Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
A randomized controlled trial of a therapeutic relational agent for reducing substance misuse during the COVID-19 pandemic

Judith J. Prochaska a,*, Erin A. Vogel a, Amy Chieng a, Michael Baiocchi b, Dale Dagar Maglalang c, Sarah Pajarito c, Kenneth R. Weingardt c, Alison Darcy c, Athena Robinson c

a Stanford Prevention Research Center, Department of Medicine, Stanford University, USA
b Department of Epidemiology & Population Health, School of Medicine, Stanford University, USA
c Woebot Health, USA

ARTICLE INFO

Keywords:
COVID-19 pandemic
Artificial intelligence
Substance-related disorders
Randomized controlled trial
Relational conversational agent

ABSTRACT

Background: The COVID-19 pandemic disrupted access to treatment for substance use disorders (SUDs), while alcohol and cannabis retail sales increased. During the pandemic, we tested a tailored digital health solution, Woebot-SUDs (W-SUDs), for reducing substance misuse.

Methods: In a randomized controlled trial, we compared W-SUDs for 8 weeks to a waitlist control. U.S. adults (N = 180) who screened positive for substance misuse (CAGE-AID > 1) were enrolled June–August 2020. The primary outcome was the change in past-month substance use occasions from baseline to end-of-treatment (EOT). Study retention was 84%. General linear models tested group differences in baseline-to-EOT change scores, adjusting for baseline differences and attrition.

Results: At baseline, the sample (age M = 40, SD = 12, 65% female, 68% non-Hispanic white) averaged 30.2 (SD = 18.6) substance occasions in the past month. Most (77%) reported alcohol problems, 28% cannabis, and 45% multiple substances; 46% reported moderate-to-severe depressive symptoms. Treatment participants averaged 920 in-app text messages (SD = 892, Median = 701); 96% of completed lessons were rated positively; and 88% would recommend W-SUDs. Relative to waitlist, W-SUDs participants significantly reduced past-month substance use occasions (M = 9.1, SE = 2.0 vs. M = 3.3, SE = 1.8; p = .039). Secondary substance use and mood outcomes did not change significantly by group; however, reductions in substance use occasions correlated significantly with increased confidence and fewer substance use problems, cravings, depression and anxiety symptoms, and pandemic-related mental health effects (p-value < .05).

Conclusions: W-SUDs was associated with significant reductions in substance use occasions. Reduction in substance use occasions was associated with better outcomes, including improved mental health. W-SUDs satisfaction was high.

1. Background

The 2016 Surgeon General’s Report on Facing Addiction in America warned that community-wide disasters can disrupt treatment services and/or increase stress, precipitating relapse or heavier substance use (U. S. Department of Health and Human Services, 2016). Before the COVID-19 pandemic, addiction treatment in the U.S. reached less than one in five individuals who met diagnostic criteria for a substance use disorder (SUD) (Substance Abuse and Mental Health Services Administration, 2019). The COVID-19 pandemic disrupted treatment access further, while retailers selling alcohol and, in some states, cannabis, were deemed essential businesses. In March 2020, compared to the year prior, alcohol retail sales increased 54% and online alcohol sales increased 262% (The Nielsen Company, 2020). Surges in cannabis sales were recorded in March 2020, and sales growth for the year was 40% stronger than 2019 (Vangst et al., 2020).

The pandemic’s extended physical distancing demands coupled with the many social, medical, political, racial/ethnic and economic stressors...
of 2020 could escalate substance misuse, relapse, and the development of SUDs in at-risk individuals (Clay and Parker, 2020). In late June 2020, 13% of U.S. adults reported initiating or increasing substance use to cope with pandemic-related stress (Czeisler et al., 2020). Demands during and following the pandemic may overwhelm already strained addiction treatment services.

Digital health applications (apps) can reduce barriers to traditional SUD treatment including stigma, costs and inadequate insurance coverage, time demands, transportation needs, lack of access to qualified providers, and challenges navigating complex treatment systems (Giroux et al., 2017). Preliminary evidence suggests that digital SUD interventions can reduce substance misuse (Boumparis et al., 2019; Giroux et al., 2017). Scalable digital SUD interventions could lessen the population-level burden of SUDs.

A novel approach for scaling therapeutic interactions without requiring human involvement is the use of artificial intelligence (AI). Programmed therapeutic relational agents can deliver a coach-like or sponsor-like experience for in-the-moment treatment delivery (Vaidyam et al., 2019). Evidence suggests that people are more likely to disclose personal information when they believe the interactions are computer-rather than human-monitored (Lucas et al., 2014); and a strong therapeutic alliance can develop in the absence of face-to-face contact (Cook and Doyle, 2002), even with a non-human app (Berry et al., 2018). While most mental health apps have high drop-off (Baumel et al., 2019; Torous et al., 2018), a meta-analysis found that conversational text-based agents may increase engagement and enjoyment in digitized mental health care (Vaidyam et al., 2019). Preliminary research indicates text-based presentation results in higher program adherence than verbal presentation (Tielman et al., 2017).

In an 8-week, single-group pre/post evaluation, we recently evaluated Woebot for Substance Use Disorders (W-SUDs) (Prochaska et al., 2021). W-SUDs was adapted from Woebot, a therapeutic relational agent found to decrease depressive symptoms (Fitzpatrick et al., 2017). Delivered via a smartphone app, W-SUDs provides psychoeducation in cognitive behavioral therapy (CBT)-based behavior change tools and encourages mood tracking and behavioral pattern insight. Participants were found to engage with W-SUDs regularly, reporting high acceptability and affective bond formation. From pre- to post-treatment, significant reductions were reported in substance use occasions, hazardous alcohol and drug use, and cravings to use substances, while confidence to resist substance use urges increased (Prochaska et al., 2021). Depression and anxiety symptoms significantly declined. Given these encouraging findings, we aimed to evaluate the efficacy of W-SUDs relative to a waitlist group in a randomized controlled trial. The trial was conducted in summer 2020, when the need for novel in-hand treatment options was amplified due to the COVID-19 pandemic.

