









A chatbot to improve adherence to internet-based cognitive–behavioural therapy among workers with subthreshold depression: a randomised controlled trial

Sakiko Yasukawa ¹, Taku Tanaka ¹, Kenji Yamane ¹, Ritsuko Kano ¹,
Masatsugu Sakata ², Hisashi Noma ³, Toshi A Furukawa ²,
Takuya Kishimoto ¹

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/bmjment-2023-300881>).

¹Technology Development Laboratories, Sony Corporation, Tokyo, Japan

²Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan

³Department of Data Science, The Institute of Statistical Mathematics, Tokyo, Japan

Correspondence to

Dr Takuya Kishimoto, Technology Development Laboratories, Sony Corporation, Tokyo, Japan; takuya.kishimoto@sony.com

Received 11 September 2023
Accepted 15 November 2023



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. Published by BMJ.

To cite: Yasukawa S, Tanaka T, Yamane K, *et al.* *BMJ Ment Health* 2024;**27**:1–7.

ABSTRACT

Background Internet-based cognitive–behavioural therapy (iCBT) is effective for subthreshold depression. However, iCBT has problems with adherence, especially when unaccompanied by human guidance. Knowledge on how to enhance adherence to iCBT without human involvement can contribute to improving the effectiveness of iCBT.

Objective This is an implementation study to examine the effect of an automated chatbot to improve the adherence rate of iCBT.

Methods We developed a chatbot to increase adherence to an existing iCBT programme, and a randomised controlled trial was conducted with two groups: one group using iCBT plus chatbot (iCBT+chatbot group) and one group not using the chatbot (iCBT group). Participants were full-time employees with subthreshold depression working in Japan (n=149, age mean=41.4 (SD=11.1)). The primary endpoint was the completion rate of the iCBT programme at 8 weeks.

Findings We analysed data from 142 participants for the primary outcome. The completion rate of the iCBT+chatbot group was 34.8% (24/69, 95% CI 23.5 to 46.0), that of the iCBT group was 19.2% (14/73, 95% CI 10.2 to 28.2), and the risk ratio was 1.81 (95% CI 1.02 to 3.21).

Conclusions Combining iCBT with a chatbot increased participants' iCBT completion rate.

Clinical implications Encouraging messages from the chatbot could improve participation in an iCBT programme. Further studies are needed to investigate whether chatbots can improve adherence to the programme in the long term and to assess their impact on depression, anxiety and well-being.

Trial registration number UMIN000047621.

BACKGROUND

Subthreshold depression has a high prevalence and risk for developing major depression (MDD).^{1 2} Depression leads to poor job performance and significant economic losses for employees.³ Therefore, treating subthreshold depression and primary

WHAT IS ALREADY KNOWN ON THIS TOPIC

- ⇒ Internet-based cognitive–behavioural therapy (iCBT) is an effective intervention for subthreshold depression.
- ⇒ Self-help intervention apps such as iCBT have problems with treatment adherence.
- ⇒ There is insufficient knowledge on how to improve adherence.

WHAT THIS STUDY ADDS

- ⇒ This study shows that a chatbot using a popular messaging application improves adherence to an iCBT programme.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- ⇒ Encouraging messages from messaging applications with a chatbot character may improve adherence to an iCBT programme and augment iCBT's effects.
- ⇒ Future studies should consider adding chatbots to iCBT to improve long-term treatment adherence; they should also check its impact on depression and anxiety.

prevention of depression in the workplace are important issues worldwide.

Evidence-based intervention, such as internet-based cognitive–behavioural therapy (iCBT), potentially improves the accessibility to the care of people with subthreshold depression that do not seek professional help. iCBT has been effective in reducing depressive symptoms and preventing MDD in adults including workers with subthreshold depression.^{4–6} Further, iCBT has the advantages of teleoperation, cost reduction and maintenance of treatment quality compared with face-to-face psychotherapy. These strengths will be particularly pertinent for the working population.

On the other hand, iCBT has its own challenges of low adherence and dropout, with reports of approximately 10% lower for iCBT compared with face-to-face.⁷ Approaches to increase adherence and retention need to be further studied,⁸ not only

because of the insufficient effectiveness of the treatment but also because of the risk for relapse.⁹ Recent reports have suggested that a combination of human and automated encouragement in iCBT can reduce treatment dropouts and increase effectiveness.¹⁰ Such strategies could be implemented in various ways, such as periodic encouraging emails, user interfaces that change in response to user inputs and in-person or artificial intelligence (AI)/chatbot feedback.¹¹

The number of digital mental health intervention studies using chatbots has increased in recent years,¹² and iCBT with chatbots has the advantage of promoting self-learning.¹³ The immediate responsiveness and human-like nature of chatbots may also benefit both human and automated encouragement, and we focused on the potential of chatbots to improve adherence. In a recent randomised controlled trial (RCT) of an iCBT with a chat-type AI agent called Woebot, conversational agent group showed reduced dropout rates by 20% over control group.¹⁴ However, no studies have yet directly examined the impact of chatbots only on promoting engagement and reducing dropout rates for iCBT.¹³ Therefore, we developed an original, lovable chatbot character that interacts with users individually according to their progress in iCBT, and studied the improvement in dropout and adherence by using the chatbot in conjunction with iCBT.

