188 (2026) 209993

Contents lists available at ScienceDirect

Journal of Substance Use and Addiction Treatment

journal homepage: www.journals.elsevier.com/journal-of-substance-use-and-addiction-treatment

A relational agent for treating substance use in adults: A randomized controlled trial with a psychoeducational comparator

Judith J. Prochaska [a,*], Maggie Wang [c], Amy Chieng [a], Jessie Moore [a], Erin A. Vogel [b], Michael Baiocchi [d], Sarah Pajarito [e], Maddison C. Pirner [e], Lauren Bullard [e], Alison Darcy [e], Athena Robinson [e]

[a] Stanford Prevention Research Center, Department of Medicine, School of Medicine, Stanford University, United States
[b] TSET Health Promotion Research Center, University of Oklahoma Health Sciences Center, United States
[c] Department of Epidemiology & Population Health, School of Medicine, Stanford University, United States
[d] Department of Biomedical Data Science, Stanford University, United States
[e] Woebot Health, United States

ARTICLE INFO

Keywords:
Digital mental health intervention
Artificial intelligence
Substance-related disorders
Randomized controlled trial
Relational conversational agent
Mobile health
Alcohol

ABSTRACT

Background: Highly accessible and scalable, digital mental health interventions can reduce barriers associated with traditional treatment. Woebot for Substance Use Disorders (W-SUD) is a smartphone application in development that uses *Woebot*, a guided self-help relational agent intended to offer support to track and reduce problematic substance use. Two formative studies supported W-SUD's feasibility, acceptability, and efficacy to reduce substance use at end-of-treatment (EOT).

Objectives: In a randomized trial with a 1-month post-treatment follow-up, we aimed to evaluate W-SUD's efficacy in reducing substance use occasions relative to a psychoeducational control.

Methods: U.S. adults ($N = 258$) with problematic substance use (CAGE-AID≥2) were recruited online and randomized to W-SUD or an email-delivered psychoeducational control. Primary and secondary outcomes were change in past-month substance use occasions from baseline to 8-weeks EOT and baseline to 1-month follow-up, respectively.

Results: The analytic sample ($N = 202$; 107 = W-SUD; 95 = psychoeducation) averaged 38.3 years of age (SD = 10.4) and was 53.5% (108/202) female and 71.3% (144/202) White. At baseline, participants averaged 33.1 (SD = 18.3) past-month substance use occasions with problematic substance use reported as alcohol (73.3%, 148/202), cannabis (30.7%, 62/202), stimulants (17.3%, 35/202), and cocaine (12.4%, 25/202). Anxiety (32.7%, 66/202) or depression (23.3%, 47/202) co-occurred. W-SUD participants sent a median of 322.5 (interquartile range: 59.8, 924.0) in-app messages during the intervention. There was no significant treatment effect on the primary (beta = −0.675, $p = 0.741$; 95%CI = -4.96–3.49) or secondary outcomes. Baseline to EOT within-group change in past-month substance use occasions were −13.6 (SD = 15.4, d = −0.879) and −12.9 (SD = 15.3, d = −0.847) for the W-SUD and psychoeducation conditions, respectively, indicating moderate to large changes in both, which was sustained at 1-month follow-up. Decreases in substance use occasions significantly correlated with decreases in anxiety and depression and with greater treatment satisfaction in both groups. There were no serious adverse events.

Conclusions: Participants in both conditions reported significant improvements in primary and secondary outcomes at EOT and 1-month post-treatment, with no statistically significant difference in improvement between conditions.

Clinicaltrials.gov (posted 2021-06-14): NCT04925570, https://clinicaltrials.gov/study/NCT04925570

* Corresponding author at: Stanford University, Stanford Prevention Research Center, 3180 Porter Drive, Room A105, Palo Alto, CA, 94304-1212, United States.
E-mail address: jpro@stanford.edu (J.J. Prochaska).

https://doi.org/10.1016/j.josat.2026.209993
Received 22 August 2025; Received in revised form 27 February 2026; Accepted 22 April 2026
Available online 24 April 2026

Abbreviations

ANOVA	Analysis of Variance statistical test to analyze differences between multiple group means
BSCQ	Brief Situational Confidence Questionnaire
CAGE-AID	Measure of problematic substance use; CAGE is an acronym from main parts of the four questions that comprise the measure (Cut down, Annoyed, Guilty, Eye-opener); AID refers to Adapted to Include Drugs, as the original CAGE assessed alcohol use
CSQ-8	Client Satisfaction Questionnaire – 8 questions
DAST-10	Drug Abuse Screening Test – 10 questions
DSMB	Data Safety Monitoring Board
EOT	End-of-treatment
GAD-7	Generalized Anxiety Disorder – 7 questions
NIDA	National Institute on Drug Abuse
PHQ-8	Patient Health Questionnaire – 8 questions
SIP-AD	Short Inventory of Problems – Alcohol and Drugs
SPS-6	Stanford Presenteeism Scale – 6 questions
SUDs	Substance Use Disorders
URP-I	User Rating Profile – Intervention
US	United States
WAI-SR	Working Alliance Inventory – Short Revised
W-SUD	*Woebot* for Substance Use Disorders

1. Introduction

Substance use disorders (SUDs) are a pervasive and global public health concern (Grodniewicz & Hohol, 2023; Rehm & Shield, 2019; Steel et al., 2014; Xiong et al., 2020). According to the 2023 National Survey on Drug Use and Health, 48.46 million people in the U.S. met criteria for a SUD in the past year: 28.86 million for alcohol use disorder, 27.15 million for a drug use disorder, and 7.55 million for both; yet, just over 1 in 4 (26.38%) had received treatment (Substance Abuse and Mental Health Services Administration & Center for Behavioral Health Statistics and Quality, 2024). Reaching and engaging people in SUD treatment requires the development of broadly accessible, affordable, and high-quality evidence-based programs (Grodniewicz & Hohol, 2023; Patel et al., 2018).

The 2016 U.S. Surgeon General report on "Facing Addiction in America" called for enhanced public education to improve awareness about substance use problems and improved access to evidence-based treatment services (U.S. Department of Health and Human Services (HHS) & Office of the Surgeon General, 2016). The report identified the many potential advantages of technology-based interventions for treating SUDs, including increased access in underserved areas; reduced costs; offsetting time demands of service providers, thereby enabling them to care for more patients; providing alternatives for people hesitant to engage in traditional in-person treatment; and systematic delivery of treatments as designed. Many health care providers remain reluctant to address addiction in their practice, citing institutional barriers (e.g., lack of support, insufficient resources, competing demands), insufficient training, and fear of harming clinical rapport (Campopiano von Klimo et al., 2024; Magnan et al., 2024).

Digital mental health interventions (DMHIs) are tools and resources that utilize digital technologies to address mental health concerns. DMHIs range from self-guided applications to programs augmented with structured receipt of human-delivered professional support. DMHIs have tremendous potential as a scalable modality to address SUDs. DMHIs can reduce barriers associated with traditional treatment, and an expanding evidence base supports DMHIs for the treatment of a variety of mental health conditions (Andrews et al., 2018; Bower et al., 2013; Firth, Torous, Nicholas, Carney, Pratap, et al., 2017; Firth, Torous, Nicholas, Carney, Rosenbaum, & Sarris, 2017; Linardon et al., 2019; Sawyer-Morris et al., 2024; Wright et al., 2019), including SUDs (Boumparis et al., 2019; Maricich, Gerwien, et al., 2021; Maricich et al., 2022; Maricich, Xiong, et al., 2021; Prochaska, Vogel, Chieng, Baiocchi, et al., 2021; Prochaska, Vogel, Chieng, Kendra, et al., 2021; Sawyer-Morris et al., 2024; Xiong et al., 2023). Recently reported findings indicate that DMHIs for SUDs are associated with significantly reduced healthcare costs and acute care resource utilization (Sawyer-Morris et al., 2024; Shah et al., 2022; Velez et al., 2022; Velez, Colman, et al., 2021; Velez, Ruetsch, & Maricich, 2021). Despite their notable benefits, people do not consistently engage with DMHIs and when they do, clinical outcomes do not always improve as expected (Donkin et al., 2011). For example, while some research linked greater DMHI engagement to improved SUD clinical outcomes (Bakker & Rickard, 2018; Garnett et al., 2018), other studies reported opposite findings, associations that vary by engagement cluster type (Graham et al., 2021; Hoffman et al., 2023; Yardley et al., 2016), or no significant relationship at all (Graham et al., 2020; Graham et al., 2021).

Engagement remains one of the central challenges in digital mental health, with most apps losing the majority of their users within days of installation (Baumel et al., 2019; Torous et al., 2018). Yet meta-analytic evidence suggests that conversational, text-based agents may buck this trend by increasing both engagement and enjoyment (Vaidyam et al., 2019). Animated, screen-based relational agents can provide in-the-moment support and deliver treatment in ways modeled after the core principles of effective human counseling, including rapport-building, trust, and therapeutic alliance (Bickmore & Gruber, 2010). Research suggests people may actually be more willing to disclose personal information when they believe they are interacting with a computer rather than a human observer (Lucas et al., 2014). A strong therapeutic alliance, it turns out, does not require face-to-face contact; it can develop through in-app interactions and conversational agents alike (Beatty et al., 2022; Darcy et al., 2021). Early research also indicates that text-based delivery tends to produce stronger program adherence than verbal presentation (Tielman et al., 2017).

The current evaluation is the third in a series of studies investigating Woebot for Substance Use Disorders (W-SUD) for evidence based treatment (Prochaska, Vogel, Chieng, Baiocchi, et al., 2021; Prochaska, Vogel, Chieng, Kendra, et al., 2021). W-SUD is a smartphone-based 8-week DMHI that encourages mood tracking, provides behavioral pattern insight, and guides users through behavior change tools rooted in cognitive behavioral therapy (CBT) techniques (Fig. 1 shows screenshots of the W-SUD app). To deliver these interventional tools, W-SUD utilizes a guided self-help relational agent ("Woebot") that is intended to present to users a friendly, helpful, self-help ally that is explicitly not a human or a therapist. W-SUD was previously evaluated in a single-arm pilot study followed by a two-arm randomized controlled trial (RCT) comparing the efficacy of W-SUD relative to a waitlist control condition (Prochaska, Vogel, Chieng, Baiocchi, et al., 2021; Prochaska, Vogel, Chieng, Kendra, et al., 2021). W-SUD was found to be feasible, acceptable, engaging, and associated with significant reductions in past-month substance use occasions and with improvement in other substance use related outcomes. Depression and anxiety symptoms among W-SUD users significantly declined. Relative to participants randomized to a waitlist control, participants receiving W-SUD reported significantly reduced past-month substance use occasions.

A systematic review found that most online interventions targeting problematic substance use produced significant short-term improvement on at least one measure of problematic substance use (Xiong et al., 2020). Longer-term follow-up evaluations are needed and with attention to multiple outcomes of interest including substance use, substance use problems and productivity effects, and factors associated with relapse (e.g., ongoing cravings and urges to use). The current two-arm RCT sought to expand on prior evidence by examining the efficacy of W-SUD relative to an attention control condition (email-delivered psychoeducation) and by evaluating the sustainability of treatment effects at 1-month post-treatment (Prochaska et al., 2023).