2. Methods

2.1. Study design

We evaluated W-SUDs in an 8-week, two-group randomized controlled trial with a waitlist comparison condition. The primary outcome was change in past-month substance use occasions self-reported at baseline and end-of-treatment (EOT). Secondary outcomes were changes in substance use problems, craving intensity, confidence to resist substance use urges, mood symptoms (depression, anxiety), pain, and pandemic-related mental health effects. Intervention engagement data were collected from the W-SUDs app. Acceptability was assessed within the app and in the EOT survey.

2.2. Recruitment and randomization

Study procedures were approved by Stanford Medicine’s Institutional Review Board. Participants were recruited June 25 to August 18, 2020 via Qualtrics Research Services, Stanford listservs, Facebook, and word-of-mouth. Recruitment materials sought adults with substance use concerns. Informed consent was required for eligibility screening and study participation.

Inclusion criteria were ages 18–65 years, residing in the U.S., scoring >1 on the CAGE-AID (Cut down, Annoyed, Guilty, Eye opener–Adapted to Include Drugs) (Brown and Rounds, 1995), owning a smartphone for accessing W-SUDs, available for the 8-week study, providing an email address, and English literacy. A cut-point of >1 on the CAGE-AID has a sensitivity of 70% and specificity of 85% for identifying individuals with SUDs (Brown and Rounds, 1995). Study exclusion criteria were prior Woebot use, anticipated pregnancy during the study period, history of severe alcohol or drug-related medical problems (e.g., delirium tremens, seizure, liver disease, hallucinations), opioid overdose requiring Narcan (naloxone), current opioid misuse without medication-assisted treatment, or past-year attempted suicide. To maximize study retention, we obtained participants’ phone numbers and provided up to U.S.$75 in Amazon gift cards for completing study assessments.

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 = 0.50, and 20% attrition, the target sample size was N = 80 per group or 160 participants. Fig. 1 shows the study consort diagram, which is missing some exclusion information lost due to a clerical error. Based on captured data, leading reasons for study exclusion were a CAGE-AID score of 0 or 1 (n = 1149) and being unable to commit to the 8-week study 2 (n = 401). All 193 eligible individuals provided informed consent for study participation; 189 completed the baseline survey; however, 5 were duplicates and 4 provided invalid contact information. In total, 180 participants were randomized to W-SUDs (n = 88) or the waitlist control (n = 92) group. Randomization was stratified on reported substance use problems on the SIP-AD measure (Blanchard et al., 2003), described below.

2.3. Study conditions

2.3.1. W-SUDs intervention

Woebot is a therapeutic relational agent that delivers CBT in the format of brief, daily text-based conversations through its own native app on iPhone and Android devices (Fitzpatrick et al., 2017). The app onboarding process introduces Woebot, explains the intended use of the device, how data are treated, and limitations of the program (e.g., that it is not a crisis service). Participants are encouraged to check-in with Woebot daily and informed that check-ins will invite weekly tracking of their mood and ratings of substance use craving and pain. Daily push notifications prompt users to check in. Woebot is introduced as a guided self-help coach and explicitly not human. Participants are educated that Woebot will offer emotional support, daily lessons, and tools for reducing substance use and manage cravings, all via the medium of text-based conversations. Conversations with Woebot are responsive in real-time to users’ inputs and crafted to offer the tool and or lesson for participants’ moment of need. After the app onboarding process, Woebot hosts the first conversation with the participant and demonstrates app navigation and tool application. Woebot was adapted for the treatment of SUDs (W-SUDs), drawing upon motivational interviewing principles, mindfulness training, dialectical behavior therapy, and CBT for relapse prevention (Prochaska et al., 2021). The W-SUDs intervention offers 66
psychoeducational lessons and psychotherapeutic skills and was designed as an 8-week treatment. Brief intervention can minimize dropout, a problem common to SUD treatment (Brorson et al., 2013). Given the sustained nature of the COVID-19 pandemic, content was added to W-SUDs to acknowledge pandemic-related stressors and isolation (Fig. 2).

2.3.2. Waitlist control
Participants randomized to the waitlist were provided access to W-SUDs after completing the 8-week study.

| Restrictions for Inclusion/Exclusion | n (Sample Size) |
|-------------------------------------|-----------------|
| CAGE (≥ 3)                          | 223             |
| Unable to commit to using app for 2 months | 401             |
| Severe relapse or psychiatric episode (≥ 5) | 305             |
| ETOH: severe shaking or hallucinations (≥ 3) | 194             |
| Non-substance use related hallucinations (≥ 3) | 237             |
| Non-substance use related paranoia (≥ 2) | 227             |
| Seizures or confusion when stopping ETOH/drugs (≥ 2) | 205             |
| No ETOH/drug use in past 12 months (≥ 2) | 160             |
| Drug-related medical problems (e.g., convulsions, GI bleed) (≥ 1) | 150             |
| Liver trouble (≥ 1) | 143             |
| No smartphone (≥ 1) | 129             |
| Using alcohol and not on medication-assisted therapy (≥ 1) | 122             |
| Opioid OD requiring Narcan/Naloxone (≥ 1) | 115             |
| Attempted suicide in past 12 mo (≥ 1) | 95              |
| Prior use of Woebot (≥ 1) | 91              |
| Anticipated pregnancy during study period (≥ 1) | 44              |
| Not English fluent (≥ 1) | 28              |
| Unable to complete study surveys (≥ 1) | 28              |
| Age 18 to 65 years (≥ 1) | 13              |
| Incomplete screening (≥ 1) | 11              |
| Not residing in UK (≥ 1) | 2               |

*Final number screened is an underestimate due to miscommunication with a third-party provider that had editor access to the survey.