OBJECTIVE

The purpose of this study was to investigate whether a chatbot add-on to iCBT,^{15 16} which has already been shown to be effective in treating depression, could increase the completion rate of iCBT programmes for workers with subthreshold depression. Therefore, the primary outcome was defined as the completion rate of iCBT programmes.

This study takes two novel approaches: the first is that the chatbot did not provide a therapeutic intervention, but instead gave users a clear role as a supporter who encouraged them to continue with iCBT. This is the first study to design a directly observable impact of chatbots on improving engagement and reducing dropout and will contribute significantly to future research on the use of chatbots in iCBT. The second is that users do not converse with the chatbot within the iCBT application, but with a messaging service commonly used in Japan. This means that additional chatbots will be used without modifying iCBT application and the familiarity with the messaging service can lower psychological hurdles for daily operation.

METHODS

Trial design

This study is RCT that investigates the improvement in completion rates of iCBT that is combined with a chatbot among workers with subthreshold depression. The study was conducted in an open-label, stratified block randomisation manner, with two arms: a group using iCBT and a chatbot (iCBT+chatbot group) and a group not using a chatbot (iCBT group). This study reports completion rates at 8 weeks and effects on psychological measures such as depression, anxiety and well-being. We followed CONSORT (Consolidated Standards of Reporting Trials) guidelines¹⁷ and completed the CONSORT checklist (online supplemental file 1). The complete study design and procedures are listed in the study protocol (online supplemental file 2).

Participants

The inclusion criteria for participants were as follows: (1) full-time employees of Sony Group Corporation and Sony

Corporation; (2) residents of Japan; (3) aged 20–60 years; (4) owned a smartphone (iPhone or Android); (5) agreed to use the iCBT app; and (6) agreed to use fitbit (Fitbit), fitabase (Small Step Labs LLC, fitbit data collection service) and LINE (LINE Corporation, messaging service).

The exclusion criteria for participants were as follows: (1) inability to read and write Japanese texts; (2) undergoing follow-up and treatment by a psychiatrist or other mental health professional; (3) a total Patient Health Questionnaire-9 (PHQ-9)^{18 19} score at the time of application of 15 or above, or 10–14 with 2 or 3 on the 9th item (suicidal ideation); and (4) plan to retire (retire or change jobs to other companies) during the participation period.

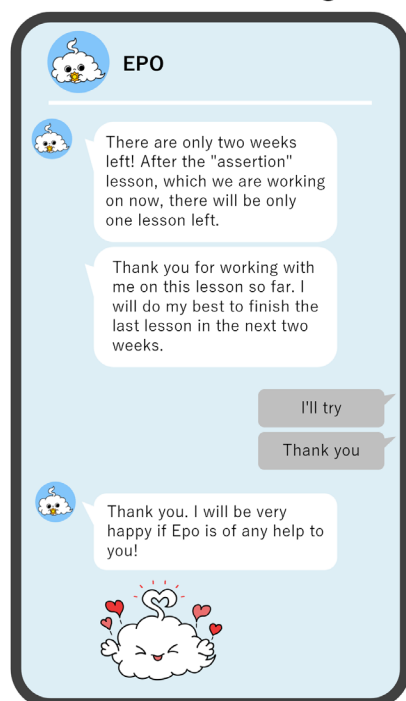
We recruited participants in April 2022, screened 334 applicants and invited 149 who met the eligibility criteria to an information session. They comprised 15 employees who scored four or less and 134 participants who scored between 5 and 9 or between 10 and 14 but scored 0 or 1 on the 9th item (suicidal ideation). The 149 applicants participated in an online information session and provided electromagnetic consent after a detailed explanation by a clinical research coordinator (CRC). Participants who completed IC (Informed Consent) were asked to complete the psychoeducation for the application during the orientation session. Participants who did not complete the psychoeducation during the orientation were asked to complete the psychoeducation lesson within a specified period. We discontinued the intervention for safety reasons if participants met the following conditions: a PHQ-9 score of 15 or higher or a score of 10–14 with a score of 2 or 3 on its 9th item (suicidal ideation) over 3 weeks.