Psychoeducation, or providing information and education on mental health (Magill et al., 2021), is an established control condition in

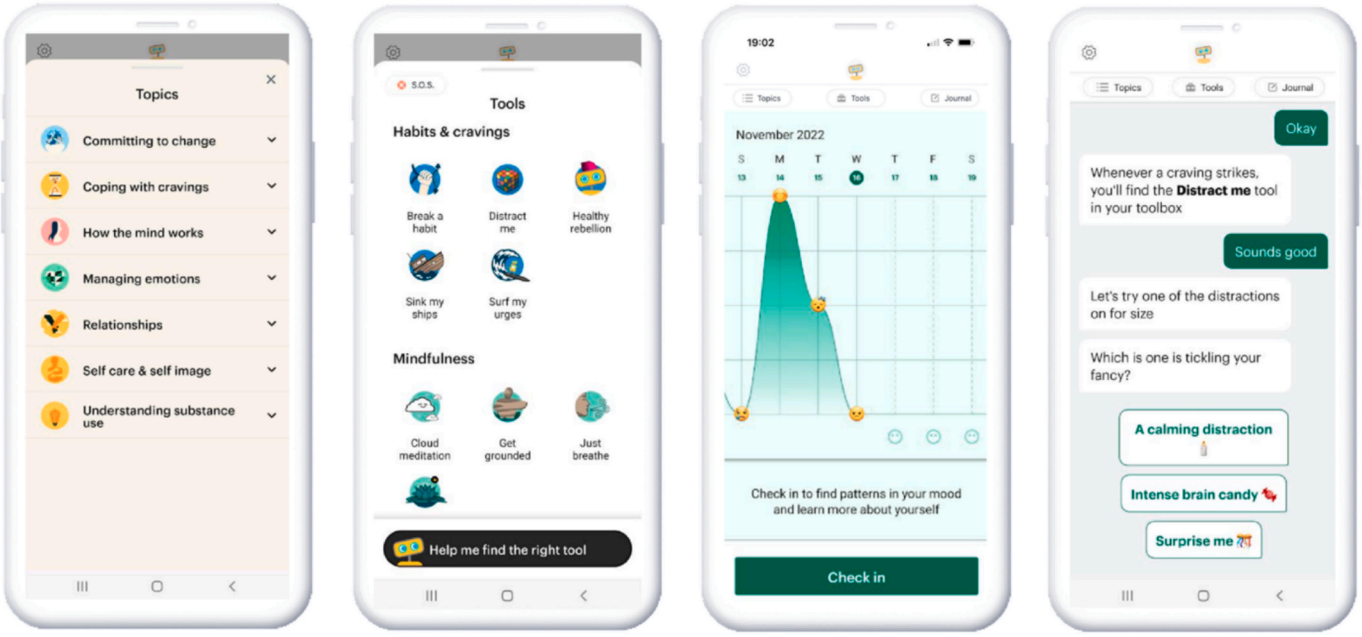


Fig. 1. Sample screen shots of W-SUD.

substance use treatment research (Brendryen & Kraft, 2008; Chiauzzi et al., 2005; Collins et al., 2002; Deady et al., 2016; Doumas et al., 2009; Doumas & Haustveit, 2008; Lau-Barraco & Dunn, 2008; Mullins et al., 2004; Naughton et al., 2012; Parsons et al., 2014; Schaub et al., 2012; Stein et al., 2018; Ybarra et al., 2013). With variability, the impact of psychoeducational controls on substance use outcomes appears to be negligible (Cohen's d / hedges' $g < 0.2$) to moderate (Cohen's d / hedges' $g \geq 0.5$ to <0.8) at primary endpoint assessments (Batterham et al., 2018; Deady et al., 2016; Doumas et al., 2009; Doumas & Haustveit, 2008; Lau-Barraco & Dunn, 2008), with some studies reporting continued improvements at follow-up (Batterham et al., 2018; Deady et al., 2016; Schaub et al., 2012).

Building upon and extending our prior evaluations, we hypothesized that U.S. adults recruited online and reporting problematic substance use who were randomized to W-SUD for 8 weeks of treatment would significantly reduce their substance use by the end of treatment (EOT) and sustain this reduction out to 1-month follow-up. The primary measure was reduction in substance use occasions as described in prior studies (Prochaska, Vogel, Chieng, Baiocchi, et al., 2021; Prochaska, Vogel, Chieng, Kendra, et al., 2021). We further hypothesized that reductions observed with W-SUD treatment would outperform the effects of emailed psychoeducational materials sent for 8 weeks.

2. Materials and methods

2.1. Study design and procedures

In a 2-group randomized controlled trial described in detail previously (Prochaska et al., 2023), we evaluated W-SUD relative to a psychoeducation control condition. Both study conditions were 8 weeks in duration, with access to intervention content discontinued after week 8 to permit a 1-month follow-up evaluation post-EOT. To prevent participant bias or differential attrition, each randomized participant received a private link to a page that described their group assignment without identifying it as the experimental or comparison condition. At study end (week 12), participants received a full description of the two groups.

Follow-up assessments were at 8 weeks (EOT) and 12 weeks (1-month follow-up). The primary outcome was the change from baseline to EOT in the past-month number of substance use occasions, summed across all substances (i.e., response at EOT minus response at baseline). Substance use occasions is the sum of the reported days of use, over the past 30 days, across substances. A secondary outcome was the change in substance use occasions from baseline to 1-month follow-up. Additional secondary outcomes were changes (baseline to EOT and baseline to 1-month follow-up) in: the percent of days abstinent from all substances, number of heavy drinking days, substance use problems, thoughts about abstinence, craving intensity, situational confidence to resist substance use urges, mood symptoms (depression, anxiety), and work productivity. Acceptability was assessed within the W-SUD program and in the EOT survey. Engagement data were collected for both conditions. Stanford Medicine's Institutional Review Board approved study procedures (protocol 58725).

The trial was fully remote. Decentralized clinical trials have grown in popularity over the past decade, accelerated by the Covid-19 pandemic (McCarthy et al., 2024; Petrini et al., 2022; Van Norman, 2021). Decentralized clinical trials offer several advantages, including increasing accessibility, sample diversity, and representativeness, while reducing participant burden and enhancing operational flexibility (Guest et al., 2021; McCarthy et al., 2024; Mitchell et al., 2020; Petrini et al., 2022; Teitcher et al., 2015; Underhill et al., 2024; Van Norman, 2021). However, methodological challenges, such as fraudulent efforts at enrollment and participation solely for research compensation, pose a threat to data quality and validity (Guest et al., 2021; Mitchell et al., 2020; Teitcher et al., 2015). Our trial undertook a number of proactive steps at recruitment and screening and reactive steps as needed post-randomization to mitigate fraudulent efforts to participate, described elsewhere (Moore et al., 2025) and detailed in the study's Consort diagram (Fig. 2).

2.2. Study conditions

2.2.1. W-SUD

W-SUD is an investigational use only DMHI utilizing *Woebot*, a guided self-help relational agent intended to offer support to track and reduce problematic substance use. W-SUD is accessible via its own native applications on iPhone and Android smartphones (Prochaska, Vogel, Chieng, Kendra, et al., 2021). Within W-SUD, *Woebot* is presented as a friendly, helpful character that is explicitly not a human, nor a therapist, but rather a guided self-help ally. Each text-based conversation with Woebot is crafted to users' reported concerns, moods, and/or

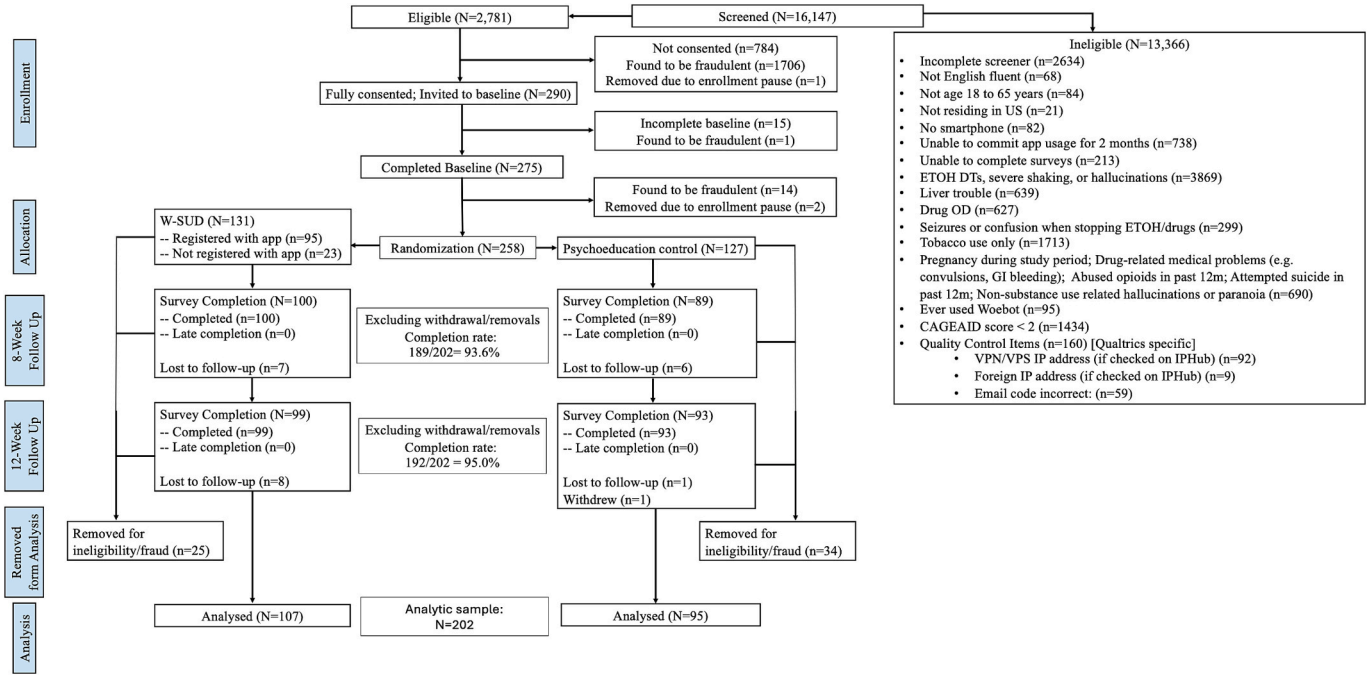


Fig. 2. Study CONSORT Diagram.

readiness to regulate their behaviors and emotions and use skills for change in the moment. Woebot delivers empathetic responses tailored to user-reported elements and offers to guide them through the in-vivo practice and application of a wide array of psychotherapeutic tools for change including cognitive techniques (e.g., thought challenging), mindfulness exercises, gratitude journaling, and mood tracking. Such techniques are primarily derived from cognitive behavioral therapy and include infused elements from motivational interviewing and dialectical behavior therapy. Through push notifications, W-SUD invites daily check-ins. W-SUD includes mood and substance use craving tracking, psychoeducational lessons on substance use (e.g., substance use impacts upon the brain, sleep, and mood; coping with cravings), and behavior change and emotion management skills derived from principles of psychotherapy. Outside of push-notifications designed to prompt optional daily check-ins, app usage is completely self-determined, and participants randomized to the intervention arm could access W-SUD at any time for the first 8 weeks of the study. The app collects counts of active days, messages sent and received, tools and stories started and completed, moods logged, and craving ratings entered.

W-SUD's onboarding explains the product's intended use, how data are treated, and service limitations that emphasize that W-SUD is not a crisis service. If crisis language is detected via the app's natural language processing algorithms, and the user confirms, W-SUD offers helpline resources (e.g., 9–1-1, suicide crisis hotlines), which also are readily available within the app navigation.

2.2.2. *Email-delivered psychoeducational control*

A psychoeducational control condition was created with information on: (a) alcohol-specific topics; (b) drug-specific topics; (c) general addiction topics; and (d) statistics relating to alcohol and substance use. The information came from the National Institute on Drug Abuse, National Institute on Alcohol Abuse and Alcoholism, and the Centers for Disease Control and Prevention. Psychoeducation is commonly provided to people with substance use problems and is intended to increase knowledge of their substances of use, including effects on the body and mind, and consequences (Magill et al., 2021). Once a week, by email, participants received a link to open a factsheet on REDCap Cloud, a secure survey platform. The link was active throughout the 8-week study period. To determine receipt, participants who opened the factsheets were asked to confirm they had finished reading by pressing the submit button at the bottom of the webpage.

2.3. *Sample recruitment, randomization, & retention*

U.S. adult residents (aged 18–65) concerned about their substance use were recruited online. Inclusion criteria were any alcohol or drug use in the past 30-days, endorsing two or more items on the CAGE-AID, owning a WiFi- or data-enabled smartphone, English literacy, and being available to participate for the duration of the 12-week study. An upper age limit of 65 years was set due to interest in assessing treatment effects on work productivity. Study exclusion criteria were prior Woebot use, current or anticipated pregnancy during the study period, tobacco use as the sole substance of concern, history of severe alcohol or drug-related medical problems (e.g., delirium tremens, seizure, liver disease, gastrointestinal bleeding, hallucinations), history of opioid overdose requiring Narcan (naloxone), current opioid misuse without medication-assisted treatment, and past-year attempted suicide. The latter four exclusions were for risk management considerations as W-SUD is not a crisis service and does not detect or treat medical emergencies. SUD treatment referrals were provided to ineligible individuals. Co-occurring psychiatric disorders were not an exclusion.