**Fig. 1.** Woebot for Substance Use Disorders (W-SUDs) Randomized Controlled Trial Consort Diagram.

**Fig. 2.** Sample Screen Shots of Woebot for Substance Use Disorders (W-SUDs).

2.4. Assessments

Study assessments were collected at baseline and EOT and were self-reported online via Qualtrics. Baseline demographic items were sex, race/ethnicity, age, marital and employment status, and zip code. Ten lifetime psychiatric diagnoses were assessed with a write-in option for others. Prior therapy experience (ever and current) and current psychiatric medication use were assessed. Engagement in SUD treatment in the past month at baseline and past 2 months at EOT (i.e., during the study period) was reported for self-help groups, outpatient, intensive outpatient, residential/inpatient, and medications to reduce substance...
Pandemic-related life disruptions were queried including changes in employment and income; engagement in COVID-19 precautionary behaviors in the past 2 weeks (4-point scale: never to most of the time, see Table 1) (Lavioie and Bacon, 2020); difficulty accessing medical treatment or getting necessities (yes/no) (Lang, 2020); and COVID-19 diagnosis.

The specified primary outcome was the change from baseline to 8-weeks (EOT) in substance use occasions reported for the past 30 days. At baseline and EOT, the number of days used in the past 30 days was assessed for each of the following: alcohol, cannabis, cocaine, prescription stimulants, methamphetamine, inhalants, sedatives or sleeping pills, hallucinogens, street opioids, and prescription opioids. Because participants used different substances and not all drank alcohol, days of reported alcohol use and days of use for each assessed drug were summed to reflect past-month substance use occasions. Past-month substance use occasions could exceed 30 if individuals used more than one substance on a day. Previously, this measure was found sensitive to treatment effects (Prochaska et al., 2021). Prior research has supported mental health effects total score had a possible range of 0–22 (Lang, 2020).

Due to an online programming glitch, DAST-10 was not administered at EOT. Craving and interference in one’s daily life was reported for the past 4 weeks, from not at all (0) to extremely (4). The Brief Situational Confidence Questionnaire (Breslin et al., 2000) assessed self-confidence to resist the urge “right now” to drink heavily (self-defined) or use drugs in different situations from 0% “not at all confident” to 100% “totally confident.”

The W-SUDs app use metrics included the number of days used; messages sent; modules completed; and mood, craving, and pain ratings submitted. Lesson acceptability was indicated thumbs up or down.

2.5. Data analyses

Paired samples t-tests and chi-square tests compared the groups on baseline variables. EOT survey completion was 84% (152/180) and significantly higher for participants identifying as non-Hispanic white (89%) compared to other racial/ethnic groups (74%) and for participants with a college degree (89%) relative to high school degree (74%) (p-values < .01). Retention also was higher for participants who at baseline reported more pandemic-related mental health effects and lower levels of craving, pain, and pain interference (p-values < .05). Of the 17 treatment participants lost to follow-up, 16 never registered with the W-SUDs app. Fig. 1 shows rates of noncompliance by condition. Two participants randomized to waitlist (2%) downloaded Woebot (not SUD-tailored) from the app store during the study period. Analyses followed the intent-to-treat protocol.

Serious adverse events occurring during the 8-week study were assessed including substance use-related hospitalization, suicide attempt, alcohol or drug overdose, and severe withdrawal (e.g., delirium tremens). Positive endorsements were queried for timing, diagnosis, and resolution and reported to the study’s Data Safety Monitoring Board within 72 h of the team learning of the event.

At EOT, participants randomized to W-SUDs completed 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 Working Alliance Inventory-Short Revised (WAI-SR) with three 4-item subscales assessing development of an affective bond in treatment and level of agreement with treatment goals and treatment tasks (Hatcher and Gil-lassy, 2006).

W-SUDs app use metrics included the number of days used; messages sent; modules completed; and mood, craving, and pain ratings submitted. Lesson acceptability was indicated thumbs up or down.
## Table 1
Baseline Participant Demographic, Mental Health, and Substance Use Characteristics by Group.

