










The preventive effect of internet-based cognitive behavioral therapy for prevention of depression during pregnancy and in the postpartum period (iPDP): a large scale randomized controlled trial

Daisuke Nishi, MD, PhD ^{1,2*} Kotaro Imamura, PhD ¹ Kazuhiro Watanabe, PhD ^{1,3}
Erika Obikane, MD, MPH, PhD ¹ Natsu Sasaki, MD, PhD ¹ Naonori Yasuma, MD, PhD ⁴ Yuki Sekiya, PhD ¹
Yutaka Matsuyama, PhD ⁵ and Norito Kawakami, MD, PhD ¹

Background: Prevention of perinatal depression beginning from the antenatal period is essential. Therefore, this study aimed to investigate the effectiveness of recently developed internet-delivered cognitive behavioral therapy (iCBT) for preventing the onset of a major depressive episode (MDE) in the third trimester and at 3 months postpartum.

Methods: This is a two-arm, parallel-group, general-information controlled, randomized controlled trial. Participants were 5017 pregnant women at 16–20 weeks' gestation without MDE at baseline. They were randomly assigned to an iCBT (intervention; $n = 2509$) or general-information (control; $n = 2508$) group, stratified by psychological distress at baseline. The primary outcomes were the numbers of new MDE onsets, measured using the World Health Organization Composite International Diagnostic Interview 3.0, at 32 weeks' gestation and at 3 months postpartum.

Results: New MDE onset was reported by 59 participants (2.35%) in the intervention group and 73 (2.91%) in the

control group during follow-up. Compared with the control group, the hazard ratio (HR) of MDE in the intervention group was 0.85 (95% CI 0.61–1.20), which was not significantly different. Among participants who scored between 5 and 8 on K6 at baseline, 10 (1.37%) in the intervention group reported new onset of MDE, compared with 28 (3.81%) in the control group, and the HR of MDE was 0.38 (95%CI 0.19–0.79).

Conclusions: No intervention effect was found for iCBT in preventing new onset of perinatal MDE. iCBT might prevent perinatal depression only among pregnant women with subthreshold depressive symptoms. Trial registration: UMIN000038190.

Keywords: antenatal depression, postpartum depression, prevention, smartphone-based cognitive behavioral therapy.

<http://onlinelibrary.wiley.com/doi/10.1111/pcn.13458/full>

A key public health challenge is the prevention of maternal depression in the perinatal period.¹ The prevalence of antenatal depression has been reported as 7.4%, 12.8%, and 12.0% in the first, second, and third trimesters, respectively,² while the prevalence of postpartum depression worldwide has been estimated at 17.7%, with substantial variation across countries.³ Depression in the antenatal period has been linked to poor nutrition, use of alcohol, tobacco, or other harmful substances, lack of antenatal care seeking, self-harm or attempted suicide, and postpartum depression. Meanwhile, postpartum depression has been linked to child abuse, both physical and psychological.^{4,5} Antenatal and postpartum depression can impact the development of children from the fetal stage to adolescence and can also contribute to paternal depression.^{6–11} For these reasons, it is critical to prevent depression from the perinatal period to the antenatal period.

Psychological interventions are recommended as the most effective approach for preventing perinatal depression.¹² A systematic review and

meta-analysis reported that the effect size of antenatal psychological intervention on universal prevention (a measure that is desirable for everyone) based on cognitive behavioral therapy (CBT) was 0.53 in an antenatal period and 0.45 in a postpartum period.¹³ In particular, fully automated internet-delivered CBT (iCBT) has advantages over face-to-face or guided iCBT because it is accessible, anonymous, and cost-effective.¹⁴

However, evidence of the effect of fully automated iCBT on perinatal depression has not been established. To date, only two randomized controlled trials (RCTs) have been conducted to investigate the effect of iCBT on perinatal depressive symptoms as universal prevention.^{15,16} One study showed a significant effect and another study failed to show a statistically significant effect, and no previous studies have assessed the number of major depressive episodes (MDEs) as outcomes. Reducing dropout and increasing adherence are critical issues. The optimal number and duration of iCBT sessions have yet