The effect size for the primary outcome of substance use occasions at EOT in a previous randomized controlled trial with a waitlist comparison was Cohen's d = 0.36 (Vogel et al., 2021). In the current trial, with an active comparator, we anticipated a slightly smaller effect (Cohen's d = 0.32). Based on power calculations with a two-sided rejection region, alpha level of 0.05, power of 80%, projected effect size of Cohen's d = 0.32 at EOT, and 20% attrition, the target sample size was $N = 194$ per group or about 400 total participants.

REDCap Cloud, a secure survey and database web-application, was used for web-based screening, collecting Part 11 compliant consent, randomization, and study assessments. Recruitment campaigns promoted the study via paid ads on social media (i.e., Instagram, Facebook); however, the effort attracted a high volume of fraudulent efforts to participate (Moore et al., 2025). Hence, we pivoted recruitment to Qualtrics Research Services survey panels and utilized Qualtrics' secure

survey platform for web-based screening only. Detection of fraudulent efforts continued post-randomization and included checking for inconsistency in birthdates provided over time, multiple screening attempts and duplicate enrollment, prior Woebot use, and registering on the W-SUD app though assigned to the psychoeducation group (i.e., likely a duplicate enrollment).

Randomization followed completion of the baseline assessment using permuted blocks (Broglio, 2018) programmed into REDCap Cloud based on CAGE-AID scores (2, 3, 4) and prior therapy experience (yes/no). The allocation ratio was 1:1, and each block size had a minimum of 4 and maximum of 6. Participants provided their phone numbers and received compensation up to US$100 for assessment completion (US$25 e-gift cards per assessment, excluding screening) to maximize study retention. A prior randomized trial of W-SUD utilizing similar strategies had 84% retention at EOT (Prochaska, Vogel, Chieng, Baiocchi, et al., 2021).

2.4. Assessments

Participants completed demographic, substance use, and mental health measures at screening and baseline; substance use and mental health measures again at EOT and 1-month post-EOT follow-up; and W-SUD feasibility and acceptability measures at EOT. Research staff tracked serious adverse events at EOT and 1-month follow-up, and collected W-SUD app usage data and factsheet "viewed" receipts throughout the 8-week intervention period (Prochaska et al., 2023).

2.4.1. Demographic characteristics

Included age; gender identity; sexual orientation; race and ethnicity; education, and marital, employment, disability, and insurance statuses.

2.4.2. Substance use measures

Included frequency of use, quantity of use for alcohol, problematic substance(s), substance use problems or consequences, craving and confidence to resist urges, and thoughts about abstinence.

The primary outcome of interest was substance use occasions reported for the past 30 days. Substances assessed were alcohol, cannabis, cocaine, prescription stimulants, methamphetamine, inhalants, sedatives, hallucinogens, street opioids, and prescription opioids. To capture the different substances used, days of reported alcohol use and days of use for each assessed drug was summed to reflect past-month substance use occasions. Past-month substance use occasions could exceed 30 if multiple substances were used in a day. This measure proved sensitive to treatment effects in earlier studies of W-SUD (Prochaska, Vogel, Chieng, Baiocchi, et al., 2021; Prochaska, Vogel, Chieng, Kendra, et al., 2021). For the current study, past 30-day tobacco use also was assessed, though not included in the calculation of substance use occasions, because W-SUD was not designed to target tobacco use.

A secondary measure of past 30-day substance use occasions was limited to one's self-reported problematic substance(s), identified as their primary, secondary, or tertiary problematic substance (other than tobacco). Days of heavy drinking and average drinks per week were measured using the Quick Drinking Screen (QDS) (Sobell et al., 2003). The QDS defines heavy drinking as having 5 or more standard drinks (for men) or 4 or more standard drinks (for women). Average drinks per week was calculated as the product of the responses to average drinks per day and the average days of drinking per week.

Substance use problems were measured using the 15-item Short Inventory of Problems—Alcohol and Drugs (SIP-AD) (possible range 0–45), which assesses substance use problems in the past 30 days (Blanchard et al., 2003). The 10-item Drug Abuse Screening Test-10 (DAST-10) (possible range 0–10) (Skinner, 1982) assessed drug-related consequences in the past 30 days, excluding alcohol and tobacco. A total score of 3 or more on the DAST-10 indicates significant drug abuse problems. The DAST-10 item concerning drug-related medical problems was a screening exclusion criterion; hence, the sample's total possible range at baseline was 0–9. Additionally, participants' perceptions of the effects of substance use on their work productivity in the past 2 weeks was measured using the Stanford Presenteeism Scale (SPS-6) (Koopman et al., 2002), a 6-item scale with response options ranging from strongly disagree (1) to strongly agree (6).

Craving of substances was assessed with a single item asking, "In the past 7 days, how much were you bothered by cravings or urges to drink alcohol or use drugs?" with response options of not at all (0), a little bit (1), moderately (2), quite a bit (3), and extremely (4). Confidence to resist the urge to use was measured using the Brief Situational Confidence Questionnaire (BSCQ) (Breslin et al., 2000), a state-dependent measure. The BSCQ assesses self-confidence to resist the urge "right now" to drink heavily (self-defined) or use drugs in different situations. Self-confidence is reported on visual analog scales (100-mm lines) anchored from 0% "not at all confident" to 100% "totally confident."

Thoughts about abstinence (TAA) were measured with 3 continuous items all with possible range of 1–10: the desire to abstain from all substances, the expectation of successfully achieving abstinence, and the perceived difficulty of avoiding relapse (Hall et al., 1990).

Engagement in mental health or substance use treatment was assessed for the past 2 months at the end-of-treatment and for the past 30 days at baseline and the 1-month follow-up. Forms of treatment included: self-help groups/meetings (AA/NA/12 Step); outpatient services (group or individual); intensive outpatient or partial hospitalization; residential/inpatient (e.g. hospital program, inpatient detoxification); or other. Follow up questions assessed the extent of attention to substances (from 0 = none at all to 10 = great extent); duration of treatment; medication use to help with reducing substance use (e.g. Antabuse, naltrexone, buprenorphine/Suboxone, methadone); and treatment satisfaction.

2.4.3. Mental health measures

Included diagnoses, treatment history, and active symptomatology. Ten lifetime psychiatric diagnoses were assessed plus a write-in option for others. A single item assessed current use of prescribed medications for a psychiatric diagnosis. Prior therapy engagement was asked as "have you ever been seen by a therapist for mental health or substance use concerns?" with affirmative responses leading to additional questions about treatment type. Depression was measured using the 8-item Patient Health Questionnaire (PHQ-8) (Kroenke et al., 2009). Anxiety was measured using the 7-item Generalized Anxiety Disorder (GAD-7) (Spitzer et al., 2006).

2.4.4. Feasibility and acceptability

Were assessed with the Usage Rating Profile-Intervention (URP—I) Feasibility (6 items) and Acceptability (6 items) scales (Briesch et al., 2013), the 8-item Client Satisfaction Questionnaire (CSQ-8) (Larsen et al., 1979), and the 12-item Working Alliance Inventory-Short Revised (WAI-SR) (Hatcher & Gillaspy, 2006). Higher scores indicate greater feasibility and acceptability and a better therapeutic alliance. The WAI-SR consists of three 4-item subscales, with 5-point rating scales, that reflect development in treatment of an affective bond and level of agreement with treatment goals and treatment tasks. URP-I item response options range from strongly disagree to strongly agree; the items are summed for a total score within each scale, with one feasibility item reverse coded. The CSQ-8 items have 4-point rating scales with response descriptors that vary (internal consistency >0.90); the total sum score ranges from 8 to 32, with higher total scores indicating higher satisfaction.

2.4.5. Serious adverse events

Occurring since study start were assessed for hospitalization related to substance use, suicide attempt, alcohol or drug overdose, and severe withdrawal (e.g., delirium tremens).

2.4.6. Measures of treatment engagement

Were collected for both study conditions. Participants' W-SUD use,

including days of use and number of messages sent, were collected via the app, as were module completion rates, lesson acceptability ratings indicated on a binary scale (i.e., a "thumbs up" or "thumbs down" emoticon), and mood impact after tools utilization (i.e., feeling "same", "better", or "worse" after completion). For each day of the intervention, the participant was considered "active" if they exchanged one or more messages on the app that day. Additionally, W-SUD assessed mood and cravings or urges to use. Current emotional states were reported through emoji selection with a default menu of 19 total moods including options for negative (angry, sad, anxious), positive (happy, content), and average mood (okay), with additional functionality to type in free text emotion words and/or self-selected emoji expressions. Cravings were assessed via the W-SUD app as: not at all (0), a little bit (1), moderately (2), quite a bit (3), or extremely (4). In the psychoeducational control condition, treatment engagement was determined by participants' button clicks to affirm reading the online factsheets.

2.5. Statistical analyses

2.5.1. Participant Retention and Sample Description

Detection of ongoing fraudulent concerns led to removal of some participants post-randomization. Threats to the equating of conditions via randomization were evaluated by testing baseline characteristic differences by group. Specifically, we characterized the magnitude of imbalance between groups on each covariate using the standardized mean difference (SMD). The SMD is most commonly used as a balance diagnostic in propensity score-matched observational samples but can also be applied to assess balance in a randomized trial setting (Austin, 2009; Stuart et al., 2013). An SMD above 0.1 is often considered an indicator for meaningful covariate imbalance, though higher thresholds can be tolerated if the imbalanced covariates are adjusted for in the analysis. Note that imbalance on prognostically relevant covariates is more important to address in analysis (e.g., through covariate adjustment) than imbalance on covariates that are not prognostic of the outcome (Ho et al., 2007).

2.5.2. Primary outcome analysis

The primary analysis compared the change in the number of substance use occasions from baseline to EOT between the W-SUD intervention group and the psychoeducational control group. To estimate the average treatment effect on the primary outcome, we fit a weighted linear regression model of substance use change score against group assignment, adjusting for a subset of observed covariates determined through a stepwise selection procedure. Because the outcome measure necessitated both baseline and EOT data, we used an intent-to-treat approach including only the subsample of participants retained, where retention was defined as remaining in the study (not withdrawing) and completing the 8-week survey. Among those deemed nonfraudulent, retention was >91% at each follow-up assessment. To account for differential dropout of participants, data were weighted by the inverse estimated probability of retention. Missing items within multi-item scales were imputed using the mean of non-missing items. Further details on the regression models (e.g., methods for variable selection, weight stabilization, and confidence interval estimation) and on handling of missing data can be found in Supplement A.1 and A.2.

In addition to comparing substance use change scores between groups, we estimated within-group Cohen's d effect sizes, calculated as the mean of the within-group substance use change scores divided by the within-group standard deviation in change scores. Lastly, we estimated pairwise Pearson correlations between the primary outcome change score and the change scores for confidence (BSCQ), depression (PHQ-8), anxiety (GAD-7), and work productivity (SPS-6).

2.5.3. Secondary outcome analysis

We estimated within-group effect sizes and treatment effects for all secondary outcomes using the same methods described above. When fitting the linear regression model, whichever outcome measure was being entered as the dependent variable was excluded from the set of possible covariates to include as independent variables. For secondary outcomes that had missing values due to branching logic (DAST-10, SPS-6, QDS), participants with missing values for the outcome were excluded (i.e., a complete-case analysis was performed). Otherwise, missing data were handled as described above and in Supplement A.2. To account for multiple hypothesis testing across secondary outcomes, we adjusted *p*-values using the Benjamini-Hochberg procedure, which controls for false discovery rate (Benjamini & Hochberg, 1995).

2.5.4. Exploratory analyses

In an exploratory subgroup analysis, we examined subgroups defined by the following variables: gender (man vs. woman or other), race (White vs. non-White), ethnicity (Hispanic/Latinx vs. not Hispanic/Latinx), primary substance (alcohol vs. not alcohol), prior therapy experience (never vs. formerly vs. currently), online screening platform (REDCap vs. Qualtrics), age (below vs. above median), past 30-day tobacco use (no vs. yes), baseline substance use occasions (below vs. above median), CAGE-AID (2 vs. 3 vs. 4), PHQ (<10 vs. ≥ 10), GAD (<10 vs. ≥ 10), DAST-10 (<3 vs. ≥ 3), BSCQ (below vs. above median), SIPAD (below vs. above median), QDS Days of Heavy Drinking (below vs. above median), QDS-Average Drinks Per Week (below vs. above median), TAA-Goal (fully quit vs. other goal), Cravings (below vs. above median), prior diagnosis of mental health condition (no vs. yes), and CSQ (high = 31–32 vs. medium = 26–30 vs. low = 8–25). Gender, race, ethnicity, primary substance, and prior therapy experience were preregistered moderators in the trial protocol.

