).

## 2.6. Demographic characteristics

Participant demographic characteristics captured at baseline included age, sex at birth, race, ethnicity, marital status, education, sexual orientation, employment status, and health insurance.

## 2.7. Clinical characteristics

### 2.7.1. Measurement of clinical symptomatology

Depression symptoms were assessed using the Patient Health Questionnaire (PHQ-8) (Kroenke et al., 2009), an 8-item self-report tool used to screen for depression and measure symptom severity. The PHQ-8 is used extensively and has demonstrated reliability as well as sensitivity to clinical change. Scores range from 0 to 24, with the following severity cutoffs: minimal (0–4), mild (5–9), moderate (10–14), moderately severe (15–19), and severe (20–27). Participants whose scores indicated minimal and mild depression and those whose scores indicated moderate, moderately severe, and severe levels of depression were distinguished analytically as subclinical and clinical, respectively.

Anxiety symptoms were assessed using the Generalized Anxiety Disorder (GAD-7) (Spitzer et al., 2006), a 7-item self-report tool that assesses the frequency and severity of anxious thoughts and behaviors over the past two weeks and can be used to identify probable cases of GAD in addition to measuring anxiety symptom severity. Scores range from 0 to 21, with the following severity cutoffs: minimal (0–4), mild (5–9), moderate (10–14), and severe (15–21). Participants with GAD-7 symptom scores greater than or equal to 10 at baseline were distinguished analytically from those without elevated symptoms of depression or anxiety.

In our analyses, “clinically elevated” baseline mood scores are defined as a baseline PHQ-8 or GAD-7 score  $\geq 10$ .

### 2.7.2. Concurrent mental health treatment

Concurrent mental health treatment was defined as mental health treatment at any point in the study. At baseline, participants were asked to indicate if they were currently seeing a therapist and/or if they were taking medications for a mental health concern or condition. At the 8-week survey assessment, participants were asked if psychotropic medications and/or psychotherapy were continued or if either was started during the course of the 8-week study intervention.

## 2.8. App utilization

App use data was used to evaluate use of WB-LIFE during at least 4 weeks of the 8-week study period. Although our prior research has shown significant improvements in depressive and anxiety symptoms in as few as two weeks (Fitzpatrick et al., 2017), in the absence of other data, opening the app on at least 4 out of 8 weeks was selected as the threshold for adherence to allow for the majority of participants sufficient time to experience clinically meaningful improvements in each of the wellness outcomes studied.

## 2.9. Analyses

All variables were descriptively characterized. Continuous variables and outcome scores were summarized on the basis of their means and standard deviations; discrete and/or ordinal variables were summarized by the sample size and proportion of participants in each category. Descriptive statistics were stratified by baseline clinical symptomatology such that participants with clinically elevated mood symptoms as indicated by a baseline PHQ-8 or GAD-7 score greater than or equal to 10 were compared to those without clinically elevated mood symptoms at baseline (i.e., those with PHQ-8 and GAD-7 scores of  $<10$  at baseline).

Linear regression modeling was used to identify characteristics associated with 8-week change scores in perceived stress and resilience. Logistic regression was used to identify characteristics associated with no longer having symptoms of burnout at 8 weeks among those with symptoms of burnout at baseline. For each outcome, bivariate regressions were fit; the independent variables were age, race/ethnicity, sex at birth, sexual orientation, education level, employment status, marital status, health insurance, baseline stress symptoms, baseline burnout, baseline resilience, and baseline depression or anxiety symptomatology. A two-sided test was used to assess the statistical significance of the covariates, the level of significance was 0.05. All independent variables were considered for inclusion in the multiple regression models. Stepwise regression was used to determine the final set of variables for each multiple regression model. An initial multiple regression model was fit; covariates with Variance Inflation Factors (VIF) greater than four were omitted in a stepwise model selection. The stepwise model selection was carried out using Akaike Information Criterion (AIC). Again, two-sided *t*-tests with a 0.05 level of significance were employed to assess significance. For each multiple regression model, model diagnostics, including evaluation of the VIF, were performed to assess model fit. Due to the exploratory nature of the study, adjustments for multiple comparisons were not employed. All analyses were conducted using the R statistical software program (R Core Team, 2021).

### 2.10. Power analysis

The sample yielded  $>99.99\%$ ,  $>99.99\%$ , and  $>99.99\%$  power to detect change in stress, resilience, and burnout, respectively.

## 3. Results

### 3.1. Participants

As shown in the CONSORT diagram (Fig. 2), 1105 potential participants were screened. Of these, 485 were excluded for not meeting eligibility criteria. Following enrollment, an additional 358 participants with unauthorized accounts (e.g., duplicate registrants or later determination of failure to meet inclusion criteria) were withdrawn from the study and omitted from the analyses. An additional six unauthorized registrants completed study procedures and were removed from the analytic sample as they were initially missed due to clerical error. Thus, the final sample included 256 participants.