Interventions

Internet-based cognitive-behavioural therapy

A smartphone app named 'Resilience Training SE (Sony Edition)' includes six iCBT components: psychoeducation (PE), behavioural activation (BA), self-monitoring (SM), cognitive restructuring (CR), assertiveness training (AT) and problem-solving (PS). The app was created for university students^{15 16} and it was necessary to modify the expressions related to school life and part-time work to those related to social life and work. During the orientation session, all participants first received psychological training on the importance of resilience to stress, CBT and weekly self-check (PHQ-9). According to PE, the app was programmed in the order of BA, SM, CR, AT and PS, each with an approximate completion time of 1 week. Each component consisted of a PE lesson describing a cognitive or a behavioural skill and a worksheet to practise what was learnt.^{15 16} Online supplemental figure 1 shows screenshots of the iCBT app. Participants were told that the test period would end after 8 weeks. The app opened weekly and prompted participants to answer the self-check. If a participant did not respond for several days, an automated email was sent to the participant, asking them to respond to the self-check. If a participant scored 15 or higher, or between 10 and 14 with a score of 2 or 3 on the 9th item (suicidal ideation), the administration sent an email advising them to contact psychological services such as occupational health. If the condition persisted for 3 consecutive weeks, the administration advised them to contact health services and informed them that the intervention would be discontinued. Each of the nine items constituting the PHQ-9 scores from 0 (not at all) to 3 (almost every day), with a range of 0–27 points. Scores of 10–14 are classified as moderate, 15–19 as moderately severe and 20–27 as severe. The administration sent participants

Conversational image



Sticker examples

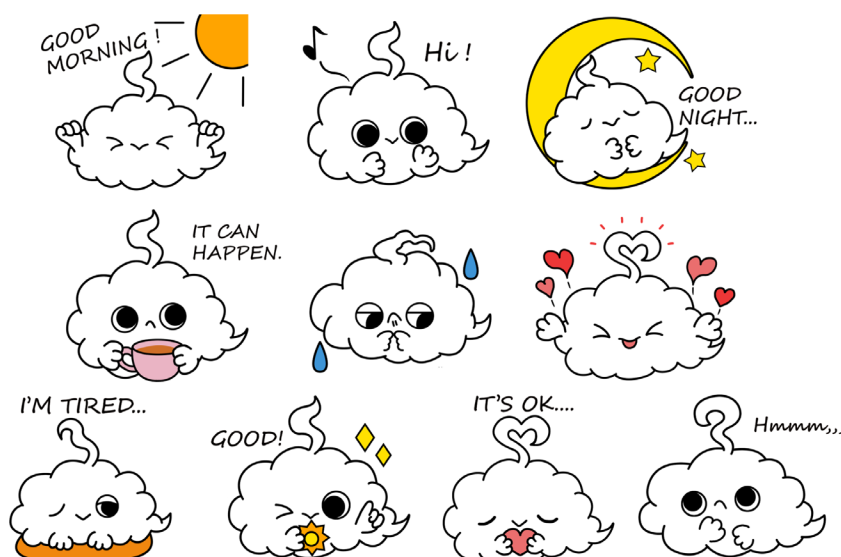


Figure 1 Chatbot character stickers, conversational images.

a web-based questionnaire during the information session and 4 and 8 weeks afterwards to collect their responses.

Chatbot

Figure 1 shows the conversation image and stamp of the chatbot named EPO, a cloud-like character designed for this study. The chatbot served as a human-like companion to participants in the iCBT+chatbot group, sending them personalised messages every morning and evening for 8 weeks, to encourage them to continue using the iCBT programme. Messages were sent through LINE, which they use for daily communication, to make the communication look more human-like. The chatbot system retrieved participants' learning progress from the online server at specific times in the morning and evening. It then retrieved messages that matched the designed progress scenarios from the dialogue system developed by Sony Group Corporation, and sent the personalised messages to each participant's messaging application. We developed a database of about 300 messages, including encouraging messages based on each participant's learning progress, surveys to adjust message frequency and daily messages to deepen communication with characters. In addition, the chatbot asked semi-open-ended questions, such as favourite lesson, to increase participants' engagement and encourage them to continue using the app. Online supplemental table 1 shows some examples of messages.

Fitbit

All participants received a fitbit charge4 (Fitbit) in advance and were instructed to wear it for 8 weeks to collect life log data. Participants could view their own sleep, step count and other information using the fitbit app. However, because the conditions on fitbit use were the same in both groups, fitbit use would not have influenced the group comparisons.

Outcomes

The primary outcome was the completion rate of the 'Resilience Training SE.' Completion rate for the app was defined as the percentage of participants who completed the lesson, which consisted of five components, to completion within 8 weeks (56 days) from the day after the end of PE. Completing the lesson to the end was defined as reading the entire lesson and completing a problem-solving component worksheet before the epilogue.

Secondary outcomes were changes from baseline to week 8 on the PHQ-9 measuring depression, Generalized Anxiety Disorder-7 (GAD-7)²⁰ measuring anxiety, CBT skills,²¹ The Satisfaction with Life Scale (SWLS)²² measuring well-being, WHO-5^{23 24} measuring well-being, Presenteeism Scale from WHO Health and Work Performance Questionnaire (Presenteeism)²⁵ measuring presenteeism, Work and Social Adjustment Scale (WSAS)²⁶ measuring social function, and Utrecht Work Engagement Scale (UWES)^{27 28} measuring work engagement.