To examine whether treatment effects for the primary outcome differed across subgroups, we used the same regression model-fitting procedure described above but included binary subgroup label(s) and interaction(s) between group assignment and subgroup label(s) as independent variables. Details on how treatment effect estimates were extracted from the fitted coefficients can be found in Supplement A.3.

Additionally, we assessed whether changes in substance use were associated with self-reported engagement in additional SUD treatment services external to the trial. Specifically, within each trial arm, we calculated the Spearman correlation between the change in the number of substance use occasions from baseline to EOT and whether the participant reported past two-month engagement in SUD treatment services at EOT. We also calculated the correlation between baseline to 1-month follow-up substance use change and whether the participant reported past 30-day engagement in SUD treatment services at follow-up. Finally, we calculated the correlation between the EOT and 1-month follow-up substance use change and whether the participant reported past 30-day engagement in SUD treatment services at follow-up.

2.5.5. Satisfaction, acceptability/feasibility, alliance, and engagement analyses

We used Kendall's tau rank correlation coefficient to measure the correlation between substance use change scores and satisfaction, acceptability, feasibility, and therapeutic alliance scores. Additionally, we ran descriptive analyses for engagement-related data in both the W-SUD and psychoeducational control groups. For the W-SUD group, we examined patterns in the app activity across the 8-week intervention period. For the psychoeducational control group, we examined the number of PDFs that were opened and "read" by participants, where "read" status was tracked through clicking a "Submit" button in REDCap Cloud.

2.5.6. Sensitivity analyses

To check whether our results were sensitive to choice of analysis methods, in addition to the weighted linear regression, we estimated treatment effects for the primary and secondary outcomes using a simple difference-in-means estimator and an unweighted linear regression. As an alternative to the stepwise selection procedure for choosing

adjustment variables in the regression, we also selected adjustment variables based on observed standardized mean differences between groups, i.e. SMD > 0.1, SMD > 0.2. We also performed a per-protocol analysis for the primary outcome, where we estimated the treatment effect using only W-SUD participants who were active for at least 4 weeks (sent at least one in-app message on at least one day per week across 4 or more weeks) and for psychoeducational control participants who opened at least 4 PDFs.

3. Results

3.1. Sample recruitment, enrollment, and exclusion

Participant recruitment spanned 11/14/2022–12/13/2022 and 3/20/2023–5/4/2023, with the final follow up assessment completed on 8/17/2023. The 3-month gap in recruitment was due to the study's need to shift recruitment methods and update IRB approval for research panel recruitment following high-volume fraudulent efforts at enrollment via social media. To meet the study's timeline and budgetary constraints, participation was closed 8/31/2023 before the target sample size was met, though lack of power did not appear to affect study conclusions, as discussed later.

Of the 258 participants randomized, 2 (0.8%) withdrew and 55 (21.3%) were removed from the study. Post-randomization removals were due to: failure to provide a consistent birthdate (28/258, 10.9%); duplicate enrollment (5/258, 1.9%); multiple screening attempts with inconsistent information (5/258, 1.9%); prior Woebot use (7/258, 2.7%); W-SUD app registration when assigned to the psychoeducation group (7/258, 2.7%); and reporting substance use at screening but not in the past 30 days at baseline (3/258, 1.2%) (see Fig. 2 Consort diagram). One participant who withdrew from the study was retained for analysis because trial eligibility criteria were met; the other 56 (21.7%, 56/258) withdrawals and removals were excluded from analyses with significantly more ($p = 0.007$) from the W-SUD group (24/131, 18.3%) than the psychoeducational control group (32/127, 25.2%).

3.2. Sample description at baseline

The analytic ($N = 202$) sample averaged 38.3 years of age (SD = 10.4), was 53.5% (108/202) female, 71.3% (144/202) White, 88.1% (178/202) non-Hispanic, 52.0% (105/202) partnered, and 66.8% (135/202) employed part- or full-time. Most participants (73.3%, 148/202) identified alcohol as their primary, secondary, or tertiary problematic substance, followed by cannabis (30.7%, 62/202); stimulants (17.3%, 35/202); and cocaine (12.4%, 25/202). At baseline, the number of substance use occasions averaged 33.1 (SD 18.3). Though W-SUD was not designed to treat tobacco use, a majority (58.9%) reported past 30-day use of tobacco.

At baseline, the sample's mean score on the SIP-AD was 18.1 (SD 11.2), 4.1 (SD 2.5) on the DAST-10; 2.0 (SD 1.1) for cravings; and 50.1% (SD 18.7%) on the BSCQ. A majority (54.5%, 110/202) reported no history of a psychiatric diagnosis; 32.7% (66/202) reported an anxiety disorder, 23.3% (47/202) unipolar depression, and 7.4% (15/202) an SUD diagnosis. Over a quarter (27.2%, 55/202) reported multiple prior psychiatric diagnoses. One in five (20.8%, 42/202) were currently taking psychiatric medication, 15.3% (31/202) were currently in therapy, and 10.4% (21/202) were actively in SUD treatment. At baseline, the sample averaged 8.7 (SD 5.9) on the PHQ-8 and 7.7 (SD 5.6) on the GAD-7, in the mild range for depression and anxiety, respectively.

Table 1 presents baseline characteristics by treatment group, $N = 105$ for W-SUD and $N = 97$ for the psychoeducational control. Some characteristics became imbalanced between the two groups after differential removal of participants post-randomization. There were apparent group differences on demographic characteristics for sex (SMD = 0.290), gender (SMD = 0.279), race (SMD = 0.556), and insurance status (SMD = 0.251). There were also apparent differences on baseline

Table 1
Baseline Demographic, Mental Health, and Substance Use Characteristics by Group ($N = 202$).

	W-SUD ($N = 107$)		Psychoeducational Control ($N = 95$)		Abs SMD
Variable	M (SD), Range	n (%)	M (SD), Range	n (%)	
Age in Years	38.2 (11.4), 18–65		38.6 (9.2), 21–63		0.04
Biological Sex					0.29
Female		50 (46.7)		58 (61.1)	
Male		57 (53.3)		37 (38.9)	
Gender					0.28
Woman		50 (46.7)		56 (58.9)	
Man		56 (52.3)		37 (38.9)	
Non-binary		1 (0.9)		2 (2.1)	
Sexual Orientation					0.13
Heterosexual		90 (84.1)		84 (88.4)	
Non-heterosexual		17 (15.9)		11 (11.6)	
Ethnicity[i]					0.008
Hispanic/Latinx		12 (11.2)		11 (11.6)	
Not Hispanic/Latinx		94 (87.9)		84 (88.4)	
Race					0.44
White		86 (80.4)		58 (61.1)	
Black/African American		17 (15.9)		28 (29.5)	
Other (e.g., Asian American, Native American, Multiracial)		3 (2.8)		9 (9.5)	
Marital Status					0.12
Married/ Cohabitating/ Partnered		53 (49.5)		52 (54.7)	
Divorced/Separated/ Widowed		13 (12.1)		9 (9.5)	
Never Married/Single		41 (38.3)		34 (35.8)	
Educational Degree					0.09
High School Degree or Less		17 (15.9)		13 (13.7)	
Some College / Technical School		37 (34.5)		36 (37.9)	
College Degree		30 (28.0)		27 (28.4)	
Graduate or Post-graduate Degree		23 (21.5)		19 (20.0)	
Employment Status					0.11
Employed Part or Full-Time		74 (69.2)		61 (64.2)	
Unemployed (homemaker, student, retired)		33 (30.8)		34 (35.8)	
Disability Status					0.10
Yes (has disability)		8 (7.5)		10 (10.5)	
No (does not have disability)		98 (91.6)		85 (89.5)	
Insurance Status					0.25
Private		56 (52.3)		47 (49.5)	
Medicaid or Medicare		32 (29.9)		33 (34.7)	
VA or TRICARE		1 (0.9)		3 (3.2)	
Don't have insurance		18 (16.8)		10 (10.5)	

(continued on next page)

Table 1 (*continued*)

Variable	W-SUD ($N = 107$) M (SD), Range	W-SUD ($N = 107$) n (%)	Psychoeducational Control ($N = 95$) M (SD), Range	Psychoeducational Control ($N = 95$) n (%)	Abs SMD
Lifetime Mental Health Conditions					
Unipolar Depression		20 (18.7)		27 (28.4)	0.23
Bipolar/Manic Depression		9 (8.4)		5 (5.3)	0.12
Anxiety Disorder		34 (31.8)		32 (33.7)	0.04
Substance Use Disorder		8 (7.5)		7 (7.4)	0.004
Other Mental Health Disorder		23 (21.5)		24 (25.3)	0.09
Multiple Diagnoses		29 (27.1)		26 (27.4)	0.006
No History of Mental Illness		62 (57.9)		48 (50.5)	0.15
Prior Therapy Experience					0.20
Never in Therapy		66 (61.7)		52 (54.7)	
In Therapy Formerly		28 (26.2)		25 (26.3)	
Currently in Therapy		13 (12.1)		18 (18.9)	
Psychiatric Medications Currently		25 (23.4)		17 (17.9)	0.14
Past 30-Day SUD Treatment					
Self-help Groups		2 (1.9)		5 (5.3)	0.18
Outpatient Services		4 (3.7)		5 (5.3)	0.07
Intensive Outpatient/ Residential		2 (1.9)		2 (2.1)	0.02
Medication use for SUD Treatment		2 (1.9)		1 (1.1)	0.07
Primary, Secondary, or Tertiary Problematic Substance					
Alcohol		76 (71.0)		72 (75.8)	0.11
Cannabis		27 (25.2)		35 (36.8)	0.25
Stimulants		20 (18.7)		15 (15.8)	0.08
Cocaine		16 (15.0)		9 (9.5)	0.17
Tobacco		36 (33.6)		34 (35.8)	0.05
Other (e.g. Heroin, Hallucinogens, Inhalants)		22 (20.6)		15 (15.8)	0.12
PHQ-8, possible range 0–24	8.7 (6.1), 0–24; $N =$ 106		8.7 (5.8), 0–23; N = 94		0.004
10+ Moderate/Severe		40 (37.4)		35 (36.8)	
GAD 7-Anxiety, possible range 0–21	7.6 (5.8), 0–21; N = 106		7.8 (5.3), 0–21; N = 94		0.04
10+ Moderate/Severe		33 (30.8)		34 (35.8)	
SPS-6, possible range 6–30					0.27
6–10 (Low presenteeism)		0		0	
11–15		6 (5.6)		4 (4.2)	
16–20		16 (15.0)		19 (20.0)	
21–25		26 (24.3)		14 (14.7)	
26–30 (High presenteeism)		26 (24.3)		24 (25.2)	
DAST, possible range 0–10	4.0 (2.4), 0–10; $N =$ 80		4.2 (2.5), 0–10; $N =$ 68		0.09
3+ Moderate/Severe		56 (52.3)		45 (47.4)	
SIP-AD, possible range 0–45	18.3 (10.9), 0–42; N = 105		17.9 (11.7), 0–45; N = 94		0.03
CAGE-AID, possible range 2–4	3.38 (0.74), 2–4; N = 105		3.33 (0.78), 2–4; $N = 94$		0.08
BSCQ, possible range 0–100%	50.4 (18.8), 3.8–93.9; $N = 103$		49.8 (18.7), 1.9–92.5; $N = 91$		0.06
Cravings Past 7 days, possible range 0–4	1.97 (1.13), 0–4; N = 107		1.96 (1.18), 0–4; N = 95		0.01
Thoughts About Abstinence, possible range 1–10 per item					
Desire to quit	6.47 (2.11), 2–10; N = 106		6.81 (2.70), 1–10; N = 95		0.14
Expected success in quitting	5.75 (2.36), 1–10; N = 107		6.01 (2.56), 1–10; N = 95		0.11
Difficulty not relapsing	7.04 (1.85), 2–10; N = 107		6.62 (2.67), 1–10; N = 95		0.18
Quick Drinking Screen Past 30 days					
Average days per week	4.01 (1.70), 0–7; N = 91		4.13 (2.03), 0–7; $N = 83$		0.07
Average standard drinks per day	4.67 (3.68), 1–27; N = 91		4.27 (2.99), 1–15; N = 83		0.12
Number of heavy drinking occasions	7.11 (7.42), 0–30; N = 91		5.92 (6.57), 0–30; N = 83		0.17
Past 30 Days of Substance Use					
Alcohol	15.1 (8.1), 0–30	91 (85.0)	16.6 (9.2), 1–30	83 (87.4)	0.15
Cannabis	16.8 (10.5), 1–30	53 (49.5)	19.8 (10.1), 2–30	59 (62.1)	0.34
Cocaine	10.5 (6.5), 2–25	17 (15.9)	13.4 (9.2), 2–30	8 (8.4)	0.12
Prescription stimulants	13.5 (11.4), 0–30	12 (11.2)	17.2 (13.5), 2–30	6 (6.3)	0.10
Methamphetamine	18.1 (10.2), 3–30	15 (14.0)	21.5 (9.5), 5–30	14 (14.7)	0.08
Inhalants	10.4 (8.4), 5–25	5 (4.7)	4.0	1 (1.1)	0.23
Sedatives	8.5 (7.7), 1–30	18 (16.8)	10.7 (9.2), 0–30	11 (11.6)	0.02
Hallucinogens	2.3 (1.5), 1–5	6 (5.6)	8.6 (8.8), 2–21	5 (5.3)	0.17
Street Opioids	5.0	1 (0.9)	8.0 (2.8), 6–10	2 (2.1)	0.13
Prescription Opioids	12.1 (9.8), 1–30	15 (14.0)	14 (8.3), 3–30	9 (9.5)	0.07
Past Month Substance Use Occasions	31.0 (18.5) 4–117		35.5 (17.9), 5–80		0.25
Past 30-day tobacco use	13.9 (14.0), 0–30	63 (58.9)	14.3 (13.9), 0–30	56 (58.9)	0.03

Abs SMD = Absolute Standardized Mean Difference. Note that for variables where participants were able to choose multiple values (e.g. Lifetime Mental Health Conditions), a separate SMD is reported for each level of that variable.