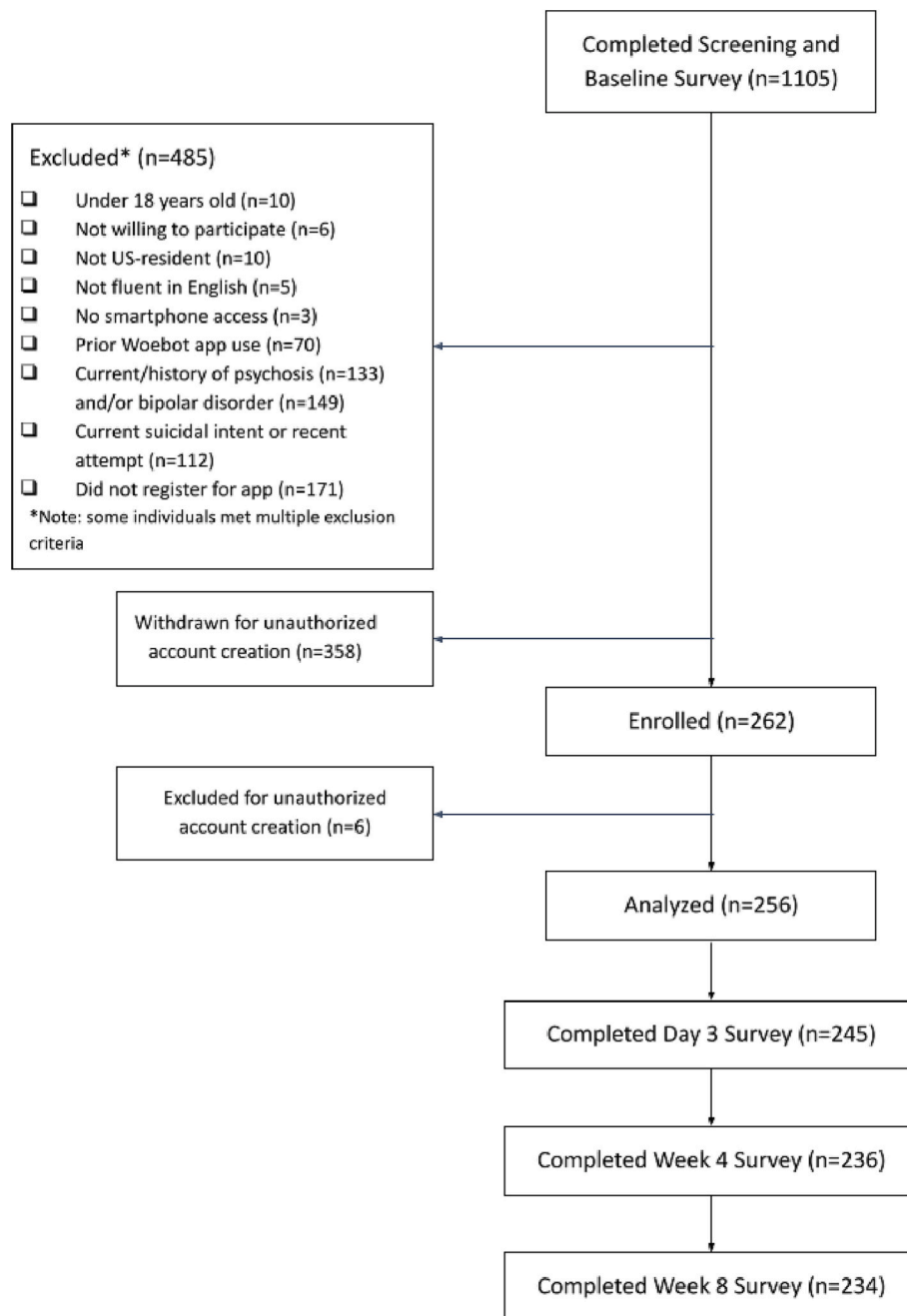


Fig. 2. CONSORT diagram.

### 3.2. Participant characteristics

Table 1 presents the sociodemographic characteristics of the sample overall and by baseline clinical symptomatology. A total of 256 participants were enrolled in the study. Participants had a mean age of 39 years (SD = 13.35), were largely female (72 %), identified as non-Hispanic White (58 %), and self-reported as heterosexual (82 %). Of the total sample, 54 % (n = 139) reported clinically elevated levels of depression or anxiety at baseline. Significantly more female and single participants reported elevated symptomatology (78 %; 43 %, respectively) than not (66 %; 25 %, respectively). There was a significantly greater percentage of non-Hispanic black participants with elevated symptomatology (28 %) than without (15 %). Conversely, there were significantly less non-Hispanic White participants with elevated symptomatology (54 %) than without (63 %). Lastly, there was a significantly

greater percentage of participants without clinical symptomatology (81 %) who engaged with the app for at least 4 of 8 weeks compared to participants with clinical symptomatology (68 %).

### 3.3. Change from baseline in outcome measures

Table 2 presents the baseline and change from baseline in outcome measures for the sample overall and by baseline clinical symptomatology (i.e., baseline PHQ-8 or GAD-7 ≥ 10 vs. baseline PHQ-8 and GAD-7 < 10). Significant reductions in perceived stress and burnout over the 8-week intervention period were found among the overall sample (d = -0.391, p < 0.001), participants with clinically elevated mood symptoms at baseline, and participants without clinically elevated mood symptoms at baseline. Participants with clinically elevated symptoms reported significantly greater reductions in stress compared