Sample size

As dropout rates improved by 20% in prior studies using a conversational agent¹⁴ or using programmes with feedback features,²⁹ we expected that participants who used chatbot support would improve their completion rates by 20%. Assuming two-sided α -level of 0.05 and 80% power, a total sample size of 124 participants was required. We recruited 150 participants to ensure statistical power in case 20% of participants did not attend the orientation session or declined to participate after hearing the explanation.

Randomisation

We used permuted block randomisation stratified by pre-assessment PHQ-9 scores (4 or less, 5 or more). For allocation, researchers who were not involved in participant recruitment

created a random allocation sequence in advance using R V.4.1.1. After participants were educated about the study at an information session and informed consent was obtained, allocation was performed. Participants were assigned to the two groups following the order of the time stamps received by the consent acquisition system, using an automated allocation system. CRC was responsible for enrolling participants and assigning them to the intervention, and the researchers with the exception of the CRC was concealed from participant assignment.

Masking

Participants and researchers were not blinded to the intervention. Secondary outcomes were self-reported by participants.

Statistical analyses

We used SAS Studio V.5.2 (SAS Institute) for the statistical analyses. Participants were analysed in the full analysis set (FAS) according to the intention-to-treat principle, regardless of the actual intervention received or study discontinuation. For the primary analysis, the completion rates for the iCBT+chatbot group and the iCBT group were compared using the χ^2 test with a two-sided significance level of 5%.

As a secondary outcome, we analysed the PHQ-9 scores using the mixed-effects model for repeated measures (MMRM). We estimated mean differences of the change scores at 1–8 weeks from baseline between the two groups. In this analysis, we restricted the analysis set to participants with baseline PHQ-9 scores of 5 or higher to assess the impact of depression on participants with subthreshold depression. The MMRM modelled the change scores at 1–8 weeks from baseline as outcomes and included the intervention condition (with or without chatbot), time of assessment (as nominal variables), age, baseline PHQ-9 scores and interaction terms between intervention and time of assessment as fixed effects. An unstructured covariance structure was used for modelling of the correlations of outcome variables. We calculated the SMD using the SD of the baseline PHQ-9 score.

We also analysed GAD-7, CBT skills, SWLS, WHO-5, Presenteeism, WSAS and UWES using MMRM for the FAS. We used

the same model to analyse PHQ-9 scores for these outcomes. GAD-7, CBT skills, SWLS, WHO-5, WSAS and UWES were measured three times (at baseline, week 4 and week 8), and presenteeism was measured twice (at baseline and at week 8).

FINDINGS

Participants characteristics

Figure 2 shows the CONSORT diagram. We randomly assigned 149 applicants, 74 to the iCBT+chatbot group and 75 to the iCBT group. Of the 149 participants, 143 were included in the analysis as an FAS, excluding 4 participants who were unable to participate in the intervention due to system issues that prevented them from logging into the iCBT application and 2 participants who did not complete the psychoeducation within a specified time period. For the primary outcome—completion rate—we included 142 of the 143 participants (follow-up rate was 99.3%, 142/143), with the exception of 1 participant for whom we discontinued the intervention because of meeting protocol-based discontinuation criteria.

Table 1 shows the baseline demographic and clinical characteristics for each group, which were balanced.

Primary analyses

Table 2 shows the completion rates for each group. The iCBT+chatbot group showed a statistically significantly higher completion rate than the iCBT group ($p<0.05$).

Secondary analyses

Online supplemental table 2 shows the results of the analyses of secondary outcomes, PHQ-9, GAD-7, CBT skills, SWLS, WHO-5, Presenteeism, WSAS and UWES. The change in PHQ-9 at week 8 in the iCBT+chatbot group was -2.21 (95% CI -3.21 to -1.22 , ES= -0.75) and that in the iCBT group was -2.30 (95% CI -3.30 to -1.30 , ES= -0.78); both groups showed significant improvements. The mean difference for the PHQ-9 at 8 weeks was 0.08 (95% CI -1.33 to 1.5 , ES= 0.03), which was insignificant. As with the PHQ-9, both groups improved CBT skills other than PS, GAD-7, WHO-5, SWLS and Presenteeism.