[i] one W-SUD participant indicated prefer not to answer.

mental health conditions (unipolar depression SMD = 0.231), substance use treatment (self-help groups SMD = 0.184), SPS-6 (SMD = 0.273), past 30-day substance use (cannabis = 0.339, inhalants = 0.228, hallucinogens = 0.168). Some covariates had an SMD only slightly greater than the 0.1 threshold for imbalance, which we interpret as weaker evidence and potentially an artifact of small sample size, as seen with

past 30-day inhalant use, rather than a meaningful sign of imbalance. Baseline imbalance was addressed in the analyses through regression adjustment and inverse probability of retention weighting, which adjusts for differential dropout.

3.3. Primary and secondary outcomes by group

Table 2 summarizes primary and secondary outcomes by group. Between baseline and EOT, the within-group change in substance use occasions was −13.6 (SD 15.4, d = −0.879) in the W-SUD group and −12.9 (SD 15.3, d = −0.847) in the psychoeducational control group, indicating a moderate to large pre- versus post-study change in both groups. Similar within-group changes were observed for substance use occasions summed only across primary, secondary, and tertiary substances (W-SUD: -11.4, SD 13.0, d = −0.884; psychoeducational control: -9.8, SD 13.1, d = −0.750). At EOT, 8 out of 100 (8%) W-SUD participants with substance use occasion data and 9 out of 90 (10%) psychoeducational control participants with substance use occasion data reported no substance use at all in the past 30 days.

In the W-SUD group, the change from baseline to EOT was medium to large for SIP-AD (−9.02, SD 11.6, d = −0.775), DAST-10 (−1.21, SD 2.33, d = −0.519), cravings (−0.66, SD 1.18, d = −0.558), heavy drinking days (−4.09, SD 6.35, d = −0.644), drinks per week (−8.67, SD 14.7, d = −0.591), and SPS (−5.42, SD 7.2, d = −0.752). In the psychoeducational control group, the change from baseline to EOT was medium to large for SIP-AD (−9.24, SD 10.2, d = −0.903), DAST-10 (−1.95, SD 1.99, d = −0.981), cravings (−0.722, SD 1.2, d = −0.602), BSCQ (11.6, SD 20.3, d = 0.571), drinks per week (−8.18, SD 11.5, d = −0.710) and SPS (−4.12, SD 6.97, d = −0.591).

Table 3 shows the estimated treatment effects for all primary and secondary outcomes. The estimated average treatment effect for change in substance use occasions between baseline and EOT (primary outcome) was −0.675 ($p = 0.741$; 95% CI -4.96-3.49, d = −0.040), and the estimated treatment effect for change in substance use occasions between baseline and 12-week follow-up (secondary outcome) was −0.339 ($p = 0.88$; 95% CI -5.1-4.1, d = −0.018). While the differences between groups were weak and not significant, negative point estimates indicate a larger average decrease in substance use occasions in W-SUD than in psychoeducational control at both 8- and 12-week follow-ups. No treatment effects for other secondary outcomes were found to be significant after adjusting for multiple hypothesis testing. The covariates included in the regression models used for estimating treatment effects can be found in Supplemental Table B7.

Null treatment effects across all primary and secondary outcomes were also observed in the per-protocol analysis (Supplemental Table B1), when different treatment effect estimation methods (unweighted linear regression, difference-in-means) were used, and when adjustment variables were chosen by standardized mean difference threshold rather than stepwise selection (Supplemental Tables B2-B4). Supplemental Table B5 shows estimated treatment effects for baseline-to-EOT change in substance use occasions across subgroups of interest. No significant between treatment group effects were found within- or between subgroups. Supplemental Table B6 shows no significant associations between change in substance use occasions and engagement in SUD treatment services.

Table 4 shows Pearson correlations between baseline-to-EOT change in substance use occasions (primary outcome) and the baseline-to-EOT change scores for confidence (BSCQ), depression (PHQ-8), anxiety (GAD-7), and work productivity (SPS-6). In the W-SUD group, baseline-to-EOT decrease in substance use occasions was significantly correlated with baseline-to-EOT increase in BSCQ confidence score ($r = -0.345$, $p < 0.01$). In the psychoeducational control group, baseline-to-EOT substance use occasions was significantly correlated with baseline-to-EOT decrease in PHQ-8 depression score ($r = 0.238$, $p < 0.05$) and GAD-7 anxiety score ($r = 0.331$, $p < 0.001$). Comparing the W-SUD group to the psychoeducational control group, there was a significant difference in correlation for the BSCQ change score (Fisher-transformed z = −0.286, p < 0.01).

3.4. Engagement and acceptability by group

Table 5 and Fig. 3 summarize engagement metrics in the W-SUD and psychoeducational control groups.

3.4.1. W-SUD group engagement

Of the 107 participants assigned to W-SUD, 92 (86.0%) registered with the app and 90 (84.1%) had at least one day of activity in the app, with a median of 322.5 [IQR 59.8–924.0] in-app messages sent by users across the 8-week intervention. Across all 8 weeks, 82 (89%, 82/92) participants completed and rated at least one psychoeducational lesson, and 47 (51%, 47/92) participants utilized and rated at least one tool for practicing psychotherapeutic skills. Participants could rate lessons either positively or negatively using a “thumbs up” or “thumbs down” emoticon and could rate tools positively, neutrally, or negatively by reporting feeling “better”, the “same”, or “worse” after tool use. Across all 8 weeks, W-SUD users completed and rated 1191 lessons and 1139 (95.6%) were rated positively (thumbs up). Across all 8 weeks, W-SUD users completed 128 total tools and 98 (76.6%) were rated positively (“feel better”), 29 (22.7%) neutrally, and 1 (<1%) worse, respectively.

Table 2
Changes in Outcome Measures from Baseline to Week 8 and Baseline to Week 12 by Group.

Measure	Baseline to Week 8 (EOT) Change				Baseline to Week 12 (1-mo follow-up) Change			
	W-SUD		Psychoed. Control		W-SUD		Psychoed. Control	
	M (SD)	Cohen's d	M (SD)	Cohen's d	M (SD)	Cohen's d	M (SD)	Cohen's d
Sub Use Occ	−13.6 (15.4)	−0.879	−12.9 (15.3)	−0.847	−15.7 (18.0)	−0.873	−15.4 (16.6)	−0.929
Pri/Sec/Ter Sub Use Occ	−11.4 (13.0)	−0.884	−9.8 (13.1)	−0.750	−12.4 (14.1)	−0.88	−12.7 (14)	−0.907
SIP-AD	−9.02 (11.6)	−0.775	−9.24 (10.2)	−0.903	−11.5 (11.5)	−0.994	−10.9 (11.5)	−0.95
DAST	−1.21 (2.33)	−0.519	−1.95 (1.99)	−0.981	–	–	–	–
Cravings	−0.66 (1.18)	−0.558	−0.722 (1.2)	−0.602	−0.816 (1.15)	−0.709	−0.892 (1.4)	−0.636
BSCQ	6.77 (20.4)	0.332	11.6 (20.3)	0.571	11.6 (21.8)	0.533	15.6 (26.1)	0.591
QDS - Heavy Drinking	−4.09 (6.35)	−0.644	−2.51 (6.02)	−0.417	−5.43 (7.21)	−0.753	−3.14 (7.75)	−0.405
QDS - Drinks Per Week	−8.67 (14.7)	−0.591	−8.18 (11.5)	−0.710	−11.1 (15.7)	−0.706	−9.43 (13.9)	−0.676
TAA - Desire to Quit	0.21 (2.33)	0.091	0.12 (2.71)	0.046	–	–	–	–
TAA - Expected Success	0.42 (2.37)	0.177	0.53 (2.72)	0.197	–	–	–	–
TAA - Difficulty Not Using	−0.64 (2.44)	−0.262	−0.28 (2.91)	−0.098	–	–	–	–
GAD	−1.88 (4.77)	−0.395	−1.72 (4.91)	−0.351	−2.09 (4.67)	−0.447	−2.57 (5.08)	−0.506
PHQ	−2.82 (4.65)	−0.607	−2.05 (5.95)	−0.345	−2.83 (4.79)	−0.591	−2.87 (6.3)	−0.455
SPS	−5.42 (7.20)	−0.752	−4.12 (6.97)	−0.591	–	–	–	–

Note: Sample sizes for analyses range from $n = 65$ (DAST) to 100 for W-SUD and $n = 54$ (DAST) to 90 for Psychoed. Control.

Table 3
Estimated Treatment Effects for Primary and Secondary Outcomes.

Measure	Beta (p; 95% CI), 8-week Change Score	Adj p, 8-week	Beta (p; 95% CI), 12-week Change Score	Adj p, 12-week
Primary Outcome				
Substance Use Occasions	−0.675 (0.741; [−4.96, 3.49]	–	–	–
Secondary Outcomes				
Substance Use Occasions	[see above, primary outcome]	–	−0.339 (0.88; [−5.1, 4.1])	0.913
BSCQ	−2.48 (0.419; [−9.05, 2.92])	0.913	−4.69 (0.157; [−11.6, 2.07])	0.896
DAST	0.893 (0.019; [0.189, 1.76])	0.608	–	–
GAD	−0.142 (0.835; [−1.37, 1.25])	0.913	0.125 (0.832; [−1.15, 1.28])	0.913
PHQ	−0.629 (0.384; [−2, 0.651])	0.913	0.0111 (0.99; [−1.61, 1.45])	0.990
SPS	−0.877 (0.405; [−2.79, 0.941])	0.913	–	–
Cravings	0.0804 (0.568; [−0.239, 0.415])	0.913	0.107 (0.546; [−0.261, 0.444])	0.913
QDS - Heavy Drinking	−1.28 (0.168; [−3.1, 0.595])	0.896	−2.03 (0.099; [−4.65, 0.249])	0.896
QDS - Drinks Per Week	−0.427 (0.843; [−4.68, 3.31])	0.913	−0.807 (0.747; [−5.96, 2.88])	0.913
TAA - Desire to Quit	0.125 (0.75; [−0.598, 0.858])	0.913	–	–
TAA - Expected Success	−0.543 (0.138; [−1.22, 0.072])	0.896	–	–
TAA - Difficulty Not Using	−0.402 (0.31; [−1.15, 0.309])	0.913	–	–
SIP-AD	−0.277 (0.87; [−2.76, 2.99])	0.913	−1.07 (0.499; [−3.94, 2.16])	0.913
Pri/Sec/Ter Sub Use	−1.28 (0.471; [−5.07, 2.36])	0.913	0.701 (0.678; [−3.26, 4.42])	0.913

Identical adjusted p-values result from the Benjamini–Hochberg (BH) correction, which preserves rank order by carrying the last (larger) adjusted value backward when smaller *p*-values would otherwise appear out of order after scaling. The correction was implemented in R using p.adjust(method = "BH").