| Variable                              | W-SUDS Group (N = 88) | Waitlist Control Group (N = 92) |
|---------------------------------------|-----------------------|---------------------------------|
|                                       | M (SD), Range         | % (n)                           | M (SD), Range         | % (n)                           |
| Age in Years                          | 40.8 (12.1), 19-65    |                                 | 39.8 (11.2), 18-64    |                                 |
| Biological Sex                        |                       |                                 |                       |                                 |
| Female                                | 65.9% (58)            |                                 | 64.1% (59)            |                                 |
| Male                                  | 34.1% (30)            |                                 | 35.9% (33)            |                                 |
| Hispanic/Latinx Ethnicity             | 13.6% (12)            |                                 | 13.0% (12)            |                                 |
| Race                                  |                       |                                 |                       |                                 |
| White                                 | 79.5% (70)            |                                 | 76.1% (70)            |                                 |
| Black/African American                | 12.5% (11)            |                                 | 12.0% (11)            |                                 |
| Asian American                        | 4.5% (4)              |                                 | 7.6% (7)              |                                 |
| American Indian or Alaska Native      | 2.3% (2)              |                                 | 2.2% (2)              |                                 |
| Multiracial                           | 1.1% (1)              |                                 | 2.2% (2)              |                                 |
| Marital Status                        |                       |                                 |                       |                                 |
| Married/Cohabiting/Partnered          | 56.8% (50)            |                                 | 55.4% (51)            |                                 |
| Divorced/Separated/Widowed            | 9.1% (8)              |                                 | 9.8% (9)              |                                 |
| Never Married/Single                  | 34.1% (30)            |                                 | 34.8% (32)            |                                 |
| Educational Degree                    |                       |                                 |                       |                                 |
| High School Degree                    | 29.5% (26)            |                                 | 29.3% (27)            |                                 |
| College Degree                        | 48.9% (43)            |                                 | 53.3% (49)            |                                 |
| Graduate or Post-graduate Degree      | 21.6% (19)            |                                 | 17.4% (16)            |                                 |
| Employment Status Prior to the Pandemic |                       |                                 |                       |                                 |
| Employed Part or Full-Time            | 75.0% (66)            |                                 | 76.9% (70)            |                                 |
| Unemployed (homemaker, student, retired) | 25.0% (22)            |                                 | 23.1% (21)            |                                 |
| Prior Therapy Experience              |                       |                                 |                       |                                 |
| Never in Therapy                      | 47.7% (42)            |                                 | 55.4% (51)            |                                 |
| In Therapy Formerly                   | 37.5% (33)            |                                 | 30.4% (28)            |                                 |
| Currently in Therapy                  | 14.8% (13)            |                                 | 14.1% (13)            |                                 |
| Prior Therapy Experience              |                       |                                 |                       |                                 |
| Psychiatric Medications CurrentlyPD   | 29.5% (26)            |                                 | 31.5% (29)            |                                 |
| Past 30 Day SUD Treatment             |                       |                                 |                       |                                 |
| Self-help groups                      | 13.0% (12)            |                                 | 11.4% (10)            |                                 |
| Outpatient Services                   | 13.0% (12)            |                                 | 12.5% (11)            |                                 |
| Intensive Outpatient / Residential    | 1.7% (3)              |                                 | 0.6% (1)              |                                 |
| Past 30 Day Medication Use for SUD Treatment | 5.7% (5)              |                                 | 8.7% (8)              |                                 |
| Lifetime Mental Health Conditions5    |                       |                                 |                       |                                 |
| Unipolar Depression                   | 30.7% (27)            |                                 | 26.1% (24)            |                                 |
| Bipolar/Manic Depression              | 8.0% (7)              |                                 | 14.1% (13)            |                                 |
| Anxiety Disorder                      | 43.2% (38)            |                                 | 37.0% (34)            |                                 |
| Other Mental Health Disorder          | 37.5% (33)            |                                 | 40.2% (37)            |                                 |
| Multiple Diagnoses                    | 39.8% (35)            |                                 | 38.0% (35)            |                                 |
| No History of Mental Illness          | 37.5% (33)            |                                 | 43.5% (40)            |                                 |
| Depressive Symptoms PHQ-8, possible range 0-24 | 8.4 (5.6), 0-20 | 9.3 (6.4), 0-24 | 45.5% (40) | 46.7% (43) |
| Anxiety Symptoms GAD 7 Anxiety, possible range 0-21 | 7.8 (5.7), 0.21 | 8.0 (6.2), 0.21 | 37.5% (33) | 41.3% (38) |
| Alcohol                               | 16.8 (9.4)            | 79 (89.8%)                       | 16.8 (10.1)           | 74 (80.4%)                       |
| Cannabis                              | 17.6 (10.3)           | 39 (44.3%)                       | 18.3 (11.4)           | 59 (64.1%)                       |
| Cocaine                               | 9.0 (5.6)             | 3 (3.4%)                         | 9.7 (8.5)             | 9 (9.8%)                         |
| Sedatives                             | 8.1 (7.2)             | 17 (19.3%)                       | 11.9 (12.2)           | 18 (19.6%)                       |
| Hallucinogens                         | 5.0 (2.8)             | 2 (2.3%)                         | 2.0 (2.0)             | 4 (4.3%)                         |
| Street Opioids                        | 3.5 (2.1)             | 2 (2.3%)                         | 3.0                  | 1 (1.1%)                         |
| Prescription Opioids                  | 13.9 (11.5)           | 7 (8.0%)                         | 16.7 (17.7)           | 9 (9.8%)                         |
| Past-Month Substance Use Occasions5   | 27.9 (16.8), 0-68     | 32.4 (20.0), 0-90                | 2.2 (2.1), 0-8        | 1.8 (1.8), 0-8                   |
| Drug Abuse Screening Test,* possible range 0-10 | 1.3 (1.8), 0-8 | 2.2 (2.1), 0-8 | 21.6% (19) | 41.3% (38) |
| Short Inventory of Problems – Alcohol and Drugs (SIP-AD), possible range 0-45 | 11.9 (10.8), 0-41 | 12.9 (10.4), 0-45 | 2.8 (0.8), 2.4 | 3.1 (0.8), 2.4 |
| CAGE-AID, possible range 2-4*         | 57.6 (23.8), 3.1-100  | 61.3 (21.8), 13.6-100            | 1.5 (1.1), 0-4        | 1.5 (1.2), 0-4                   |

*Participants responding “not applicable” are treated as missing.

**Mean days of use are calculated among those who report any use of that substance in the past 30 days.

N = 179 reporting.
Most (77%) identified alcohol as their problematic substance, followed by cannabis (28%), stimulants (12%), and other (14%, e.g., club drugs, opioids, sedatives). A majority (59%) reported a lifetime psychiatric diagnosis, most commonly anxiety disorders (40%) and unipolar depression (28%); 12% reported a SUD diagnosis; 40% reported multiple diagnoses. Few (14%) were currently in therapy; 31% were currently taking psychiatric medication. Engagement in SUD treatment in the past 30 days was reported by 28% of participants, including 7% taking medications to reduce substance use.

The full sample’s mean scores at baseline were 30 (SD = 19) for past-month substance use occasions; 12 (SD = 11) on the SIP-AD; 2 (SD = 2) on the DAST-10; 2 (SD = 1) for craving; 59% (SD = 23%) for confidence to resist substance use urges; and 12 (SD = 5) for pandemic-related mental health effects. Sample baseline mean scores on the PHQ-8 (SD = 9), GAD-7 (SD = 8), pain rating (SD = 32), and pain interference (SD = 1) were in the mild range. Baseline characteristics by group are shown in Tables 1 and 2. Waitlist participants had significantly higher CAGE-AID and DAST-10 scores than participants randomized to W-SUDs. All other baseline group comparisons were not statistically significant.