¹ Department of Mental Health, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

² Department of Public Mental Health Research, National Institute of Mental Health, National Center of Neurology and Psychiatry, Tokyo, Japan

³ Department of Public Health, Kitasato University School of Medicine, Sagamihara, Japan

⁴ Ageonomori Clinic, Ageo, Saitama, Japan

⁵ Department of Biostatistics, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

* Correspondence: Email: d-nishi@m.u-tokyo.ac.jp

to be established. For instance, a previous study program involved 44 10-min sessions, of which 11 were during the antenatal period and 33 during the postnatal period,¹⁶ but it might be difficult for most pregnant women to participate in such a large number of sessions. According to a recent meta-analysis, the number of iCBT sessions for treating subthreshold depression ranged from 6 to 12, with 6 being the most common.¹⁷

Moreover, another study reported that iCBT had a preventive effect on subthreshold depressive symptoms.¹⁸ The effect of iCBT may differ depending on the degree of psychological distress at baseline, but this has not been fully investigated.

The primary aim of this large-scale RCT was to investigate the effectiveness of a recently developed smartphone-based automated iCBT program comprising six sessions of 5–10 min aimed at preventing the onset of MDE in the third trimester as well as at 3-months postpartum in pregnant women in the second trimester at the time of the study. The secondary aim was to clarify the difference of the effect based on psychological distress at baseline.

Methods

Trial design

The study protocol¹⁹ was registered in the Clinical Trials Registry (UMIN-CTR ID: UMIN000038190) of the University Hospital Medical Information Network (UMIN), and the study was conducted as a two-arm parallel-group RCT. Participants were allocated to the intervention and control groups at a ratio of 1:1. Pregnant women using the app Luna Luna Baby by MTI Ltd. were recruited. The app provides information on fetal development as well as expected changes in pregnant women according to their gestational age. The participants registered the date of their last menstruation before pregnancy in the app, and this date was used to determine the number of weeks of pregnancy.

Follow-up assessments were conducted in the third trimester (32 weeks' gestation) and at 3 months postpartum. This paper follows the Consolidated Standards of Reporting Trials (CONSORT) for RCTs.²⁰

The Research Ethics Review Board of the Graduate School of Medicine/Faculty of Medicine, the University of Tokyo (2019150NI) approved the study plan.

Participants

Pregnant women (primiparous and multiparous) who had a Luna Luna Baby user ID were invited to participate in the RCT if they satisfied the following criteria¹: 20 years of age or older,² 16–20 weeks of gestation,³ no diagnosis of an MDE in the past month by the web-based self-administered version of the WHO Composite International Diagnostic Interview 3.0 (WHO-CIDI 3.0),²³ and⁴ no diagnosis of lifetime bipolar disorder by the WHO-CIDI 3.0.

Recruitment

MTI Ltd. sent an invitation message *via* the app to potentially eligible pregnant women, explaining the study and providing information on the eligibility criteria. After reading the explanation of the study, potential participants were invited to participate in the study, provide their consent in the app, and then complete and return the baseline survey. MTI Ltd. also sent invitations to the study participants to complete follow-up assessments. The recruitment of participants was limited to users of the app.

Participants assigned to the intervention group were asked to continue participating in the intervention program until 32 weeks of gestation. They were asked not to share the contents of the program *via* social media. A popup message was used to remind the participants to complete the program if they have not already done so. The intervention programs were terminated at 32 weeks of gestation.

Participants assigned to the control group did not receive any intervention during the baseline and follow-up periods. General information about mental health during pregnancy was provided to participants in both the intervention and control groups. Specifically, when

the app user opens the app, pregnant women can obtain information on the daily condition of the baby (fetus) according to the number of weeks of pregnancy, and changes that are likely to occur in the pregnant woman's body and mind. In addition, the users can ask their concerns with other users if they wish. All participants used the app to respond to questionnaires. A gift code for JPY 500 (USD 4.3) was sent for each completed survey.