**Table 1**  
Participant characteristics, overall and by baseline clinical symptomatology.

	Total Sample (N = 256)	PHQ-8 and GAD-7 < 10 (N = 117)	PHQ-8 or GAD ≥ 10 (N = 139)	P value
Age (years), mean (SD)	39 (13.35)	41 (13.22)	38 (13.42)	0.198
Race/ethnicity, n (%)				0.033
Non-Hispanic Black	56 (22 %)	17 (15 %)	39 (28 %)	
Non-Hispanic White	149 (58 %)	74 (63 %)	75 (54 %)	
Other	51 (20 %)	26 (22 %)	25 (18 %)	
Sex at Birth n (%)				0.043
Female	184 (72 %)	76 (66 %)	108 (78 %)	
Male	71 (28 %)	40 (34 %)	31 (22 %)	
Sexual orientation, n (%)				0.992
Not heterosexual	45 (18 %)	21 (18 %)	24 (17 %)	
Heterosexual	210 (82 %)	95 (82 %)	115 (83 %)	
Education, n (%)				0.401
Some college or technical school	47 (19 %)	18 (16 %)	29 (21 %)	
College graduate	97 (39 %)	43 (38 %)	54 (40 %)	
Graduate or postgraduate degree	74 (30 %)	39 (35 %)	35 (26 %)	
High school only (grade 9–12)	30 (12 %)	12 (11 %)	18 (13 %)	
Employment, n (%)				0.304
Employed full-time	136 (55 %)	68 (61 %)	68 (50 %)	
Employed part-time	26 (11 %)	12 (11 %)	14 (10 %)	
Not employed	56 (23 %)	21 (19 %)	35 (26 %)	
Other	28 (11 %)	10 (9 %)	18 (13 %)	
Marital Status, n (%)				0.011
Divorced/separated/widowed	25 (10 %)	11 (10 %)	14 (10 %)	
Married/partnered/cohabiting	136 (55 %)	73 (65 %)	63 (47 %)	
Single	85 (35 %)	28 (25 %)	57 (43 %)	
Health insurance, n (%)				0.06
Government insurance	90 (37 %)	39 (36 %)	51 (38 %)	
Private insurance	123 (51 %)	62 (57 %)	61 (46 %)	
None/prefer not to answer	30 (12 %)	8 (7 %)	22 (16 %)	
Use of WB-LIFE for at least 4 of 8 weeks	189 (74 %)	95 (81 %)	94 (68 %)	0.020
Concurrent mental health treatment	112 (44 %)	43 (37 %)	69 (50 %)	0.052

PHQ-8 = Patient Health Questionnaire-8 item scale; GAD-7 = Generalized Anxiety Disorder-7 item scale; SD = standard deviation.

to those without clinically elevated symptoms. Meanwhile, significant increases in resilience were found over the same period among the overall sample ( $d = 0.436$ ,  $p < 0.001$ ), participants with clinically elevated symptoms, and participants without clinically elevated symptoms. However, there was no statistically significant difference in change of resilience scores between participants with clinically elevated symptoms and those without. Over half of those with burnout at baseline did not continue to have burnout at 8-weeks ( $p = 53\%$ ,  $p < 0.001$ , Cohen's  $h = 1.64$ ); no differences in this outcome variable were found for those with versus without clinically elevated symptoms at baseline.

**Table 2**  
Baseline and change from baseline in stress, resilience, and burnout, overall and by baseline clinical symptomatology.

	Total sample (N = 256)	PHQ-8 and GAD-7 < 10 (N = 117)	PHQ-8 or GAD ≥ 10 (N = 139)	P value for < 10 vs ≥ 10
Perceived stress, mean (SD)				
Baseline	21.36 (5.61)	18.68 (5.13)	23.60 (4.99)	<0.001
8 weeks	18.71 (5.90)	17.10 (6.19)	20.07 (5.30)	<0.001
Change from baseline	-2.62 (6.69)*	-1.52 (6.80)*	-3.54 (6.48)*	0.022
Resilience, mean (SD)				
Baseline	2.97 (0.88)	3.27 (0.84)	2.72 (0.83)	<0.001
8 weeks	3.30 (0.89)	3.53 (0.94)	3.11 (0.79)	<0.001
Change from baseline	0.32 (0.74)*	0.24 (0.69)*	0.39 (0.77)*	0.128
Burnout, n/N (%)				
Baseline (≥3) (n = 255)	152/255 (60 %)	41/116 (35 %)	111/139 (80 %)	<0.001
8 weeks (n = 233)	87/233 (37 %)*	26/106 (25 %)	61/127 (48 %)*	<0.001
Change (n = 139) <sup>a</sup>	74/139 (53 %)*	22/37 (59 %)*	52/102 (51 %)*	0.38

PHQ-8 = Patient Health Questionnaire-8 item scale; GAD-7 = Generalized Anxiety Disorder-7 item scale; SD = standard deviation.

\* Within column group change  $p < 0.05$ .

<sup>a</sup> Among those with burnout at baseline, improved to no burnout at 8 weeks.

### 3.4. Bivariate associations between sociodemographic variables and outcome measures

Unadjusted bivariate linear and logistic regression was used to evaluate demographic factors associated with changes in stress (see Table 3a) and resilience (see Table 3b), and burnout at 8 weeks among those with burnout at baseline (see Table 3c) over the 8-week study period. Results indicate that age was inversely associated with reduced stress such that smaller reductions in stress was observed with increasing age. Females were more likely to continue to have burnout at 8 weeks than males. Participants not currently employed had less decline in stress than those employed full-time. Non-Hispanic Black participants showed a greater decrease in stress and were more likely to show improvement in burnout compared to their non-Hispanic White counterparts. Insured participants had less decline in stress and were less likely to show improvement in burnout than those without insurance. Meanwhile, non-Hispanic Black participants showed less decline in resilience than non-Hispanic White participants, and insured participants had a greater increase in resilience than uninsured.

### 3.5. Bivariate associations between clinical characteristics and outcome measures

To evaluate clinical factors associated with changes in stress (see Table 3a), resilience (see Table 3b), and burnout (see Table 3c) across the intervention period, unadjusted bivariate linear and logistic regression models were estimated. Results show that participants reporting moderate and high levels of stress at baseline showed a greater decrease in stress than those reporting low levels of stress. The magnitude of the decrease in stress was greater for participants with burnout at baseline than those without. Participants with normal baseline levels of resilience showed a greater decrease in stress than those with low levels, while participants with normal to high levels of resilience showed less decline in resilience and were more likely to show improvement in burnout compared to those with low resilience levels. Participants receiving concurrent mental health treatment showed less decline in stress and were less likely to show improvement in burnout than those not receiving treatment. Participants with clinically elevated levels of