### 3.2. Changes in substance use and mental health outcomes by group

Table 3 summarizes the results of eight general linear models testing group differences in the primary outcome and secondary outcomes. From baseline to EOT, participants randomized to W-SUDs reduced their substance use occasions significantly more than the waitlist group ($F_{(1,148)} = 4.53, p = .035$). The estimated marginal mean reductions in past 30-day substance use occasions were -9.6 (SE = 2.3) for W-SUDs versus -3.9 (SE = 2.2) for the waitlist control. Confidence gains were on

### Table 2

**COVID-19 Pandemic-Related Measures by Group at Baseline.**

| Variable                                                                 | W-SUDS Group (N = 88) | Waitlist Group (N = 92) |
|--------------------------------------------------------------------------|-----------------------|-------------------------|
| Employment Status Due to the Pandemic                                    |                       |                         |
| Unaffected by the Pandemic                                               | 44.3% (39)            | 44.6% (41)              |
| Working from Home due to the Pandemic                                    | 27.3% (24)            | 25.0% (23)              |
| Employment Temporarily or Permanently Stopped                           | 12.5% (11)            | 18.5% (17)              |
| Added or Changed Jobs                                                    | 4.5% (4)              | 6.5% (6)                |
| Other / Not Applicable (e.g., retired, homemaker)                       | 11.4% (10)            | 5.4% (5)                |
| Designated an Essential Service Worker in the Pandemic                   |                       |                         |
| Yes                                                                      | 25.0% (22)            | 32.6% (30)              |
| No/Don’t Know                                                            | 75.0% (66)            | 67.4% (62)              |
| Changes in Income due to the Pandemic                                    |                       |                         |
| No Change in Income / Don’t Know                                        | 55.7% (49)            | 55.4% (51)              |
| Reduced Income                                                           | 35.2% (31)            | 32.7% (31)              |
| Increased Income                                                         | 9.1% (8)              | 10.9% (10)              |
| COVID Precautions Past 2 Weeks (% most of time)NA                        |                       |                         |
| Staying 6 ft. away from other People (n=179)                            | 71.6% (63)            | 71.4% (65)              |
| Staying/working at home (n=153)                                          | 58.9% (43)            | 57.5% (46)              |
| Avoiding large social gatherings (n=174)                                 | 83.5% (71)            | 80.9% (72)              |
| Avoiding small social gatherings (n=175)                                 | 40.7% (35)            | 61.8% (55)              |
| Avoiding any non-essential travel (n=170)                               | 56.6% (47)            | 56.3% (49)              |
| Avoiding bars (n=163)                                                   | 79.5% (62)            | 76.5% (65)              |
| Difficulty getting Necessities (food, medication, medical help) in the   |                       |                         |
| Past 2 Weeks due to the Pandemic                                        | 18.2% (16)            | 23.9% (22)              |
| Pandemic-related Mental Health Effects                                  |                       |                         |
| Preoccupation with the COVID-19 Virus (often/most of time)               | 62.5% (55)            | 50.0% (46)              |
| Worrying about Health of Self & Others (often/most of time)              | 52.3% (46)            | 59.8% (55)              |
| Stressfulness of Changes in Social Contacts (mod-to-extreme)             | 71.6% (63)            | 69.6% (64)              |
| Stressfulness of Changes in Lifestyle (mod-to-extreme)                   | 62.5% (55)            | 62.0% (57)              |
| Worsening of Mental/Emotional Health (mod-to-extreme)                    | 64.5% (57)            | 63.0% (58)              |
| Pandemic Effects on Sleep                                                |                       |                         |
| Sleeping a Lot More                                                      | 9.1% (8)              | 6.5% (6)                |
| Sleeping a Little More                                                  | 9.1% (8)              | 8.7% (8)                |
| No Change in Sleep                                                      | 31.8% (28)            | 33.7% (31)              |
| Sleeping a Little Less                                                  | 34.1% (30)            | 30.4% (28)              |
| Sleeping a Lot Less                                                     | 15.9% (14)            | 20.7% (19)              |
| Mean (SD)                                                               |                       |                         |
| Pandemic-related Mental Health Effects Score                             | 12.1 (4.3)            | 12.3 (5.1)              |
average two-fold greater for W-SUDs than the waitlist group, but not statistically significant (p = .175). Substance use problems, craving, depressive symptoms, anxiety, pain ratings, and pandemic-related mental health effects declined over time with no statistically significant difference between groups.

At EOT, 24% (37/152) of participants reported engaging in SUD treatment (other than W-SUDs) during the 2-month study period with no significant difference by study condition ($X^2(1,152) = 0.52, p = .752$). Engagement in SUD treatment during the study period, added as a covariate to the model of substance use occasions, was not significant ($F(1,147) = 0.31, p = .577$) and did not alter the treatment effect for W-SUDs ($F(1,147) = 4.53, p = .035$). In the models of secondary outcomes, SUD treatment engagement was significantly associated with only anxiety. Engaging in SUD treatment during the study period was associated with less reductions in anxiety (p < .001).

Reduction in substance use occasions was significantly associated with reduction in substance use problems and both measures were significantly associated with gains in confidence and decreased cravings, depressive and anxiety symptoms, and pandemic-related mental health effects (Table 4). Confidence gains were significantly associated with gains in confidence and decreased cravings, with reduction in substance use problems and both measures were correlated.

3.3. Serious adverse events

Of the 152 participants completing the EOT assessment, one reported a serious adverse event. The individual, randomized to waitlist, was hospitalized for alcohol detoxification. The DSMB deemed the event to be unrelated to study involvement.