3.4.2. Psychoeducational control group engagement

Of the 90 participants assigned to psychoeducational control with primary outcome data at 8-weeks, 84 (93.3%) opened and clicked submit on at least one PDF through the REDCap Cloud interface. The mean number of PDFs opened across all psychoeducational control participants was 5.8 (SD 2.4). Out of the psychoeducational control participants, 73 (81.1%) opened four or more PDFs, and 29 (32.2%) opened all eight PDFs.

Table 6 summarizes satisfaction, acceptability, and feasibility ratings in both the W-SUD and psychoeducational control groups. A higher satisfaction score at EOT was significantly associated with a greater baseline-to-EOT decrease in substance use occasions in both the W-SUD group (tau = −0.174, $p = 0.015$) and in the psychoeducational control group (tau = −0.203, $p = 0.009$). There was a significant association between feasibility and baseline-to-EOT decrease in substance use occasions in the psychoeducational control group (tau = −0.167, $p = 0.031$) but not in the W-SUD group. In the W-SUD group, higher scores on the goals (tau = −0.190; $p = 0.008$), tasks (tau = −0.190; $p = 0.017$), and affective bond formation (tau = −0.158; $p = 0.029$) WAI-SR subscales were significantly associated with greater baseline-to-EOT decrease in substance use occasions.

3.5. Safety outcomes

No study participants reported a serious adverse event during the study period. Of the 202 participants, 8 selected "prefer not to answer" on one of four serious adverse event assessment questions, of whom 3 confirmed with study staff by phone that there was no adverse event. The other 5 did not respond to study staff and were provided with a resource list. The IRB was consulted and the selection of "Prefer not to answer" was not considered to be an adverse event.

Table 4
Correlation Between Substance Use Change Score and Confidence, Depression, Anxiety, and Work Productivity Change Scores from baseline to EOT in W-SUD versus Psychoeducational Control.

Pearson correlation between Sub Use Occ and:	W-SUD	Psychoed. Control	Fisher-transformed *Z*-score
BSCQ	−0.345**	0.07	−0.286**
PHQ	0.158	0.238*	−0.568
GAD	0.177	0.331***	−1.12
SPS	−0.071	−0.23	0.878

* $p < 0.05$,.
** $p < 0.01$.
*** $p < 0.001$.

4. Conclusions

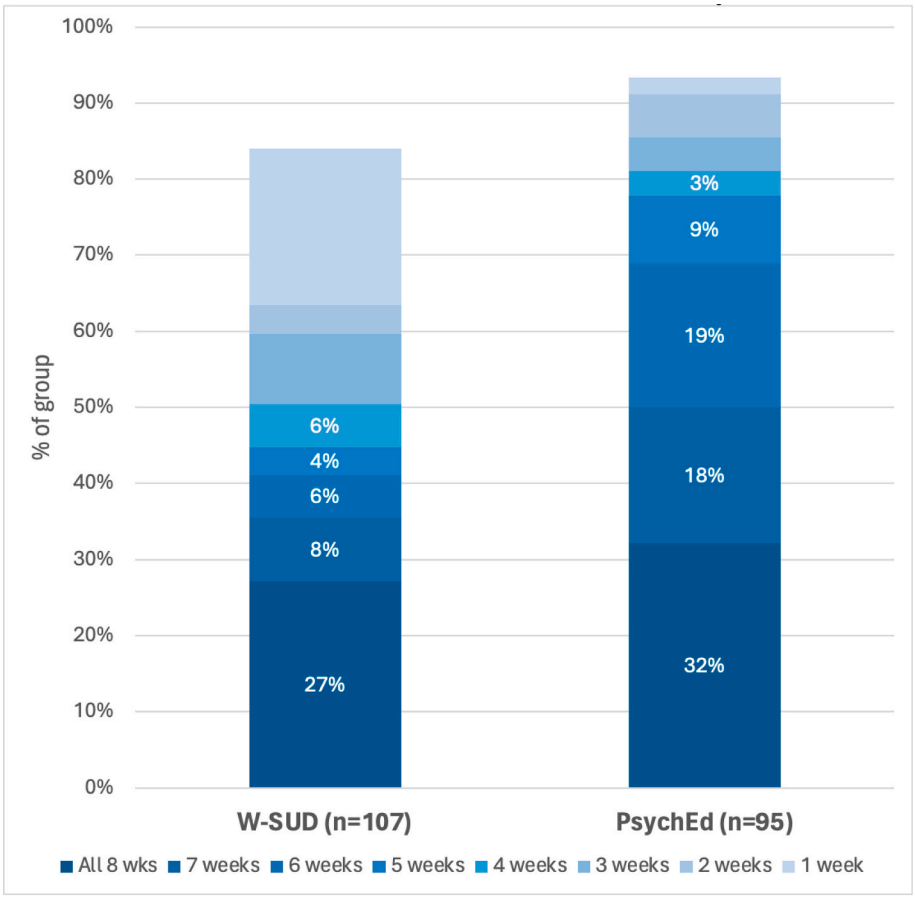
Building upon prior findings of feasibility, acceptability, and short-term efficacy relative to a waitlist control, the current study examined the sustained effects of W-SUD for reducing problematic substance use at 1-month post-treatment and tested against a psychoeducation control condition. Significant within-group reductions in substance use occasions were observed from baseline to EOT, replicating prior findings and with comparable effect sizes for changes over time. Further, the changes were sustained at 1-month follow-up. Yet, the between-group comparisons indicated that W-SUD's effects did not significantly differ from the psychoeducational control condition. Notably, both conditions had moderate to large pre- versus post-treatment change (>0.80), which for the psychoeducational control condition was higher than anticipated based on the literature (Batterham et al., 2018; Deady et al., 2016; Doumas et al., 2009; Doumas & Haustveit, 2008; Lau-Barraco & Dunn, 2008).

It is possible that both conditions reduced substance use behaviors and roughly equally so. The high engagement, acceptability, and feasibility scores corroborate this. Selection of a control group comparison is nuanced. Here, we chose one that was perhaps more active than desired for a trial intended to test superiority. Further, the psychoeducation condition may have activated some of the same mechanisms of short-term improvement as W-SUD, such as increased self-monitoring,

Table 5
Engagement in W-SUD and Psychoeducational Control Groups Across all 8 Weeks.

W-SUD (N = 107)	N Any (%)	M (SD)	Mdn
Days Active	90 (84.1)	16.2 (13.6)	12.5
Messages Sent by Users	90 (84.1)	322.5 (446.9)	322.5
Lessons completed and rate	1191		
% Completed Lessons Rated Positive	1139 (95.6%)		
% Completed Lessons Rated Negative	52 (4.3%)		
Tools Completed and Rated	128		
% Completed Tools Rated Positive	98 (76.6%)		
% Completed Tools Rated Neutral	29 (22.7%)		
% Completed Tools Rated Negative	1 (<1%)		

Psychoeducation (N = 95)	N Any (%)	M(SD)	Mdn
Opened PDF (per submission in REDCap Cloud)	84 (93.3)	5.8 (2.4)	6.5



*15 (14.0%) participants randomized to W-SUD did not register with the app

Fig. 3. Distribution of Weeks of Engagement with W-SUD and the Psychoeducational Control Group.

Table 6
Within Group Correlations for the Primary Outcome (Substance Use Change Score, baseline to EOT) with Satisfaction, Acceptability, Feasibility, and Alliance.

		W-SUD (N = 107)				Psychoeducational Control (N = 95)			
Scale	Sub-scales	N	M (SD)	Min, Max	Tau (p-value)	N	M (SD)	Min, Max	Tau (p-value)
CSQ		99	26.9 (5.2)	13.3, 32	**−0.174 (0.015)**	87	28.1 (4.5)	14, 32	**−0.203 (0.009)**
URP-I									
	Acceptability	99	30.1 (5.6)	12, 36	−0.134 (0.061)	89	30.3 (5.8)	11, 36	−0.118 (0.120)
	Feasibility	99	30.6 (4.5)	16, 36	−0.016 (0.822)	89	32.6 (3.5)	23, 36	**−0.167 (0.031)**
WAI-SR									
	Goals	97	14.8 (14.4)	4, 20	**−0.190 (0.008)**		–	–	–
	Tasks	98	14.4 (4.5)	4, 20	**−0.190 (0.017)**		–	–	–
	Bond	97	15.2 (4.5)	4, 20	**−0.158 (0.029)**		–	–	–

expectancy effects, and structured weekly engagement. With theoretical basis and changes observed over time, W-SUD's core components may still be clinically meaningful. Also plausible is that exogenous factors unrelated to either condition, such as transitioning back into usual routines after easing of COVID pandemic restrictions, may have contributed to the larger-than-usual reported decreases in substance use.

The noise introduced by fraudulent activity also remains a concern. Even with careful filtering, we may not have caught all fraudulent enrollments (Moore et al., 2025), and any remaining in the W-SUD treatment group could have biased findings toward the null. Notably, more participants in the comparison arm were excluded post-randomization than in the intervention arm, largely because they were caught having enrolled into W-SUD, suggesting double enrollment. Discussed in detail elsewhere (Moore et al., 2025), additional strategies for defeating double enrollment include live video consenting with ID verification, comparing IP address to physical address, and budgeting for more extensive identity verification software or services.

The high volume surges in fraudulent efforts at study enrollment, due in large part to social media recruitment, took staff attention with study time delays (Moore et al., 2025). Ultimately, the size of the analytic cohort was roughly half the target sample size ($N = 107$ in the W-SUD group and $N = 95$ in the psychoeducation group compared to the target sample size of $N = 194$ participants in each group). A smaller sample size reduces the probability of detecting a true effect. However, the observed treatment effect for baseline to EOT (d = −0.040) was far weaker than forecasted (d = 0.32), and a larger sample was unlikely to change the study conclusion. Despite not meeting the target sample size, we were still able to observe significant within-group pre- versus post-treatment change in substance use occasions and in most secondary measures of interest.

Another study limitation was reliance on self-report of substance use, though a strength was assessment of various aspects of substance use,

cravings/urges, and substance use problems/consequences, and findings were highly consistent with detection of change over time within but not between groups. Testing for phosphatidylethanol (PEth), a biomarker of alcohol use, was deemed unsuitable for our large, decentralized trial addressing varied SUDs. In a remotely conducted trial testing a non-human therapeutic, demand characteristics to underreport use or misreport abstinence are low, and any misreporting of substance use changes over time are not anticipated to differ by condition. A study strength was high engagement in both treatment conditions. No serious adverse events were reported despite a population potentially at increased risk for them.

While participants in both conditions reported significant improvements in the primary and secondary outcomes from baseline to EOT and sustained at 1-month post-treatment, with no statistically significant difference in the magnitude of improvement between conditions, follow-up studies may benefit from a setting shift, implementing a real-world effectiveness study rather than an efficacy trial. Investigation within a mental health and/or SUD-specific telehealth treatment clinic may, for example, lend insights into other elements of each program's utility, uptake, retention, and sustained impact. Such implementation science methodology could track and evaluate relevant clinical deployment metrics between the two programs such as referral to either program by providers, provider opinion regarding either program (at baseline and end of intervention), longer-term follow-up periods, and/or health claims data highlighting health care resource utilization. These types of real-world data elements were not captured in the current study and remain relevant to ultimate success of program deployment within healthcare systems, especially in an ambulatory care clinical program design.

Lastly, while engagement with Woebot suggested a respectable dose, with most sending over 300 in-app messages during the 8-week intervention (more than 5 a day), previous research on Woebot specifically identified participant clusters meaningfully different in engagement patterns which related to mood and anxiety outcomes (Hoffman et al., 2023). Recent research recommendations indeed suggest that digital mental health intervention research ought to explore optimal engagement patterns associated with key outcomes, rather than broadly applying a 'one size fits all' metric to engagement (Boucher & Raiker, 2024). A sufficiently powered prospective trial of W-SUD may benefit from further analysis of such specified clusters to better define the quality and type of engagement over time as well as potential associations with outcomes. Such analyses may illuminate patterns otherwise unobservable when the entire group is analyzed together.