**Table 3a**  
Unadjusted bivariate linear regression models of 8-week change scores in perceived stress.

Variables	Perceived Stress Scale (PSS): 8-week Change Scores		
	Estimates	95 % CI	P-value
Use of WB-LIFE at least 4 of 8 weeks	0.55	-1.5, 2.6	0.6
Age	0.11	0.04, 0.18	0.001
Race/ethnicity			
Non-Hispanic White		Reference level	
Non-Hispanic Black	-4.0	-6.1, -1.9	<0.001
Other	-1.4	-3.6, 0.76	0.2
Sex at birth			
Male		Reference level	
Female	1.3	-0.62, 3.1	0.2
Sexual orientation			
Heterosexual		Reference level	
Not Heterosexual	1.4	-0.92, 3.6	0.2
Education level			
High School		Reference level	
College degree	0.78	-1.9, 3.5	0.6
Graduate or postgraduate degree	-2.2	-5.0, 0.62	0.13
Some college or technical school	1.6	-1.5, 4.8	0.3
Employment			
Full Time		Reference level	
Not Employed	2.7	0.61, 4.8	0.012
Other	1.7	-0.93, 4.3	0.2
Part Time	2.5	-0.27, 5.2	0.077
Marital status			
Married/partnered/cohabiting		Reference level	
Divorced/separated/widowed	0.20	-2.9, 3.3	0.9
Single	-1.4	-3.3, 0.44	0.13
Health insurance			
No insurance/prefer not to answer		Reference level	
Government based insurance	4.3	1.7, 7.0	0.001
Private insurance	3.8	1.2, 6.3	0.004
BL stress symptoms			
Low		Reference level	
Moderate	-10	-13, -7.6	<0.001
High	-15	-18, -12	<0.001
BL burnout	-2.5	-4.2, -0.75	0.005
BL resilience			
Low		Reference level	
Normal	-3.4	-5.1, -1.6	<0.001
High	1.1	-2.3, 4.5	0.5
BL BL symptomatology			
PHQ-8 and GAD-7 < 10		Reference level	
PHQ-8 or GAD-7 ≥ 10	-2.0	-3.7, -0.30	0.022
Concurrent mental health treatment			
No		Reference level	
Yes	2.5	0.79, 4.2	0.004

PHQ-8 = Patient Health Questionnaire-8 item scale; GAD-7 = Generalized Anxiety Disorder-7 item scale; BL = baseline.

**Table 3b**  
Unadjusted bivariate linear regressions of 8-week change scores in resilience.

Predictors	Brief Resilience Scale (BRS): 8-week change scores		
	Estimates	95 % CI	P-value
Use of WB-LIFE at least 4 of 8 weeks	0.14	-0.08, 0.36	0.2
Age	0.00	-0.01, 0.01	0.8
Race/ethnicity			
Non-Hispanic White		Reference level	
Non-Hispanic Black	-0.27	-0.50, -0.04	0.021
Other	0.07	-0.17, 0.32	0.5
Sex at birth			
Male		Reference level	
Female	0.20	0.00, 0.41	0.055
Sexual orientation			
Heterosexual		Reference level	
Not Heterosexual	0.10	-0.15, 0.35	0.4
Education level			
High school		Reference level	
College degree	-0.22	-0.53, 0.09	0.2
Graduate or postgraduate degree	-0.23	-0.55, 0.09	0.2
Some college or technical school	-0.18	-0.55, 0.18	0.3
Employment			
Full Time		Reference level	
Not Employed	0.14	-0.11, 0.39	0.3
Other	-0.05	-0.37, 0.27	0.8
Part	0.06	-0.27, 0.39	0.7
Marital status			
Married/partnered/cohabiting		Reference level	
Divorced/separated/widowed	-0.10	-0.44, 0.24	0.6
Single	-0.10	-0.31, 0.11	0.3
Health insurance			
No insurance/prefer not to answer		Reference level	
Government based insurance	0.32	0.00, 0.63	0.047
Private insurance	0.38	0.08, 0.68	0.013
BL stress symptoms			
Low		Reference level	
Moderate	0.05	-0.28, 0.38	0.8
High	-0.19	-0.58, 0.20	0.3
BL burnout	0.15	-0.04, 0.34	0.13
BL resilience			
Low		Reference level	
Normal	-0.50	-0.68, -0.31	<0.001
High	-0.73	-1.1, -0.37	<0.001
BL symptomatology			
BL PHQ8 and GAD7 < 10		Reference level	
BL PHQ8 or GAD7 ≥ 10	0.15	-0.04, 0.33	0.13
Concurrent mental health treatment			
No		Reference level	
Yes	-0.10	-0.29, 0.09	0.3

PHQ-8 = Patient Health Questionnaire-8 item scale; GAD-7 = Generalized Anxiety Disorder-7 item scale; BL = baseline.