Interventions

The first author (DN) developed a smartphone-based six-module iCBT program for pregnant women together with four of the co-authors (EO, KI, NS, and YS). The copyright for the program belongs to the first author, and MTI Ltd. wrote the programming code to implement the program in the app. Topics of interest to pregnant women were extracted and used to tailor the program to them. The details of the program's development have been reported elsewhere.¹⁹ Although there are only six sessions, our previous RCT demonstrated that an iCBT program with the same number of sessions was effective in preventing depression in workers.²¹ The program was delivered to the participants *via* the Luna Luna Baby app; therefore, they did not have to download a separate app. Some parts of the modules were based on our previous iCBT program, which successfully prevented the onset of MDE in office workers.²¹ Other parts, including self-compassion, mindfulness, and values-based behavioral activation, were newly developed. The six modules—(1) psychoeducation, (2) case formulation based on a cognitive-behavioral model, (3) behavioral activation, (4) self-compassion, (5) mindfulness, and (6) problem-solving—were presented at a rate of one per week, with each module taking about 5 min to complete.

Outcomes

Primary outcome

The primary outcomes were onset of MDE by 32 weeks of gestation and by 3 months postpartum. Onset of MDE during the follow-up period was assessed using the web-based self-administered Japanese version of the WHO-CIDI 3.0 depression section, in accordance with the DSM-IV-TR criteria. The web version has been reported to agree well with the clinical diagnosis of MDE.²² An MDE incident was considered to have occurred if the participant reported an MDE episode at either 32 weeks of gestation or at 3 months postpartum after the baseline. The participant was requested to state the month during which onset of the MDE episode occurred.

Secondary outcomes

The Edinburgh Postnatal Depression Scale

Depressive symptoms were assessed using the Edinburgh Postnatal Depression Scale (EPDS), Japanese version.^{23,24} The EPDS is the most frequently used scale for screening perinatal depression because of its focus on the cognitive symptoms of depression rather than somatic items that can generate false positives during and after pregnancy. It comprises 10 items each scored on a 4-point scale (0–3), for a total possible score of 30. Higher scores indicate more severe depressive symptoms. The EPDS was conducted at baseline, 32 weeks of gestation, 1 week postpartum, and 3 months postpartum.

Kessler's Psychological Distress Scale

Psychological distress was assessed using Kessler's Psychological Distress Scale (K6), Japanese version.^{25,26} K6 comprises six items that assess how frequently respondents have experienced symptoms of psychological distress during the past 30 days. Possible responses range from 0 (none of the time) to 4 (all of the time), for a maximum total score of 24. Higher scores indicate more severe psychological distress. K6 was administered at baseline, 32 weeks of gestation, 1 week postpartum, and 3 months postpartum.

Participants who had not responded within a week received a popup message requesting them to complete the assessment.

Sample size calculation

The sample size necessary to evaluate the primary outcome was calculated as follows. New onset of MDE in Japan during the observation period and the effect size of the hazard ratio (HR) were estimated to be 5% and 0.65, respectively, based on a meta-analysis.²⁷ In addition, 30% of the participants were expected to drop out prior to follow-up, based on a previous study.²¹ Assuming an α level of 0.05 (two-tailed) and a β level of 0.20, power Cox analysis was performed with STATA 14.0, giving a calculated sample size of 4812. Because many participants using the app were recruited in 1 day, it was difficult to stop recruiting immediately when the sample size was reached. Thus, the sample size was set at 5000.

Randomization

Participants who met the inclusion criteria were assigned at random to the intervention and control groups and were then grouped according to their K6 score in the baseline survey into a lower stratum (four or less) or a higher stratum (five or more).^{19,27} MTI Ltd. sent the baseline data to our research team. We analyzed not only the data of the whole sample (to examine the universal intervention effect) but also the data of prespecified subgroups (to examine the selective intervention effect). An independent biostatistician used a computer-generated random allocation sequence to create a stratified permuted-block random table and the block size of the RCT was set to 4. The stratified permuted-block random table was password-protected and blinded to the research team. Only the research assistant was able to access it during the random allocation. MTI Ltd. then allocated the study participants to the intervention group in accordance with the allocation results provided by the research assistant.