3.4. W-SUDs use and acceptability

Table 5 summarizes treatment participants’ use of the W-SUDs app. Of the 88 participants randomized to W-SUDs, 77 (88%) registered with the app, using the app an average of 34.9 days (SD = 28.9, Median = 31), and sending on average 1052 in-app text messages (SD = 878, Median = 881). Including the 11 participants who did not register with W-SUDs in the denominator, the treatment group averaged 920 in-app text messages (SD = 892, Median = 701). Engagement ranged from 77 participants sending on average 172 messages (SD = 105, Median = 182) at week 1–50 participants averaging 126 messages (SD = 89, Median = 111) at week 8; 52 participants (68% of the 77 who registered) continued to message W-SUDs after 8 weeks. The correlation between the total count of in-app text messages and reduction in self-reported substance use occasions from baseline to EOT was $r = -.23, p = .06, n = 68$.

Most treatment participants (77%–84%) completed modules and

### Table 4

Correlations among change scores (baseline to end-of-treatment) for substance use occasions, substance use problems, confidence, depressive symptoms, and anxiety (N = 152).

| Dependent Variable | SIP-AD | Confidence | Craving | PHQ-8 | GAD-7 | Pandemic MH effects |
|--------------------|--------|------------|---------|-------|-------|---------------------|
| Substance Use Occasions | .34*** | −.29*** | .35*** | .26** | .23** | .20** |
| Substance Use Problems, SIP-AD | 1 | −.39*** | .55*** | .50*** | .49*** | .33** |
| Confidence Score | 1 | −.51*** | −.41*** | −.43*** | −.16 |
| Craving Past 7-days | 1 | .35*** | .41*** | .41*** | .24** |
| Depressive Symptoms, PHQ-8 | 1 | .65*** | .39*** | .39*** | .25** |
| Anxiety Symptoms, GAD-7 | 1 | .26** | .26** | .26** | .26** |

*p < .05, **p < .01, ***p < .001. MH = mental health.
submitted mood, craving, and pain ratings; 96% of completed lessons received a thumbs-up rating. The most frequently completed lesson, “COVID-19 Perspective,” received thumbs-up from 61 of 63 treatment participants (97%). Highest-rated lessons (100% thumbs up), completed by a majority of treatment participants, were on: urge surfing, labeling, and met most/almost all of their needs (74%). Additionally, 80% were effective engagement in SUD treatment services during the study period did not predict reductions in substance use occasions; affect the strength of treatment effect of W-SUDs; or correlate with measures of W-SUDs use, acceptability, feasibility, or satisfaction. W-SUDs appears useful and appropriate for adults with substance use concerns across the severity continuum.

The current findings support the utility of AI applications for reflecting tailored empathy, at scale, and support and extend a growing body of literature on the use of conversational agents (or chatbots) to support behavioral health. 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 (Giroux et al., 2017). The current findings also contribute to research on digital health interventions for delivering mental health and substance use treatment to communities affected by natural disasters or pandemic-related transitions such as gradual reopening of services. A novel moment in time, pandemic-related secular trends are possibly anomalous and unpredictable making it challenging to generalize findings. Though changes in secondary outcomes did not differ significantly by group, changes in secondary outcomes were significantly associated with improvements in the primary outcome of substance use occasions.

The W-SUDs app registration rate was 88%, better than our initial study (Prochaska et al., 2021) and comparable to or better than other successful mobile health interventions (Cliffe et al., 2020). More participants used W-SUDs early in treatment; however, app activity remained high through 8 weeks and for many participants extended past 8-weeks. W-SUDs improvements since our initial pilot include content refinement based on previous W-SUDs’ participant feedback and integration of a refined app-onboarding process. Compared to our pilot, app use metrics in the current study nearly doubled and acceptability ratings increased on all measures. As found in our previous study, W-SUDs scored highest on the WAI-SR on affective bond formation, which is notable given that Woebot is expressly non-human. That the sum count of total in-app texts was weakly correlated with reduction in substance use occasions is consistent with acknowledgement in the digital health space of the need for promoting “effective engagement” rather than simply more engagement to achieve intended behavior change outcomes (Yardley et al., 2016). In the current study, reduction in substance use occasions was associated with W-SUDs treatment acceptability, feasibility, and satisfaction.

Engagement in SUD treatment services during the study period did not predict reductions in substance use occasions; affect the strength of the treatment effect of W-SUDs; or correlate with measures of W-SUDs use, acceptability, feasibility, or satisfaction. W-SUDs appears useful and appropriate for adults with substance use concerns across the severity continuum.

The current findings support the utility of AI applications for reflecting tailored empathy, at scale, and support and extend a growing body of literature on the use of conversational agents (or chatbots) to support behavioral health. 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 (Giroux et al., 2017). The current findings also contribute to research on digital health interventions for delivering mental health and substance use treatment to communities affected by natural disasters or other widespread stressors (Ruggiero et al., 2006, 2012; Strudwick et al., 2021).

Prior studies of digital therapeutics, including programs addressing substance misuse, have indicated lower participation among people

### Table 5

| W-SUD App Usage among Treatment Participants (N = 88). |
|-----------------------------------------------------|
| **N (%) of participants** | **Engaged** |
| Mean (SD) | Median | IQR |
|-----------|---------|------|
| Days Used | 77 (88%) | 34.9 (28.9) | 31.0 | 12.0, 50.5 |
| In-App Text | 77 (88%) | 1052.0 | 881 | 351.5 |
| Messages | (878.0) | 1626.5 |
| Completed Modules | 74 (84%) | 20.0 (12.0) | 17.5 | 9.8, 31.0 |

**Submitted Ratings**

- Mood: 70 (80%) | 35.2 (27.6) | 32.0 | 14.8, 54.3
- Craving: 70 (80%) | 6.3 (2.8) | 7.5 | 4.0, 9.0
- Pain: 68 (77%) | 6.3 (2.7) | 7.5 | 5.0, 9.0