**Table 3c**  
Unadjusted bivariate logistic regressions of 8-week improvement in burnout.

Predictors	Burnout: No symptoms of burnout at 8 weeks		
	Odds Ratio	95 % CI	p-value
Use of WB-LIFE at least 4 of 8 weeks	1.03	0.48, 2.20	>0.9
Age	0.98	0.96, 1.01	0.2
Race/ethnicity		Reference level	
Non-Hispanic White		1.60, 10.1	0.004
Non-Hispanic Black	3.84	0.51, 2.94	0.6
Other	1.23		
Sex at birth		Reference level	
Male		0.16, 0.81	0.015
Female	0.37		
Sexual orientation		Reference level	
Heterosexual		0.40, 2.30	>0.9
Not heterosexual	0.96		
Education level		Reference level	
High school		0.39, 3.33	0.8
College degree	1.13	0.92, 8.35	0.075
Graduate or postgraduate degree	2.70	0.81, 9.31	0.11
Some college or technical school	2.67		
Employment		Reference level	
Full time		0.26, 1.42	0.3
Not employed	0.62	0.18, 1.61	0.3
Other	0.54	0.25, 2.76	0.7
Part	0.81		
Marital status		Reference level	
Married/partnered/cohabiting		0.23, 2.44	0.7
Divorced/separated/widowed	0.77	0.74, 3.24	0.2
Single	1.54		
Health insurance		Reference level	
No insurance/prefer not to answer		0.12, 1.03	0.064
Government-based insurance	0.37	0.12, 0.88	0.031
Private insurance	0.34		
BL stress symptoms		Reference level	
Low		0.24, 3.07	0.8
Moderate	0.87	0.30, 4.90	0.8
High	1.22		
Baseline burnout severity		Reference level	
3		0.20, 1.23	0.14
4	0.50	0.59, 2.72	0.6
5	1.26		
BL resilience		Reference level	
Low		3.05, 14.6	<0.001
Normal/high	6.48		
BL Symptomatology		Reference level	
BL PHQ8 and GAD7 < 10		0.33, 1.51	0.4
BL PHQ8 or GAD7 ≥ 10	0.71		
Concurrent mental health treatment		Reference level	
No		0.23, 0.91	0.028
Yes	0.46		

PHQ-8 = Patient Health Questionnaire-8 item scale; GAD-7 = Generalized Anxiety Disorder-7 item scale; BL = baseline.

mood symptoms at baseline showed a significantly greater decrease in stress than those without. No differences in change in resilience or burnout were found between these groups.

**3.6. Multiple linear regression model of 8-week change scores in stress, resilience, and burnout**

Stepwise linear and logistic regression models were estimated to determine whether changes in stress (see Table 4a), resilience (see Table 4b), and burnout (see Table 4c) vary by sociodemographic factors, clinical characteristics, and clinical symptomatology at baseline.

To identify variables associated with change in stress, the following variables were retained in the final change in stress model: age; sexual orientation; education level; health insurance; baseline levels of stress, burnout, and resilience; and concurrent mental health treatment ( $F_{13,201} = 11.32, p < 0.001, R^2 = 39\%$ ). Results indicate that age was inversely associated with reduced stress such that a smaller reduction in stress was observed with increasing age. The magnitude of the decrease in stress also was increased among participants with burnout at baseline than those without. A greater decrease in stress was seen among participants with government-based insurance, moderate and high baseline levels of stress, and normal levels of resilience compared to those without insurance and with low levels of stress and resilience, respectively.

The following variables were retained in the final change in resilience model: age, baseline levels of stress and resilience, and concurrent

**Table 4a**  
Multiple linear regression model of 8-week change in perceived stress.

Characteristic	Perceived Stress Scale (PSS): 8-week Change Scores		
	Estimate	95 % CI	P-value
Age	0.07	0.01, 0.13	0.023
Sexual orientation		Reference level	
Heterosexual		-0.43, 3.4	0.13
Not heterosexual	1.5		
Education level		Reference level	
High school (grades 9–12)		-1.5, 3.0	0.5
College degree	0.73	-3.9, 0.98	0.2
Graduate or postgraduate degree	-1.5	-0.99, 4.2	0.2
Some college or technical school	1.6		
Health insurance		Reference level	
No insurance/prefer not to answer		-5.8, -0.20	0.036
Government based insurance	-3	-4.2, 0.93	0.2
Private insurance	-1.7		
BL stress symptoms		Reference level	
Low		-12, -6.8	<0.001
Moderate	-9.2	-16, -10	<0.001
High	-13		
BL burnout	-2.4	-3.9, -0.87	0.002
BL resilience		Reference level	
Low		-4.4, -1.2	<0.001
Normal	-2.8	-3.5, 2.4	0.7
High	-0.53		
Concurrent mental health treatment	1.4	-0.07, 2.8	0.062

PHQ-8 = Patient Health Questionnaire-8 item scale; GAD-7 = Generalized Anxiety Disorder-7 item scale; BL = baseline.

Note: A stepwise regression approach was used to determine the final form of the multiple regression model. Akaike Information Criterion (AIC) was used to include (or exclude) variables from the model.



**Table 4b**  
Multiple linear regression model of 8-week change in resilience (BRS).

Characteristic	Brief Resilience Scale (BRS): 8-week change scores		
	Estimate	95 % CI	P-value
Age	-0.01	-0.01, 0.00	0.11
BL stress symptoms		Reference level	
Low		-0.37, 0.28	0.8
Moderate	-0.04		
High	-0.3	-0.68, 0.08	0.13
BL resilience		Reference level	
Low		-1.2, -0.39	<0.001
High	-0.78		
Normal	-0.54	-0.74, -0.34	<0.001
Concurrent mental health treatment	-0.17	-0.37, 0.02	0.084

PHQ-8 = Patient Health Questionnaire-8 item scale; GAD-7 = Generalized Anxiety Disorder-7 item scale; BL = baseline.

Note: A stepwise regression approach was used to determine the final form of the multiple regression model. Akaike Information Criterion (AIC) was used to include (or exclude) variables from the model.

**Table 4c**  
Multiple logistic regression model of 8-week change in burnout among those with burnout at baseline.

Characteristic	Burnout:		
	Odds ratio	95 % CI	P-value
BL resilience		Reference level	
Low		2.60, 13.1	<0.001
Normal/high	5.67		
Concurrent mental health treatment	0.48	0.22, 1.04	0.064

PHQ-8 = Patient Health Questionnaire-8 item scale; GAD-7 = Generalized Anxiety Disorder-7 item scale; BL = baseline.