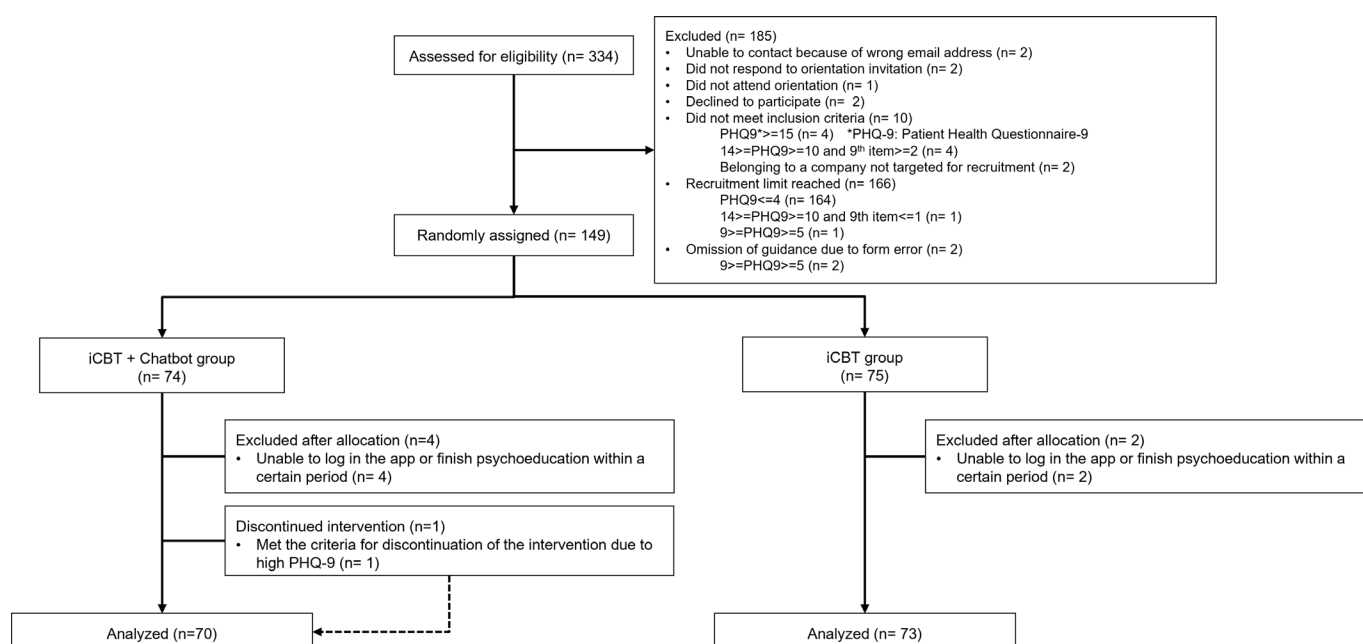


Figure 2 CONSORT (Consolidated Standards of Reporting Trials) diagram. iCBT, internet-based cognitive–behavioural therapy.

Table 1 Baseline characteristics of all participants (N=143) and by each component

Demographics	All participants	iCBT+chatbot group	iCBT group
n (%)	143 (100%)	70 (49%)	73 (51%)
Gender			
Women, n (%)	52 (36%)	27 (39%)	25 (34%)
Men, n (%)	90 (63%)	42 (60%)	48 (66%)
Unanswered, n (%)	1 (0.7%)	1 (1.4%)	0 (0%)
Age, mean (SD)	41.4 (11.1)	41.8 (10.4)	41.0 (11.9)
Marital status			
Marriage/cohabitation, n (%)	94 (66%)	48 (69%)	46 (63%)
Unmarried, n (%)	43 (30%)	19 (27%)	24 (33%)
Divorce/bereavement, n (%)	6 (4.2%)	3 (4.3%)	3 (4.1%)
Years of employment, mean (SD)	17.3 (11.4)	17.8 (10.6)	16.9 (12.2)
Occupation			
R&D, n (%)	85 (59%)	45 (64%)	40 (55%)
Other, n (%)	58 (41%)	35 (36%)	33 (45%)
Smartphone OS used in the research			
Android, n (%)	75 (52%)	38 (54%)	37 (51%)
iPhone, n (%)	68 (48%)	32 (46%)	36 (49%)
Clinical characteristics			
PHQ-9, mean (SD)	6.96 (3.57)	7.19 (3.67)	6.74 (3.48)
PHQ-9, mean (SD) excluded baseline PHQ-9≤4	8.50 (2.94)	8.69 (2.97)	8.31 (2.96)
GAD-7, mean (SD)	5.62 (3.49)	5.93 (3.46)	5.33 (3.51)
Cognitive and behavioural skills			
SM skills, mean (SD)	9.36 (3.15)	9.13 (3.37)	9.58 (2.92)
BA skills, mean (SD)	13.1 (3.58)	12.6 (3.60)	13.6 (3.52)
CR skills, mean (SD)	9.38 (3.23)	8.80 (3.18)	9.95 (3.20)
AT skills, mean (SD)	10.3 (3.75)	9.43 (3.34)	11.2 (3.94)
PS skills, mean (SD)	11.0 (2.88)	11.1 (2.73)	10.8 (3.03)
Well-being			
WHO-5, mean (SD)	12.0 (4.06)	11.9 (4.05)	12.2 (4.10)
SWLS, mean (SD)	21.7 (6.12)	21.5 (6.00)	21.9 (6.27)
Function			
Presenteeism, mean (SD)	5.92 (1.96)	5.60 (2.14)	6.23 (1.74)
WSAS, mean (SD)	15.8 (8.29)	15.9 (7.48)	15.6 (9.05)
Work engagement			
UWES, mean (SD)	27.6 (8.66)	26.8 (8.21)	28.3 (9.06)