CRediT authorship contribution statement

Judith J. Prochaska: Writing – review & editing, Writing – original draft, Supervision, Resources, Methodology, Formal analysis, Data curation, Conceptualization. **Maggie Wang:** Writing – review & editing, Writing – original draft, Validation, Resources, Methodology, Formal analysis, Data curation. **Amy Chieng:** Writing – review & editing, Software, Project administration, Investigation, Data curation. **Jessie Moore:** Writing – review & editing, Investigation, Data curation. **Erin A. Vogel:** Writing – review & editing, Methodology. **Michael Baiocchi:** Writing – review & editing, Validation, Resources, Methodology, Formal analysis, Data curation. **Sarah Pajarito:** Writing – review & editing, Visualization, Software, Project administration, Investigation, Data curation. **Maddison C. Pirner:** Writing – review & editing, Writing – original draft, Visualization, Software, Resources, Project administration, Investigation, Data curation. **Lauren Bullard:** Writing – review & editing, Project administration. **Alison Darcy:** Resources, Funding acquisition, Conceptualization. **Athena Robinson:** Writing – review & editing, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

Funding statement

The National Institute on Drug Abuse (NIDA) with a Small Business Innovations Research Award [R44-DA048712] and a mentored-career development award [K01 DA055073] supported the research. The article's contents are solely the responsibility of the authors and do not necessarily represent the official views of NIDA.

Declaration of competing interest

Sarah Pajarito, MA, Maddison C. Pirner, MPS, Lauren Bullard, PhD, Alison Darcy, PhD, and Athena Robinson, PhD were Woebot Health employees. All other authors declared no conflicts of interest.

Acknowledgements

We acknowledge the study's Data Safety Monitoring Members Travis Cook, PhD, Julie Hamlin, LCSW, Danielle Ramo, PhD, and Kelly Young-Wolff, PhD, MPH for their oversight of the study procedures and outcomes. We are grateful for the guidance from Leonardo Angelone, PhD and Victor Prikhodko, MBA at NIDA. We are also grateful to Molly Bowdring, PhD, for her input on the study.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.josat.2026.209993.

Data availability

Data available from authors upon reasonable request. Minimal dataset available at: doi:10.25740/qt012tb3777

References

Andrews, G., Basu, A., Cuijpers, P., Craske, M. G., McEvoy, P., English, C. L., & Newby, J. M. (2018). Computer therapy for the anxiety and depression disorders is effective, acceptable and practical health care: An updated meta-analysis. *Journal of Anxiety Disorders, 55*, 70–78. https://doi.org/10.1016/j.janxdis.2018.01.001

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine, 28*(25), 3083–3107. https://doi.org/10.1002/sim.3697

Bakker, D., & Rickard, N. (2018). Engagement in mobile phone app for self-monitoring of emotional wellbeing predicts changes in mental health: MoodPrism. *Journal of Affective Disorders, 227*, 432–442. https://doi.org/10.1016/j.jad.2017.11.016

Batterham, P. J., Calear, A. L., Farrer, L., McCallum, S. M., & Cheng, V. W. S. (2018). FitMindKit: Randomised controlled trial of an automatically tailored online program for mood, anxiety, substance use and suicidality. *Internet Interventions, 12*, 91–99. https://doi.org/10.1016/j.invent.2017.08.002

Baumel, A., Muench, F., Edan, S., & Kane, J. M. (2019). Objective user engagement with mental health apps: Systematic search and panel-based usage analysis. *Journal of Medical Internet Research, 21*(9), Article e14567. https://doi.org/10.2196/14567

Beatty, C., Malik, T., Meheli, S., & Sinha, C. (2022). Evaluating the therapeutic Alliance with a free-text CBT conversational agent (Wysa): A mixed-methods study. *Frontiers in Digital Health, 4*, Article 847991. https://doi.org/10.3389/fdgth.2022.847991

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B: Methodological, 57*(1), 289–300. http://www.jstor.org/stable/2346101.

Bickmore, T., & Gruber, A. (2010). Relational agents in clinical psychiatry. *Harvard Review of Psychiatry, 18*(2), 119–130. https://doi.org/10.3109/10673221003707538

Blanchard, K. A., Morgenstern, J., Morgan, T. J., Labouvie, E. W., & Bux, D. A. (2003). Assessing consequences of substance use: Psychometric properties of the inventory of drug use consequences. *Psychology of Addictive Behaviors, 17*(4), 328–331. https://doi.org/10.1037/0893-164X.17.4.328

Boucher, E. M., & Raiker, J. S. (2024). Engagement and retention in digital mental health interventions: A narrative review. *BMC Digital Health, 2*(1), 52. https://doi.org/10.1186/s44247-024-00105-9

Boumparis, N., Schulte, M. H. J., & Riper, H. (2019). Digital mental health for alcohol and substance use disorders. *Current Treatment Options in Psychiatry, 6*(4), 352–366. https://doi.org/10.1007/s40501-019-00190-y

Bower, P., Kontopantelis, E., Sutton, A., Kendrick, T., Richards, D. A., Gilbody, S., … Liu, E. T. (2013). Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. *Bmj, 346*, Article f540. https://doi.org/10.1136/bmj.f540

Brendryen, H., & Kraft, P. (2008). Happy ending: A randomized controlled trial of a digital multi-media smoking cessation intervention. *Addiction, 103*(3), 478–484. https://doi.org/10.1111/j.1360-0443.2007.02119.x

Breslin, F. C., Sobell, L. C., Sobell, M. B., & Agrawal, S. (2000). A comparison of a brief and long version of the situational confidence questionnaire. *Behaviour Research and Therapy, 38*(12), 1211–1220. https://doi.org/10.1016/S0005-7967(99)00152-7

Briesch, A. M., Chafouleas, S. M., Neugebauer, S. R., & Riley-Tillman, T. C. (2013). Assessing influences on intervention implementation: Revision of the usage rating profile-intervention. *Journal of School Psychology, 51*(1), 81–96. https://doi.org/10.1016/j.jsp.2012.08.006

Broglio, K. (2018). Randomization in clinical trials: Permuted blocks and stratification. *JAMA, 319*(21), 2223–2224. https://doi.org/10.1001/jama.2018.6360

Campopiano von Klimo, M., Nolan, L., Corbin, M., Farinelli, L., Pytell, J. D., Simon, C., … Compton, W. M. (2024). Physician reluctance to intervene in addiction: A systematic review. *JAMA Network Open, 7*(7), Article e2420837. https://doi.org/10.1001/jamanetworkopen.2024.20837

Chiauzzi, E., Green, T. C., Lord, S., Thum, C., & Goldstein, M. (2005). My student body: A high-risk drinking prevention web site for college students. *Journal of American College Health, 53*(6), 263–274. https://doi.org/10.3200/jach.53.6.263-274

Collins, S. E., Carey, K. B., & Sliwinski, M. J. (2002). Mailed personalized normative feedback as a brief intervention for at-risk college drinkers. *Journal of Studies on Alcohol, 63*(5), 559–567. https://doi.org/10.15288/jsa.2002.63.559

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 Formative Research, 5*(5), Article e27868. https://doi.org/10.2196/27868

Deady, M., Mills, K. L., Teesson, M., & Kay-Lambkin, F. (2016). An online intervention for co-occurring depression and problematic alcohol use in Young people: Primary outcomes from a randomized controlled trial. *Journal of Medical Internet Research, 18*(3), Article e71. https://doi.org/10.2196/jmir.5178

Donkin, L., Christensen, H., Naismith, S. L., Neal, B., Hickie, I. B., & Glozier, N. (2011). A systematic review of the impact of adherence on the effectiveness of e-therapies. *Journal of Medical Internet Research, 13*(3), Article e52. https://doi.org/10.2196/jmir.1772

Doumas, D. M., & Haustveit, T. (2008). Reducing heavy drinking in intercollegiate athletes: Evaluation of a web-based personalized feedback program. *The Sport Psychologist, 22*(2), 212–228. https://scholarworks.boisestate.edu/cgi/viewcontent.cgi?article=1004&context=counsel_facpubs.

Doumas, D. M., McKinley, L. L., & Book, P. (2009). Evaluation of two web-based alcohol interventions for mandated college students. *Journal of Substance Abuse Treatment, 36*(1), 65–74. https://doi.org/10.1016/j.jsat.2008.05.009

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

Firth, J., Torous, J., Nicholas, J., Carney, R., Rosenbaum, S., & Sarris, J. (2017). Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. *Journal of Affective Disorders, 218*, 15–22. https://doi.org/10.1016/j.jad.2017.04.046

Garnett, C., Perski, O., Tombor, I., West, R., Michie, S., & Brown, J. (2018). Predictors of engagement, response to follow up, and extent of alcohol reduction in users of a smartphone app (drink less): Secondary analysis of a factorial randomized controlled trial. *JMIR mHealth and uHealth, 6*(12), Article e11175. https://doi.org/10.2196/11175

Graham, A. K., Greene, C. J., Kwasny, M. J., Kaiser, S. M., Lieponis, P., Powell, T., & Mohr, D. C. (2020). Coached Mobile app platform for the treatment of depression and anxiety among primary care patients: A randomized clinical trial. *JAMA Psychiatry, 77*(9), 906–914. https://doi.org/10.1001/jamapsychiatry.2020.1011

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

Grodniewicz, J. P., & Hohol, M. (2023). Waiting for a digital therapist: Three challenges on the path to psychotherapy delivered by artificial intelligence [hypothesis and theory]. *Frontiers in Psychiatry, 14*. https://doi.org/10.3389/fpsyt.2023.1190084

Guest, J. L., Adam, E., Lucas, I. L., Chandler, C. J., Filipowicz, R., Luisi, N., … Sullivan, P. S. (2021). Methods for authenticating participants in fully web-based Mobile app trials from the iReach project: Cross-sectional study. *JMIR mHealth and uHealth, 9*(8), Article e28232. https://doi.org/10.2196/28232

Hall, S. M., Havassy, B. E., & Wasserman, D. A. (1990). Commitment to abstinence and acute stress in relapse to alcohol, opiates, and nicotine. *Journal of Consulting and Clinical Psychology, 58*(2), 175–181. https://doi.org/10.1037//0022-006x.58.2.175

Hatcher, R. L., & Gillaspy, J. A. (2006). Development and validation of a revised short version of the working alliance inventory. *Psychotherapy Research, 16*(1), 12–25. https://doi.org/10.1080/10503300500352500

Ho, D. E., Imai, K., King, G., & Stuart, E. A. (2007). Matching as nonparametric preprocessing for reducing model dependence in parametric causal inference. *Political Analysis, 15*(3), 199–236. https://doi.org/10.1093/pan/mpl013

Hoffman, V., Flom, M., Mariano, T. Y., Chiauzzi, E., Williams, A., Kirvin-Quamme, A., … Perski, O. (2023). User engagement clusters of an 8-week digital mental health intervention guided by a relational agent (Woebot): Exploratory study. *Journal of Medical Internet Research, 25*, Article e47198. https://doi.org/10.2196/47198

Koopman, C., Pelletier, K. R., Murray, J. F., Sharda, C. E., Berger, M. L., Turpin, R. S., … Bendel, T. (2002). Stanford presenteeism scale: Health status and employee productivity. *Journal of Occupational and Environmental Medicine, 44*(1), 14–20. https://doi.org/10.1097/00043764-200201000-00004

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. *Journal of Affective Disorders, 114*(1), 163–173. https://doi.org/10.1016/j.jad.2008.06.026

Larsen, D. L., Attkisson, C. C., Hargreaves, W. A., & Nguyen, T. D. (1979). Assessment of client/patient satisfaction: Development of a general scale. *Evaluation and Program Planning, 2*(3), 197–207. https://doi.org/10.1016/0149-7189(79)90094-6

Lau-Barraco, C., & Dunn, M. E. (2008). Evaluation of a single-session expectancy challenge intervention to reduce alcohol use among college students. *Psychology of Addictive Behaviors, 22*(2), 168–175. https://doi.org/10.1037/0893-164x.22.2.168

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

Lucas, G. M., Gratch, J., King, A., & Morency, L.-P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior, 37*, 94–100. https://doi.org/10.1016/j.chb.2014.04.043

Magill, M., Martino, S., & Wampold, B. (2021). The principles and practices of psychoeducation with alcohol or other drug use disorders: A review and brief guide. *Journal of Substance Abuse Treatment, 126*, Article 108442. https://doi.org/10.1016/j.jsat.2021.108442

Magnan, E., Weyrich, M., Miller, M., Melnikow, J., Moulin, A., Servis, M., … Henry, S. G. (2024). Stigma against patients with substance use disorders among health care professionals and trainees and stigma-reducing interventions: A systematic review. *Academic Medicine, 99*(2), 221–231. https://doi.org/10.1097/acm.0000000000005467