Note: A stepwise regression approach was used to determine the final form of the multiple regression model. Akaike Information Criterion (AIC) was used to include (or exclude) variables from the model.

mental health treatment ( $F_{6,210} = 7.31, p < 0.001, R^2 = 15\%$ ). Baseline levels of resilience emerged as the only significant predictor of change in resilience in the model: compared to participants with low levels of resilience at baseline, those with normal and high levels showed less increase in resilience across the study period as compared to those with low baseline levels.

Resilience levels at baseline and concurrent mental health treatment comprised the final model to identify factors associated with change in burnout. In assessing model fit, the difference in the null and model deviance is 25.67 (179.45–153.77) with 129 and 127 degrees of freedom, respectively. The corresponding *p*-value, based on the Chi-Square test, is <0.001. The percent of variance explained is 14 %. Results reveal that participants with normal to high levels of resilience were more likely to show improvement in burnout over the study period compared to those with low levels.

## 4. Discussion

### 4.1. Principal findings

In this single-arm exploratory trial of WB-LIFE, we sought to assess the magnitude of change in perceived stress, burnout, and resilience over an 8-week study period as well as explore potential relationships between demographic and clinical characteristics and change in each

outcome. Exposure to WB-LIFE was associated with significant reductions in perceived stress and burnout and significantly increased resilience over the 8-week study period. A greater reduction in stress was observed among those with clinically elevated mood symptoms (i.e., PHQ-8 or GAD-7 scores  $\geq 10$ ) at baseline compared to those without clinically elevated symptoms; however, differences in the improvement in resilience scores and burnout between the two groups were not statistically significant. While the magnitude of change in stress varied for those with and without clinically elevated mood symptoms at baseline, significant improvements in stress, burnout, and resilience over the 8-week study period were observed for both groups.

Baseline levels of stress, burnout, and resilience were generally associated with the magnitude of 8-week changes in those measures. These findings are consistent with expectations given that greater absolute scores at baseline offer more room for change than smaller scores. Interestingly, participants reporting concurrent mental health treatment experienced smaller reductions in stress and reduced likelihood of showing improvement in burnout than those not reporting concurrent mental health treatment. One potential explanation for these findings is that those receiving concurrent mental health treatment could be doing so for a condition or concerns that are weakly related to the outcomes of this analysis.

Despite the identification of several significant demographic characteristics in the bivariate analyses, most of the significance was not retained in the multiple regression models that adjusted for baseline levels of each outcome. Baseline resilience was the only factor associated with changes in each of the outcomes in both the bivariate and multivariate analyses, consistent with research showing that resilience is negatively correlated with psychological distress and positively correlated with indicators of psychological well-being (Hu et al., 2015). That most of the characteristics found to be significantly associated with the outcomes in the bivariate analysis were not significant in the multivariate analyses that controlled for baseline levels of the outcomes suggests that the benefits of WB-LIFE may extend broadly to those across different demographic subgroups.

Results of this study add to the nascent and growing body of empirical research investigating the impact of mental health relational agents on psychological well-being. Our results are consistent with those of previous studies demonstrating support for mental health conversational agents to reduce stress and depression symptoms in both clinical and non-clinical samples (Abd-Alrazaq et al., 2020). To date, no studies of which we are aware have investigated the impact of mental health conversational agents on burnout or resilience.

### 4.2. Limitations

Results of this study must be considered in light of several limitations. First, due to its observational design and lack of a control group, causal links between changes in stress, burnout, and resilience, and the intervention cannot be inferred. A randomized controlled trial should be conducted to determine efficacy in improving these outcomes. Second, the extent to which the statistically significant improvements in the outcomes are clinically meaningful is unknown. To our knowledge, there are no accepted minimal clinically important differences (MCIDs), or magnitudes of change that are meaningful to individuals, for the PSS-10 (measuring perceived stress), BRS (measuring resilience), or the single-item burnout measure evaluated in our study. Third, while our sample was relatively diverse for a naturalistic sample, it was disproportionately female and well-educated, with 69 % having college or advanced degrees. Additional studies with more representative samples are warranted to ensure generalizability. Fourth, we did not find a significant association between the dichotomous indicator of app utilization (i.e., use during 4 of 8 weeks) and any of the outcomes examined in adjusted models. Additional analyses including more fine-grained utilization and additional engagement variables would permit us to determine if the null finding may be the result of the crude measure used

in the current study or if perhaps there is no linear relationship that would be captured in analytic models. Finally, this study examined symptom changes over an 8-week intervention period and did not examine durability of observed changes. In future research, we plan to follow participants beyond the intervention period to investigate longer term outcomes.

## 5. Conclusion

Results of this single-arm, exploratory study suggest that WB-LIFE, a relational agent-guided mental health intervention, may be associated with reduced stress and burnout and increased resilience in populations presenting with clinical and non-clinical levels of mood and anxiety symptoms. Response to such interventions irrespective of baseline clinical symptomatology or baseline demographic or clinical characteristics suggests that delivery to a broad population may be most beneficial to public health. Hypothesis-testing studies are warranted to draw conclusions about efficacy.

## Funding

This work was funded by Woebot Health.

## Ethics approval and consent to participate

WIRB-Copernicus IRB Group (WCG) Institutional Review Board (WCG-IRB) reviewed and approved this study on January 20, 2022. Informed consent was voluntarily given by all participants prior to their participation in the study.

## Declaration of competing interest

ED, MP, SR, AW, AR, and VF-H are employees of Woebot Health.

## Acknowledgements

The authors wish to thank Robbert Zusterzeel, M.D., Ph.D., MPH, Tim Campellone, Ph.D., Alison Darcy, Ph.D. for their helpful comments and Adam Platt, MSc for technical support in the preparation of this manuscript submission.