AT, assertiveness training; BA, behavioural activation; CR, cognitive restructuring; GAD-7, Generalised Anxiety Disorder; iCBT, internet-based cognitive-behavioural therapy; PHQ-9, Patient Health Questionnaire-9; Presenteeism, WHO Health and Work Performance Questionnaire; PS, problem-solving; SM, self-monitoring; SWLS, Satisfaction with Life Scale; UWES, Utrecht Work Engagement Scale; WSAS, Work and Social Adjustment Scale.

No improvement was observed for PS in CBT skills in either group and for WSAS and UWES in the iCBT+chatbot group. As with PHQ-9, no significant difference was observed between the two groups for secondary outcomes except for UWES, but contrary to expectations, the iCBT group improved significantly more than the iCBT+chatbot group for UWES ($p<0.05$).

Table 2 Completion rates of iCBT

	iCBT completion rate (95% CI) (n)	Risk ratio (95% CI)	Risk difference (95% CI)	P value
iCBT+chatbot group (n=69 *)	34.8% (23.5% to 46.0%) (n=24)	1.81 (1.02 to 3.21%)	15.6% (1.19% to 30.0%)	0.0358
iCBT group (n=73)	19.2% (10.2% to 28.2%) (n=14)			

*One participant was excluded as the person met the protocol-based discontinuation criteria and was discontinued from the intervention. iCBT, internet-based cognitive-behavioural therapy.

Online supplemental table 3 shows the change from baseline in PHQ-9 at 1–8 weeks (adjusted and unadjusted).

Adverse events

We informed the Sony Bioethics Committee that one of the participants had been hospitalised during the study because of a traffic accident (which was judged unlikely to have been caused by this study). Apart from this, none of the participants had serious adverse events.

DISCUSSION

A messaging application with a lovable chatbot character significantly increased the completion rates of iCBT during the 8-week intervention period. The group that used the chatbot was 15.6 percentage points (95% CI 1.19 to 30.0) more likely to complete all lessons within 8 weeks than the group without the chatbot.

The chatbot that we developed for this study sent fully automated encouraging messages according to the individual's

programme progress. We designed the chatbot character to be friendly and expressive of emotions and used 10 chatbot character stickers along with the messages to combine automation with human-like qualities. A study of a digital smoking cessation programme reported that the addition of a chatbot more than doubled user engagement compared with a traditional programme.³⁰ Our study extended these findings to iCBT for subclinical depression.

The higher completion rate when using the chatbot could also be because the pace of the lessons was controlled by messages based on the user's individual iCBT progress. Online supplemental table 4 shows iCBT completion rates after ten weeks (8 weeks of intervention plus 2 weeks of allowance), suggesting that the chatbot helped control lesson pace. In the table, the risk difference at 8 weeks was 15.6%, whereas it decreased to 7.9% at 10 weeks, with no significant difference between the two groups. The lower risk difference at 10 weeks may simply reflect the fact that the messaging app was no longer active after 8 weeks. Or it may have been due to the email sent by the management office at 8 weeks to inform participants of the end of the study period. In the email, we wrote that we would allow participants to continue to use the application 2 weeks after notification of the end of the study. The fact that fewer participants in the iCBT+chatbot group completed the app within 2 weeks than in the iCBT group suggests that the chatbot was able to control the pace to complete the lessons within 8 weeks. The usability survey also showed that more than half (51%) of those who used the chatbot indicated that a message from the chatbot was the trigger for their use of iCBT.

However, the average 10-week completion rate for both groups was almost 40 percentage points lower than we had expected. The most likely reason for the lower than expected completion rate was that the amount of learning in the iCBT programme was too much for the duration of the experience, in addition to the participants' background of being busy with work and family duties. When asked in the usability survey why they were unable to complete the lessons, approximately 51% (38/75) indicated that they were too busy to take the time to use the application. We think 2 months is not enough time to learn the five components, and there are too many components when using a self-help app. The appropriate amount and duration of learning for the target audience is an important issue.

Secondary outcomes such as PHQ-9 showed improvement for subthreshold depressed employees, with or without chatbot use. Previous studies have shown that human or automated encouragement can reduce depression,¹⁰ since the iCBT+chatbot group had a higher completion rate than the iCBT group, we expected the chatbot to improve PHQ-9 scores more, but there was no significant difference in the amount of change between the two groups. Of the secondary outcomes, UWES improved in the iCBT group compared with the iCBT+chatbot group, but the reasons are unknown. We attribute the lack of significant differences between the groups in changes on the PHQ-9 to the following. First, participation in the study and initial experience of PE and BA may have been sufficient to improve depression. Online supplemental table 4 shows that the PHQ-9 of both groups had already improved by more than -2 points at 2-3 weeks. Second, the weekly self-check and visualisation of the life log by wearing the fitbit all the time, performed independently from the classes, might have contributed to the improvement of depression (68% of all respondents were satisfied with the use of the fitbit). Online supplemental table 5 shows the survey response rate for each week for the secondary outcomes, which was high, ranging from 70% to 90%. The effect of self-check

on improving depression has been demonstrated in previous studies.¹⁰ Further research is needed to examine whether there is a difference in depression when the completion rate is further increased and whether the contribution of chatbots to lesson progress control improves outcomes in the long term rather than the short term.