Maricich, Y. A., Gerwien, R., Kuo, A., Malone, D. C., & Velez, F. F. (2021). Real-world use and clinical outcomes after 24 weeks of treatment with a prescription digital therapeutic for opioid use disorder. *Hospital Practice, 49*(5), 348–355. https://doi.org/10.1080/21548331.2021.1974243

Maricich, Y. A., Nunes, E. V., Campbell, A. N. C., Botbyl, J. D., & Luderer, H. F. (2022). Safety and efficacy of a digital therapeutic for substance use disorder: Secondary analysis of data from a NIDA clinical trials network study. *Substance Abuse, 43*(1), 937–942. https://doi.org/10.1080/08897077.2022.2060425

Maricich, Y. A., Xiong, X., Gerwien, R., Kuo, A., Velez, F., Imbert, B., … Williams, K. (2021). Real-world evidence for a prescription digital therapeutic to treat opioid use disorder. *Current Medical Research and Opinion, 37*(2), 175–183. https://doi.org/10.1080/03007995.2020.1846023

McCarthy, M. W., Lindsell, C. J., Rajasingham, R., Stewart, T. G., Boulware, D. R., & Naggie, S. (2024). Progress toward realizing the promise of decentralized clinical trials. *Journal of Clinical and Translational Science, 8*(1), Article e19. https://doi.org/10.1017/cts.2023.706

Mitchell, J. W., Chavanduka, T. M. D., Sullivan, S., & Stephenson, R. (2020). Recommendations from a descriptive evaluation to improve screening procedures for web-based studies with couples: Cross-sectional study. *JMIR Public Health and Surveillance, 6*(2), Article e15079. https://doi.org/10.2196/15079

Moore, J. B., Chieng, A., Pirner, M. C., Pajarito, S., Vogel, E. A., Bowdring, M. A., … Prochaska, J. J. (2025). Mitigating fraud in a fully decentralized clinical trial of a digital health intervention. *Annals of Behavioral Medicine, 59*(1). https://doi.org/10.1093/abm/kaaf047

Mullins, S. M., Suarez, M., Ondersma, S. J., & Page, M. C. (2004). The impact of motivational interviewing on substance abuse treatment retention: A randomized control trial of women involved with child welfare. *Journal of Substance Abuse Treatment, 27*(1), 51–58. https://doi.org/10.1016/j.jsat.2004.03.010

Naughton, F., Prevost, A. T., Gilbert, H., & Sutton, S. (2012). Randomized controlled trial evaluation of a tailored leaflet and SMS text message self-help intervention for pregnant smokers (MiQuit). *Nicotine & Tobacco Research, 14*(5), 569–577. https://doi.org/10.1093/ntr/ntr254

Parsons, J. T., Lelutiu-Weinberger, C., Botsko, M., & Golub, S. A. (2014). A randomized controlled trial utilizing motivational interviewing to reduce HIV risk and drug use in young gay and bisexual men. *Journal of Consulting and Clinical Psychology, 82*(1), 9–18. https://doi.org/10.1037/a0035311

Patel, V., Saxena, S., Lund, C., Thornicroft, G., Baingana, F., Bolton, P., … UnÜtzer, J. (2018). The Lancet commission on global mental health and sustainable development. *Lancet, 392*(10157), 1553–1598. https://doi.org/10.1016/s0140-6736(18)31612-x

Petrini, C., Mannelli, C., Riva, L., Gainotti, S., & Gussoni, G. (2022). Decentralized clinical trials (DCTs): A few ethical considerations. *Frontiers in Public Health, 10*, Article 1081150. https://doi.org/10.3389/fpubh.2022.1081150

Prochaska, J. J., Vogel, E. A., Chieng, A., Baiocchi, M., Maglalang, D. D., Pajarito, S., … Robinson, A. (2021). A randomized controlled trial of a therapeutic relational agent for reducing substance misuse during the COVID-19 pandemic. *Drug and Alcohol Dependence, 227*, Article 108986. https://doi.org/10.1016/j.drugalcdep.2021.108986

Prochaska, J. J., Vogel, E. A., Chieng, A., Baiocchi, M., Pajarito, S., Pirner, M., … Robinson, A. (2023). A relational agent for treating substance use in adults: Protocol for a randomized controlled trial with a psychoeducational comparator. *Contemporary Clinical Trials, 127*, Article 107125. https://doi.org/10.1016/j.cct.2023.107125

Prochaska, J. J., Vogel, E. A., Chieng, A., Kendra, M., Baiocchi, M., Pajarito, S., & Robinson, A. (2021). A therapeutic relational agent for reducing problematic substance use (Woebot): Development and usability study. *Journal of Medical Internet Research, 23*(3), Article e24850. https://doi.org/10.2196/24850

Rehm, J., & Shield, K. D. (2019). Global burden of disease and the impact of mental and addictive disorders. *Current Psychiatry Reports, 21*(2), 10. https://doi.org/10.1007/s11920-019-0997-0

Sawyer-Morris, G., Wilde, J. A., Molfenter, T., & Taxman, F. (2024). Use of digital health and digital therapeutics to treat SUD in criminal justice settings: A review. *Current Addiction Reports, 11*(1), 149–162. https://doi.org/10.1007/s40429-023-00523-1

Schaub, M., Sullivan, R., Haug, S., & Stark, L. (2012). Web-based cognitive behavioral self-help intervention to reduce cocaine consumption in problematic cocaine users: Randomized controlled trial. *Journal of Medical Internet Research, 14*(6), Article e166. https://doi.org/10.2196/jmir.2244

Shah, N., Velez, F. F., Colman, S., Kauffman, L., Ruetsch, C., Anastassopoulos, K., & Maricich, Y. (2022). Real-world reductions in healthcare resource utilization over 6 months in patients with substance use disorders treated with a prescription digital therapeutic. *Advances in Therapy, 39*(9), 4146–4156. https://doi.org/10.1007/s12325-022-02215-0

Skinner, H. A. (1982). The drug abuse screening test. *Addictive Behaviors, 7*(4), 363–371. https://doi.org/10.1016/0306-4603(82)90005-3

Sobell, L. C., Agrawal, S., Sobell, M. B., Leo, G. I., Young, L. J., Cunningham, J. A., & Simco, E. R. (2003). Comparison of a quick drinking screen with the timeline followback for individuals with alcohol problems. *Journal of Studies on Alcohol, 64*(6), 858–861. https://doi.org/10.15288/jsa.2003.64.858

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, 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

Steel, Z., Marnane, C., Iranpour, C., Chey, T., Jackson, J. W., Patel, V., & Silove, D. (2014). The global prevalence of common mental disorders: A systematic review and meta-analysis 1980–2013. *International Journal of Epidemiology, 43*(2), 476–493. https://doi.org/10.1093/ije/dyu038

Stein, M. D., Caviness, C. M., Morse, E. F., Grimone, K. R., Audet, D., Herman, D. S., … Anderson, B. J. (2018). A developmental-based motivational intervention to reduce alcohol and marijuana use among non-treatment-seeking young adults: A randomized controlled trial. *Addiction, 113*(3), 440–453. https://doi.org/10.1111/add.14026

Stuart, E. A., Lee, B. K., & Leacy, F. P. (2013). Prognostic score-based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. *Journal of Clinical Epidemiology, 66*(8 Suppl), S84–S90.e81. https://doi.org/10.1016/j.jclinepi.2013.01.013

Substance Abuse and Mental Health Services Administration, & Center for Behavioral Health Statistics and Quality. (2024). 2023 NSDUH Detailed Tables. https://www.samhsa.gov/data/report/2023-nsduh-detailed-tables.

Teitcher, J. E., Bockting, W. O., Bauermeister, J. A., Hoefer, C. J., Miner, M. H., & Klitzman, R. L. (2015). Detecting, preventing, and responding to "fraudsters" in internet research: Ethics and tradeoffs. *The Journal of Law, Medicine & Ethics, 43*(1), 116–133. https://doi.org/10.1111/jlme.12200

Tielman, M. L., Neerincx, M. A., van Meggelen, M., Franken, I., & Brinkman, W. P. (2017). How should a virtual agent present psychoeducation? Influence of verbal and textual presentation on adherence. *Technology and Health Care, 25*(6), 1081–1096. https://doi.org/10.3233/thc-170899

Torous, J., Nicholas, J., Larsen, M. E., Firth, J., & Christensen, H. (2018). Clinical review of user engagement with mental health smartphone apps: Evidence, theory and improvements. *Evidence-Based Mental Health, 21*(3), 116–119. https://doi.org/10.1136/eb-2018-102891

U.S. Department of Health and Human Services (HHS), & Office of the Surgeon General. (2016). Facing Addiction in America: The Surgeon General's Report on Alcohol, Drugs, and Health. https://www.ncbi.nlm.nih.gov/books/NBK424857/.

Underhill, C., Freeman, J., Dixon, J., Buzza, M., Long, D., Burbury, K., … Woollett, A. (2024). Decentralized clinical trials as a new paradigm of trial delivery to improve equity of access. *JAMA Oncology, 10*(4), 526–530. https://doi.org/10.1001/jamaoncol.2023.6565

Vaidyam, A. N., Wisniewski, H., Halamka, J. D., Kashavan, M. S., & Torous, J. B. (2019). Chatbots and conversational agents in mental health: A review of the psychiatric landscape. *The Canadian Journal of Psychiatry, 64*(7), 456–464. https://doi.org/10.1177/0706743719828977

Van Norman, G. A. (2021). Decentralized clinical trials: The future of medical product development?*. *JACC Basic Transl Sci, 6*(4), 384–387. https://doi.org/10.1016/j.jacbts.2021.01.011

Velez, F. F., Anastassopoulos, K. P., Colman, S., Shah, N., Kauffman, L., Murphy, S. M., … Maricich, Y. A. (2022). Reduced healthcare resource utilization in patients with opioid use disorder in the 12 months after initiation of a prescription digital therapeutic. *Advances in Therapy, 39*(9), 4131–4145. https://doi.org/10.1007/s12325-022-02217-y

Velez, F. F., Colman, S., Kauffman, L., Ruetsch, C., Anastassopoulos, K., & Maricich, Y. A. (2021). Comparison of healthcare resource utilization between patients who engaged or did not engage with a prescription digital therapeutic for opioid use disorder. *ClinicoEconomics and Outcomes Research, 13*, 909–916. https://doi.org/10.2147/ceor.S334274

Velez, F. F., Ruetsch, C., & Maricich, Y. (2021). Evidence of long-term real-world reduction in healthcare resource utilization following treatment of opioid use disorder with reSET-O, a novel prescription digital therapeutic. *Expert Review of Pharmacoeconomics & Outcomes Research, 21*(4), 519–520. https://doi.org/10.1080/14737167.2021.1939687

Vogel, E. A., Chieng, A., Robinson, A., Pajarito, S., & Prochaska, J. J. (2021). Associations between substance use problems and stress during COVID-19. *Journal of Studies on Alcohol and Drugs, 82*(6), 776–781. https://doi.org/10.15288/jsad.2021.82.776

Wright, J. H., Mishkind, M., Eells, T. D., & Chan, S. R. (2019). Computer-assisted cognitive-behavior therapy and Mobile apps for depression and anxiety. *Current Psychiatry Reports, 21*(7), 62. https://doi.org/10.1007/s11920-019-1031-2

Xiong, J., Lipsitz, O., Nasri, F., Lui, L. M. W., Gill, H., Phan, L., … McIntyre, R. S. (2020). Impact of COVID-19 pandemic on mental health in the general population: A systematic review. *Journal of Affective Disorders, 277*, 55–64. https://doi.org/10.1016/j.jad.2020.08.001

Xiong, X., Braun, S., Stitzer, M., Luderer, H., Shafai, G., Hare, B., … Maricich, Y. (2023). Evaluation of real-world outcomes associated with use of a prescription digital therapeutic to treat substance use disorders. *The American Journal on Addictions, 32*(1), 24–31. https://doi.org/10.1111/ajad.13346

Yardley, L., Spring, B. J., Riper, H., Morrison, L. G., Crane, D. H., Curtis, K., … Blandford, A. (2016). Understanding and promoting effective engagement with digital behavior change interventions. *American Journal of Preventive Medicine, 51*(5), 833–842. https://doi.org/10.1016/j.amepre.2016.06.015

Ybarra, M. L., Holtrop, J. S., Prescott, T. L., Rahbar, M. H., & Strong, D. (2013). Pilot RCT results of stop my smoking USA: A text messaging-based smoking cessation program for young adults. *Nicotine & Tobacco Research, 15*(8), 1388–1399. https://doi.org/10.1093/ntr/nts339