## References

- Abd-alrazaq, A.A., Alajlani, M., Alalwan, A.A., Bewick, B.M., Gardner, P., Househ, M., 2019. An overview of the features of chatbots in mental health: a scoping review. *Int. J. Med. Inf.* 132, 103978 <https://doi.org/10.1016/j.ijmedinf.2019.103978>.
- Abd-Alrazaq, A.A., Rababeh, A., Alajlani, M., Bewick, B.M., Househ, M., 2020. Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. *J. Med. Internet Res.* 22 (7), e16021 <https://doi.org/10.2196/16021>.
- Ahola, K., Toppinen-Tanner, S., Seppänen, J., 2017. Interventions to alleviate burnout symptoms and to support return to work among employees with burnout: systematic review and meta-analysis. *Burn. Res.* 4, 1–11. <https://doi.org/10.1016/j.burn.2017.02.001>.
- Baumel, A., Muench, F., Edan, S., Kane, J.M., 2019. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. *J. Med. Internet Res.* 21 (9), e14567 <https://doi.org/10.2196/14567>.
- Boucher, E.M., Harake, N.R., Ward, H.E., et al., 2021. Artificially intelligent chatbots in digital mental health interventions: a review. *Expert Rev. Med. Devices* 18 (sup1), 37–49. <https://doi.org/10.1080/17434440.2021.2013200>.
- Bureau of Health Workforce, Health Resources and Services Administration (HRSA), 2022. U.S. Department of Health & human services, designated health professional shortage areas statistics: designated HPSA quarterly summary, as of October 25, 2022. Published October 25. Accessed October 27, 2022. <https://data.hrsa.gov/topics/health-workforce/shortage-areas>.
- Cohen, S., Kamarck, T., Mermelstein, R., 1983. A global measure of perceived stress. *J. Health Soc. Behav.* 24 (4), 385. <https://doi.org/10.2307/2136404>.
- Colodro Conde, L., Couvy-Duchesne, B., Zhu, G., et al., 2019. A direct test of the diathesis-stress model for depression. *Eur. Neuropsychopharmacol.* 29, S805–S806. <https://doi.org/10.1016/j.euroneuro.2017.08.045>.
- Connor, K.M., Zhang, W., 2006. Resilience: determinants, measurement, and treatment responsiveness. *CNS Spectr.* 11 (S12), 5–12. <https://doi.org/10.1017/S1092852900025797>.
- Cooke, J.E., Eirich, R., Racine, N., Madigan, S., 2020. Prevalence of posttraumatic and general psychological stress during COVID-19: a rapid review and meta-analysis. *Psychiatry Res.* 292, 113347 <https://doi.org/10.1016/j.psychres.2020.113347>.
- Darcy, A., Daniels, J., Salinger, D., Wicks, P., Robinson, A., 2021. Evidence of human-level bonds established with a digital conversational agent: cross-sectional, retrospective observational study. *JMIR Form. Res.* 5 (5), e27868 <https://doi.org/10.2196/27868>.
- Dolan, E.D., Mohr, D., Lempa, M., et al., 2015. Using a single item to measure burnout in primary care staff: a psychometric evaluation. *J. Gen. Intern. Med.* 30 (5), 582–587. <https://doi.org/10.1007/s11606-014-3112-6>.
- Firth, J., Torous, J., Nicholas, J., et al., 2017a. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. *World Psychiatry* 16 (3), 287–298. <https://doi.org/10.1002/wps.20472>.
- Firth, J., Torous, J., Nicholas, J., Carney, R., Rosenbaum, S., Sarris, J., 2017b. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. *J. Affect. Disord.* 218, 15–22. <https://doi.org/10.1016/j.jad.2017.04.046>.
- Fitzpatrick, K.K., Darcy, A., Vierhile, M., 2017. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. *JMIR Ment. Health.* 4 (2), e19 <https://doi.org/10.2196/mental.7785>.
- Harris, M.L., Oldmeadow, C., Hure, A., Luu, J., Loxton, D., Attia, J., 2017. Stress increases the risk of type 2 diabetes onset in women: a 12-year longitudinal study using causal modelling. *PLoS One* 12 (2), e0172126. <https://doi.org/10.1371/journal.pone.0172126>.
- Hassard, J., Teoh, K.R.H., Visockaite, G., Dewe, P., Cox, T., 2018. The cost of work-related stress to society: a systematic review. *J. Occup. Health Psychol.* 23 (1), 1–17. <https://doi.org/10.1037/ocp0000069>.
- Hofmann, S.G., Asnaani, A., Vonk, I.J.J., Sawyer, A.T., Fang, A., 2012. The efficacy of cognitive behavioral therapy: a review of meta-analyses. *Cogn. Ther. Res.* 36 (5), 427–440. <https://doi.org/10.1007/s10608-012-9476-1>.
- Hu, T., Zhang, D., Wang, J., 2015. A meta-analysis of the trait resilience and mental health. *Personal. Individ. Differ.* 76, 18–27. <https://doi.org/10.1016/j.paid.2014.11.039>.
- Joyce, S., Shand, F., Tighe, J., Laurent, S.J., Bryant, R.A., Harvey, S.B., 2018. Road to resilience: a systematic review and meta-analysis of resilience training programmes and interventions. *BMJ Open* 8 (6). <https://doi.org/10.1136/bmjopen-2017-017858>.
- Kroenke, K., Strine, T.W., Spitzer, R.L., Williams, J.B.W., Berry, J.T., Mokdad, A.H., 2009. The PHQ-8 as a measure of current depression in the general population. *J. Affect. Disord.* 114 (1–3), 163–173. <https://doi.org/10.1016/j.jad.2008.06.026>.
- Lau, N., O'Daffer, A., Colt, S., et al., 2020. Android and iPhone mobile apps for psychosocial wellness and stress management: systematic search in app stores and literature review. *JMIR MHealth UHealth.* 8 (5), e17798 <https://doi.org/10.2196/17798>.
- Lee, E.H., 2012. Review of the psychometric evidence of the perceived stress scale. *Asian Nurs. Res.* 6 (4), 121–127. <https://doi.org/10.1016/j.anr.2012.08.004>.
- Lim, S.M., Shiao, C.W.C., Cheng, L.J., Lau, Y., 2022. Chatbot-delivered psychotherapy for adults with depressive and anxiety symptoms: a systematic review and meta-regression. *Behav. Ther.* 53 (2), 334–347. <https://doi.org/10.1016/j.beth.2021.09.007>.
- Linardon, J., Cuijpers, P., Carlbring, P., Messer, M., Fuller-Tyszkiewicz, M., 2019. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. *World Psychiatry* 18 (3), 325–336. <https://doi.org/10.1002/wps.20673>.
- Ly, K.H., Ly, A.M., Andersson, G., 2017. A fully automated conversational agent for promoting mental well-being: a pilot RCT using mixed methods. *Internet Interv.* 10, 39–46. <https://doi.org/10.1016/j.invent.2017.10.002>.
- Martin, D.J., Garske, J.P., Davis, M.K., 2000. Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review. *J. Consult. Clin. Psychol.* 68 (3), 438–450.
- Maslach, C., Jackson, S., Leiter, M., 1997. *The Maslach burnout inventory manual*. In: *Third Edition, Zalaquett, C.P., Wood, J.R. (Eds.), Evaluating Stress: A Book of Resources*. Scarecrow Education, pp. 191–218.
- Prochaska, J.J., Vogel, E.A., Chieng, A., et al., 2021a. A therapeutic relational agent for reducing problematic substance use (Woebot): development and usability study. *J. Med. Internet Res.* 23 (3), e24850 <https://doi.org/10.2196/24850>.
- Prochaska, J.J., Vogel, E.A., Chieng, A., et al., 2021b. A randomized controlled trial of a therapeutic relational agent for reducing substance misuse during the COVID-19 pandemic. *Drug Alcohol Depend.* 227, 108986 <https://doi.org/10.1016/j.drugalcdep.2021.108986>.
- R Core Team, 2021. R: a language and environment for statistical computing. Published online. <https://www.r-project.org/>.
- Ramachandran M, Suharwardy S, Leonard S, et al. Acceptability of postpartum mood management through a smartphone-based automated conversational agent. Presented at: 40th Annual Pregnancy Meeting of the Society for Fetal and Maternal Medicine; February 2020; Grapevine, Texas.
- Richardson, S., Shaffer, J.A., Falzon, L., Krupka, D., Davidson, K.W., Edmondson, D., 2012. Meta-analysis of perceived stress and its association with incident coronary heart disease. *Am. J. Cardiol.* 110 (12), 1711–1716. <https://doi.org/10.1016/j.amjcard.2012.08.004>.
- Rutten, B.P.F., Hammels, C., Geschwind, N., et al., 2013. Resilience in mental health: linking psychological and neurobiological perspectives. *Acta Psychiatr. Scand.* 128 (1), 3–20. <https://doi.org/10.1111/acps.12095>.