CONCLUSION AND IMPLICATIONS

This is the first RCT that attempts to examine whether a human-like automated guidance function enabled by a chatbot could increase adherence to an iCBT programme for subthreshold depression. The results suggest that the personalised messages sent by the chatbot helped participants control their pace in attending lessons and improve programme adherence without human guidance. Despite the improved completion rates, and contrary to expectations, PHQ-9 and GAD-7 scores at 8 weeks were similarly improved in both groups with and without the use of the chatbot.

This study has two limitations. First, some users felt that the iCBT programme, which involved PE and five components for 2 months, required too much learning and, thus, the completion rate was lower than expected. Second, because the study was conducted with Sony employees as an in-house study, it may be that our sample had higher digital literacy than the general population. Even among such people, a chatbot messaging app helped increase the adherence. Lastly, this study was not designed to test the chatbot's efficacy in improving subthreshold depression symptoms.

Future studies should review the iCBT programme's structure and its experience and continuously improve the chatbot to enable it to eventually promote clinical indicators, such as PHQ-9 scores. For example, we believe that in addition to progress-based messages, individualised messages that capture the user's personality and characteristics could provide more detailed support.

Contributors TT and TK conceived the study. TT, TK, KY, RK and SY designed the study. TAF and HN supervised the study design. RK constructed the chatbot intervention. KY constructed the system for linking iCBT and the chatbot. MS supervised chatbot messages and intervention pacing. TT, TK, KY, RK and SY managed the study and acquired the data. SY analysed the data. HN supervised the statistical analysis. SY wrote the first draft of the manuscript, and all the other authors revised the text critically and approved the final manuscript. TK acts as guarantor, is fully responsible for the data and content of the manuscript, and will manage the correspondence related to the article.

Funding The study was funded by Sony Group Corporation and Sony Corporation.

Competing interests TK, SY, TT, KY and RK are employees of Sony Corporation. TAF, HN and MS acted as advisors during the study. TAF reports personal fees from Boehringer Ingelheim, DT Axis, Kyoto University Original, MSD, Shionogi and Sony, and a grant from Shionogi, beyond the submitted work. In addition, TAF has patents 2020-548587 and 2022-082495 pending, and intellectual properties for Kokoro-app licensed to Mitsubishi-Tanabe. HN reports personal fees from Boehringer Ingelheim, Kyowa Kirin, Toyota Motor Corporation, GlaxoSmithKline, Ono Pharmaceutical, Sony and Terumo beyond the submitted work. MS reports personal fees from Sony beyond the submitted work. All other authors declare no conflict of interest.

Patient consent for publication Not applicable.

Ethics approval This study was approved by Sony Bioethics Committee (#21-17-0001). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. After the publication of the primary findings, the deidentified and completely anonymised individual participant-level dataset will be posted on the UMIN-ICDR website (<https://www.umin.ac.jp/icdr/index-j.html>) for access by qualified researchers.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those

of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Sakiko Yasukawa <http://orcid.org/0000-0003-2375-609X>