**Possible Range**

| Usage Rating Profile-Intervention (URP-I) | Possible Range | Mean (SD) |
|-----------------------------------------|---------------|-----------|
| Acceptability (6 items) | 6 – 36 | 29.0 (4.8) |
| Feasibility (6 items) | 6 – 36 | 32.0 (3.6) |
| Client Satisfaction Questionnaire (CSQ-8, 8 items) | 8 – 32 | 25.5 (5.0) |
| Working Alliance Inventory-Short Revised (WAI-SR) | 12 – 60 | 44.0 (12.5) |
| Agreement with treatment goals (4 items) | 4 – 20 | 14.4 (4.7) |
| Agreement with treatment tasks (4 items) | 4 – 20 | 13.9 (4.5) |
| Affective bond formation in treatment (4 items) | 4 – 20 | 15.7 (4.6) |

**In-app thumbs up ratings correlated with CSQ-8 (r = .52, p < .001) and WAI-SR (r = .44, p < .001). URP-I feasibility correlated with days of W-SUDs app use (r = .27, p = .030) and the number of in-app text messages (r = .30, p = .015), mood ratings (r = .34, p = .007) and completed modules (r = .25, p = .048).**

**4. Discussion**

In a randomized controlled trial conducted in summer 2020, during the COVID-19 pandemic, W-SUDs, a therapeutic relational agent, had decreased substance use occasions relative to a waitlist control group. Participants who reduced their substance use occasions from baseline to EOT reported increases in confidence to resist urges and fewer substance use problems, cravings, depressive and anxiety symptoms, and pandemic-related mental health effects. Hence, successfully reducing substance use, even if not achieving complete abstinence, was associated with fewer negative consequences and mental health concerns.

Improvements observed in secondary substance use and mood outcomes did not differ by group. Prior studies of W-SUDs and Woebot have found significant treatment effects on mood (Fitzpatrick et al., 2017; Prochaska et al., 2021). Failure to replicate may relate to the sample’s lower depression and anxiety scores at baseline or may relate to pandemic-related transitions such as gradual reopening of services. A novel moment in time, pandemic-related secular trends are possibly anomalous and unpredictable making it challenging to generalize findings. Though changes in secondary outcomes did not differ significantly by group, changes in secondary outcomes were significantly associated with improvements in the primary outcome of substance use occasions.

The W-SUDs app registration rate was 88%, better than our initial study (Prochaska et al., 2021) and comparable or better than other successful mobile health interventions (Cliffe et al., 2020). More participants used W-SUDs early in treatment; however, app activity remained high through 8 weeks and for many participants extended past 8-weeks. W-SUDs improvements since our initial pilot include content refinement based on previous W-SUDs’ participant feedback and integration of a refined app-onboarding process. Compared to our pilot, app use metrics in the current study nearly doubled and acceptability ratings increased on all measures. As found in our previous study, W-SUDs scored highest on the WAI-SR on affective bond formation, which is notable given that Woebot is expressly non-human. That the sum count of total in-app texts was weakly correlated with reduction in substance use occasions is consistent with acknowledgement in the digital health space of the need for promoting “effective engagement” rather than simply more engagement to achieve intended behavior change outcomes (Yardley et al., 2016). In the current study, reduction in substance use occasions was associated with W-SUDs treatment acceptability, feasibility, and satisfaction.

Engagement in SUD treatment services during the study period did not predict reductions in substance use occasions; affect the strength of the treatment effect of W-SUDs; or correlate with measures of W-SUDs use, acceptability, feasibility, or satisfaction. W-SUDs appears useful and appropriate for adults with substance use concerns across the severity continuum.

The current findings support the utility of AI applications for reflecting tailored empathy, at scale, and support and extend a growing body of literature on the use of conversational agents (or chatbots) to support behavioral health. 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 (Giroux et al., 2017). The current findings also contribute to research on digital health interventions for delivering mental health and substance use treatment to communities affected by natural disasters or other widespread stressors (Ruggiero et al., 2006, 2012; Strudwick et al., 2021).

Prior studies of digital therapeutics, including programs addressing substance misuse, have indicated lower participation among people
Trial registration

ClinicalTrials.Gov NCT04096001; http://clinicaltrials.gov/ct2/show/NCT04460027.

Declaration of Competing Interest

Sarah Pajarito, MA, Ken Weingardt, PhD, Alison Darcy, PhD, and Athena Robinson, PhD are Woebot Health employees. Other authors: None.

Acknowledgments

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 ongoing guidance from Leonardo Angeline, PhD and Victor Prikhodko, MBA at NIDA.