- Salvagioni, D.A.J., Melanda, F.N., Mesas, A.E., González, A.D., Gabani, F.L., Andrade, S. M. de, 2017. Physical, psychological and occupational consequences of job burnout: a systematic review of prospective studies. *PLoS One* 12 (10), e0185781. <https://doi.org/10.1371/journal.pone.0185781>.
- Smith, B.W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., Bernard, J., 2008. The brief resilience scale: assessing the ability to bounce back. *Int. J. Behav. Med.* 15 (3), 194–200.
- Song, H., Fang, F., Tomasson, G., et al., 2018. Association of stress-related disorders with subsequent autoimmune disease. *JAMA*. 319 (23), 2388. <https://doi.org/10.1001/jama.2018.7028>.
- Southwick, S.M., Charney, D.S., 2012. The science of resilience: implications for the prevention and treatment of depression. *Science* 338 (6103), 79–82. <https://doi.org/10.1126/science.1222942>.
- Southwick, S.M., Litz, B.T., Charney, D., Friedman, J.M. (Eds.), 2011. *Resilience and Mental Health: Challenges across the Lifespan*. Cambridge University Press.
- Spitzer, R.L., Kroenke, K., Williams, J.B.W., Löwe, B., 2006. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch. Intern. Med.* 166 (10), 1092–1097. <https://doi.org/10.1001/archinte.166.10.1092>.
- Suharwardy S, Ramachandran M, Leonard S, et al. Effect of an automated conversational agent on postpartum mental health: a randomized, controlled trial. Poster presented at: 40th Annual Pregnancy Meeting of the Society for Fetal and Maternal Medicine; February 2020; Grapevine, Texas.
- Waller, R., Gilbody, S., 2009. Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence. *Psychol. Med.* 39 (5), 705–712. <https://doi.org/10.1017/S0033291708004224>.
- West, C.P., Dyrbye, L.N., Sloan, J.A., Shanafelt, T.D., 2009. Single item measures of emotional exhaustion and depersonalization are useful for assessing burnout in medical professionals. *J. Gen. Intern. Med.* 24 (12), 1318–1321. <https://doi.org/10.1007/s11606-009-1129-z>.
- World Health Organization (WHO), 2019. Burn-out an “occupational phenomenon”: international classification of diseases. Published May 28. Accessed May 12, 2023. <https://www.who.int/news/item/28-05-2019-burn-out-an-occupational-phenomenon-international-classification-of-diseases>.
- World Health Organization (WHO), 2023. Stress. Published February 21. Accessed May 12, 2023. <https://www.who.int/news-room/questions-and-answers/item/stress>.