Taku Tanaka <http://orcid.org/0009-0004-5910-7387>

Kenji Yamane <http://orcid.org/0009-0005-1756-2987>

Ritsuko Kano <http://orcid.org/0009-0000-7273-4191>

Masatsugu Sakata <http://orcid.org/0000-0002-5358-5263>

Hisashi Noma <http://orcid.org/0000-0002-2520-9949>

Toshi A Furukawa <http://orcid.org/0000-0003-2159-3776>

Takuya Kishimoto <http://orcid.org/0009-0001-1624-0602>

REFERENCES

- Cuijpers P, Smit F. Subthreshold depression as a risk indicator for major depressive disorder: a systematic review of prospective studies. *Acta Psychiatr Scand* 2004;109:325–31.
- Cuijpers P, de Graaf R, van Dorsselaer S. Minor depression: risk profiles, functional disability, health care use and risk of developing major depression. *J Affect Disord* 2004;79:71–9.
- Kessler RC. The costs of depression. *Psychiatr Clin North Am* 2012;35:1–14.
- He R, Wei J, Huang K, et al. Nonpharmacological interventions for subthreshold depression in adults: A systematic review and network meta-analysis. *Psychiatry Res* 2022;317:50165-1781(22)00488-7.
- Reins JA, Buntrock C, Zimmermann J, et al. Efficacy and Moderators of Internet-based interventions in adults with subthreshold depression: an individual participant data meta-analysis of randomized controlled trials. *Psychother Psychosom* 2021;90:94–106.
- Zhou T, Li X, Pei Y, et al. Internet-based cognitive behavioural therapy for subthreshold depression: a systematic review and meta-analysis. *BMC Psychiatry* 2016;16:356.
- Kambeitz-Ilanovic L, Rzyavaya U, Völkel L, et al. A systematic review of Digital and face-to-face cognitive behavioral therapy for depression. *NPJ Digit Med* 2022;5:144.
- Melville KM, Casey LM, Kavanagh DJ. Dropout from Internet-based treatment for psychological disorders. *Br J Clin Psychol* 2010;49:455–71.
- Klein EB, Stone WN, Hicks MW, et al. Understanding dropouts. *J Ment Health Couns* 2003;25:89–100.
- Furukawa TA, Suganuma A, Ostinelli EG, et al. Dismantling, Optimising, and Personalising Internet cognitive behavioural therapy for depression: a systematic review and component network meta-analysis using individual participant data. *Lancet Psychiatry* 2021;8:500–11.
- Denecke K, Schmid N, Nüssli S. Implementation of cognitive behavioral therapy in E-mental health Apps. *J Med Internet Res* 2022;24:e27791.
- Milne-Ives M, de Cock C, Lim E, et al. The effectiveness of artificial intelligence conversational agents in health care. *J Med Internet Res* 2020;22:e20346.
- Boucher EM, Harake NR, Ward HE, et al. Artificially intelligent Chatbots in Digital mental health interventions: a review. *Expert Rev Med Devices* 2021;18:37–49.
- Fitzpatrick KK, Darcy A, Vierhile M. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Ment Health* 2017;4:e19.
- Uwatoko T, Luo Y, Sakata M, et al. Healthy campus trial: a Multiphase optimization strategy (MOST) fully factorial trial to optimize the Smartphone cognitive behavioral therapy (CBT) App for mental health promotion among university students: study protocol for a randomized controlled trial. *Trials* 2018;19:353.
- Sakata M, Toyomoto R, Yoshida K, et al. Components of Smartphone cognitive-behavioural therapy for subthreshold depression among 1093 university students: a factorial trial. *Evid Based Ment Health* 2022;25:e18–25.
- Schulz KF, Altman DG, Moher D, et al. Statement: updated guidelines for reporting parallel group randomised trials. *BMC Med* 2010;8:18.
- Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001;16:606–13.
- Muramatsu K, Miyaoka H, Kamijima K, et al. The patient health questionnaire, Japanese version: validity according to the mini-International neuropsychiatric interview-plus. *Psychol Rep* 2007;101:952–60.
- Spitzer RL, Kroenke K, Williams JBW, et al. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med* 2006;166:1092–7.
- Sakata M, Toyomoto R, Yoshida K, et al. Development and validation of the cognitive behavioural therapy skills scale among college students. *Evid Based Ment Health* 2021;24:70–6.
- Diener E, Emmons RA, Larsen RJ, et al. The satisfaction with life scale. *J Pers Assess* 1985;49:71–5.
- Awata S, Bech P, Koizumi Y, et al. Validity and utility of the Japanese version of the WHO-five well-being index in the context of detecting suicidal ideation in elderly community residents. *Int Psychogeriatr* 2007;19:77–88.
- Awata S, Bech P, Yoshida S, et al. Reliability and validity of the Japanese version of the world health organization-five well-being index in the context of detecting depression in diabetic patients. *Psychiatry Clin Neurosci* 2007;61:112–9.
- Kessler RC, Barber C, Beck A, et al. The world health organization health and work performance questionnaire (HPQ). *J Occup Environ Med* 2003;45:156–74.
- Mundt JC, Marks IM, Shear MK, et al. The work and social adjustment scale: a simple measure of impairment in functioning. *Br J Psychiatry* 2002;180:461–4.
- Shimazu A, Schaufeli WB, Kosugi S, et al. Work engagement in Japan: validation of the Japanese version of the Utrecht work engagement scale. *Applied Psychology* 2008;57:510–23. 10.1111/j.1464-0597.2008.00333.x Available: <https://iaap-journals.onlinelibrary.wiley.com/toc/14640597/57/3>
- Schaufeli W, Bakker A. *UWES - Utrecht Work Engagement Scale: Test Manual*. Utrecht University, Department of Psychology, 2003.
- Torous J, Lipschitz J, Ng M, et al. Dropout rates in clinical trials of Smartphone Apps for depressive symptoms: A systematic review and meta-analysis. *J Affect Disord* 2020;263:413–9.
- Perski O, Crane D, Beard E, et al. Does the addition of a supportive Chatbot promote user engagement with a smoking cessation App? an experimental study. *Digit Health* 2019;5:2055207619880676.