References

Azzopardi-Muscat, N., Sersen, K., 2019. Towards an equitable digital public health era: promoting equity through a health literacy perspective. Eur. J. Public Health 29 (Supplement 3), 13–17.
Baumel, A., Muench, F., Edan, S., Kane, J.M., 2019. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J. Med. Internet Res. 21 (9), e14567.
Berry, K., Salter, A., Morris, R., James, S., Bucci, S., 2018. Assessing therapeutic alliance in the context of mHealth interventions for mental health problems: development of the Mobile Agnew Relationship Measure (mARM) questionnaire. J. Med. Internet Res. 20 (4), e90.
Bickmore, T., Gruber, A., 2010. Relational agents in clinical psychiatry. Harv. Rev. Psychiatry 18 (2), 119–130.
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. Psychol. Addict. Behav. 17 (4), 328–331.
Boumparis, N., Schulte, M.H.J., Riper, H., 2019. Digital mental health for alcohol and substance use disorders. Curr. Treat. Options Psychiatry 6 (4), 352–366.
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. Behav. Res. Ther. 38 (12), 1211–1220.
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. J. Sch. Psychol. 51 (1), 81–96.
Brown, R.L., Rounds, L.A., 1995. Conjoint screening questionnaires for alcohol and other drug abuse: criterion validity in a primary care practice. Wis. Med. J. 94 (3), 135–145.
Clay, J.M., Parker, M.O., 2020. Alcohol use and misuse during the COVID-19 pandemic: a potential public health crisis? Lanceit Public Health 5 (5), e259.
Cliffe, B., Croker, A., Denne, M., Smith, J., Stallard, P., 2020. Digital cognitive behavioural therapy for insomnia for adolescents with mental health problems: feasibility open trial. JIMIR Ment. Health 7 (3), e14842.
Cook, J.E., Doyle, C., 2002. Working alliance in online therapy as compared to face-to-face therapy: preliminary results. Cyberpsychology Behav. 5 (2), 95–105.
Czeisler, M.E., Lane, R.I., Petersky, E., Wiley, J.F., Christensen, A., Najj, R., Weaver, M.D., Robbins, R., Facer-Childs, E.R., Barger, L.K., Czeisler, C.A., Howard, M.E., Rajaratnam, S.M.W., 2020. Mental health, substance use, and suicidal ideation during the COVID-19 pandemic—united States, June 24–30, 2020. MMWR 69 (32), 1049–1057.
Fitzpatrick, K.K., Darcy, A., Vierhile, M., 2017. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): a randomized controlled trial. JIMIR Ment. Health 4 (2), e19.
Giroux, I., Goullet, A., Mercier, J., Jacques, C., Bouchard, S., 2017. Online and mobile interventions for problem gambling, alcohol, and drugs: a systematic review. Front. Psychol. 8 (954).
Hatcher, R.L., Gillaspy, J.A., 2006. Development and validation of a revised short version of the working alliance inventory. Psychother. Res. 16 (1), 12–25.
Kroenke, K., Strine, T.W., Spitzer, R.L., Williams, J.B.W., Berry, J.T., Mokdad, A.H., 2009. The PHQ-8 as a measure of current depression in the general population. J. Affect. Disord. 114 (1), 163–173.
Lang, A.J., 2020. Complementary and Integrative Research (CAIR) Pandemic Impact Questionnaire (C-PIQ). San Diego, CA.
Larsen, D.L., Attkisson, C.C., Hargreaves, W.A., Nguyen, T.D., 1979. Assessment of client/patient satisfaction: development of a general scale. Eval. Program Plann. 2 (3), 197–207.

Author contributions

JJP, AR and AD designed the study and acquired funding. JJP and AR supervised the study recruitment and implementation and coordinated the research activities. EV and SP provided intellectual input to the study assessment. KRW aided in intervention content development. AC led institutional approvals. AC and SP led study recruitment, data accrual, and data management. JJP, AC, EAV, DDM and MB had access to the study data downloaded from Qualtrics. MB created the randomization and advised on data analyses. JJP, EAV and DDM performed the data analyses. JJP drafted the manuscript and incorporated feedback from all coauthors.
Lavoie, K., Bacon, S., 2020. International COVID-19 Awareness and Responses Evaluation (iCARE) Study. Montreal Behavioural Medicine Centre.

Lennox, R., Dennis, M.L., Scott, C.K., Funk, R., 2006. Combining psychometric and biometric measures of substance use. Drug Alcohol Depend. 83 (2), 95-103.

Lucas, G.M., Gratch, J., King, A., Morency, L.-P., 2014. It’s only a computer: virtual humans increase willingness to disclose. Comput. Human Behav. 37, 94–100.

Moss, H.B., Chen, C.M., Yi, H.Y., 2012. Measures of substance consumption among substance users, DSM-IV abusers, and those with DSM-IV dependence disorders in a nationally representative sample. J. Stud. Alcohol Drugs 73 (5), 820–828.

Potdar, R., Thomas, A., DiMeglio, M., Mohiuddin, K., Djibo, D.A., Laudanski, K., Dourado, C.M., Leighton, J.C., Ford, J.G., 2020. Access to internet, smartphone usage, and acceptability of mobile health technology among cancer patients. Support. Care Cancer 28 (11), 5455-5461.

Prochaska, J.J., Vogel, E.A., Chiou, A., Kendra, M., Baizachi, M., Pajarito, S., Robinson, A., 2021. A therapeutic relational agent for reducing problematic substance use (Woebot): development and usability study. J. Med. Internet Res. 23 (3), e24850.

Ruggiero, K.J., Resnick, H.S., Acierno, R., Coffey, S.P., Carpenter, M.J., Ruscio, A.M., Stephens, R.S., Kilpatrick, D.G., Stasiewicz, P.R., Roffman, R.A., Budzisalas, M., Galea, S., 2006. Internet-based intervention for mental health and substance use problems in disaster-affected populations: a pilot feasibility study. Behav. Ther. 37 (2), 190–205.

Ruggiero, K.J., Resnick, H.S., Paul, L.A., Gross, K., McCasley, J.L., Acierno, R., Morgan, M., Galea, S., 2012. Randomized controlled trial of an internet-based intervention using random-digit-dial recruitment: the Disaster Recovery Web project. Contemp. Clin. Trials 33 (1), 237–244.

Skinner, H.A., 1982. The drug abuse screening test. Addict. Behav. 7 (4), 363–371.

Spitzer, R.L., Kroenke, K., Williams, J.B.W., Lowe, B., 2006. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch. Intern. Med. 166 (10), 1092–1097.

U.S. Department of Health and Human Services, 2016. Facing Addiction in America: The Surgeon General’s Report on Alcohol, Drugs, and Health. HHS, Washington, DC.

Vaidyam, A.N., Wiśniewski, H., Halamka, J.D., Kashavan, M.S., Torous, J.B., 2019. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. Can. J. Psychiatry 64 (7), 456–464.

Vangst, LeafLink, Flowhub, 2020. The State of the Cannabis Industry 2020 (Accessed 4/8/2021). https://www.stateofthecannabisindustry.com/.

Yardley, L., Spring, B.J., Riper, H., Morrison, L.G., Crane, D.H., Curtis, K., Merchant, G.C., Naughton, F., Blandford, A., 2016. Understanding and promoting effective engagement with digital behavior change interventions. Am. J. Prev. Med. 51 (5), 833–842.