Statistical methods

Main analysis

As reported elsewhere,^{28,29} we conducted a survival analysis to determine the effectiveness of the intervention on time to MDE onset while controlling for censoring effects resulting from differing follow-up durations or the completion of follow-up without MDE onset. The length of follow-up for each participant was represented by either the number of months between the baseline survey and onset of MDE or the completion of the follow-up period (3 months postpartum, or 32 weeks of gestation if the participant dropped out at the 3-month postpartum follow-up), whichever came first. We used the Kaplan–Meier method to estimate the cumulative incidences of MDE at 32 weeks of gestation and at 3 months postpartum, as well as event-free survivals at every follow-up month. The difference in cumulative proportions of MDE at 32 weeks of gestation and at 3 months postpartum between the intervention and control groups was analyzed. The difference in survival probabilities between the two groups was analyzed by the log-rank. The difference in incidence of MDEs between the groups and the estimated HR with 95% confidence intervals (Cis) was analyzed by a single covariate Cox discrete-time hazard model. We also estimated the intervention effect, adjusting for dependent censoring and using the inverse probability of the censoring weighted (IPCW) method to conduct a sensitivity analysis.²⁸ The analysis was carried out on an intention-to-treat (ITT) basis.

Secondary analyses

Mixed models for repeated measures analyses were conducted for the secondary outcomes (EPDS and K6 scores), using a group (intervention or control) \times time (baseline, 32 weeks of gestation, 1 week postpartum, or 3 months postpartum) interaction as an indicator of the intervention effect while accounting for missing data within the statistical model. Statistical significance for all analyses in this study was set at the 0.05 level (two-tailed), and the 95% Cis were calculated. Effect size was estimated as follows: first the MIXED procedure was used to convert to an effect size by dividing by a pooled SD at baseline and follow-ups in order to estimate the regression coefficient for each interaction of group (intervention group vs. control group) \times time (baseline,

32 weeks of gestation, 1 week postpartum, and 3 months postpartum); second, Cohen's *d* was calculated for completers at baseline for each follow-up. This analysis was also carried out on an ITT basis.

Subgroup analysis

We speculated that the effectiveness of the programs might differ depending on the initial severity of psychological distress. Therefore, we analyzed the data according to the prespecified subgroups (i.e. participants who scored four or less vs. five or more on the K6 during the baseline survey). Moreover, post-hoc analysis of the subgroup with K6 scores ranging from 5 to 8 was performed because iCBT might be particularly effective for people with subthreshold depressive symptoms.²⁶ The number needed to treat (NNT) to prevent one case of new MDE onset was calculated when the results were statistically significant.

All statistical analyses were performed using SPSS Statistics 26.0 (IBM Corp., Armonk, NY) and R 4.0.0 (R Foundation for Statistical Computing, Vienna, Austria); the “survival” and “mets” packages were used for the IPCW method.

Ethical and safety considerations

Informed consent to collect data from the app was obtained from all participants after informing them about the study's purpose. They were told that participation was voluntary and could be withdrawn for any reason at any time, and that withdrawal would not have any negative repercussions. We expected no adverse health effects from this intervention; on the contrary, we expected that depressive symptoms would possibly improve. In addition, we sent messages to participants who met the criteria for MDE in the past month or for lifetime bipolar disorders at baseline, encouraging them to visit a psychiatrist.

Changes to the protocol

We planned to analyze the data by prespecified subgroups (i.e., participants who scored four or less vs. five or more on K6 at the baseline survey) in the registered protocol.¹⁹ In addition, we performed a post-hoc subgroup analysis to compare three groups based on their K6 score at baseline (4 or less, 5–8, and 9 or more), because the two latest individual participant data meta-analyses have shown that psychotherapy including iCBT is more effective with more severe baseline

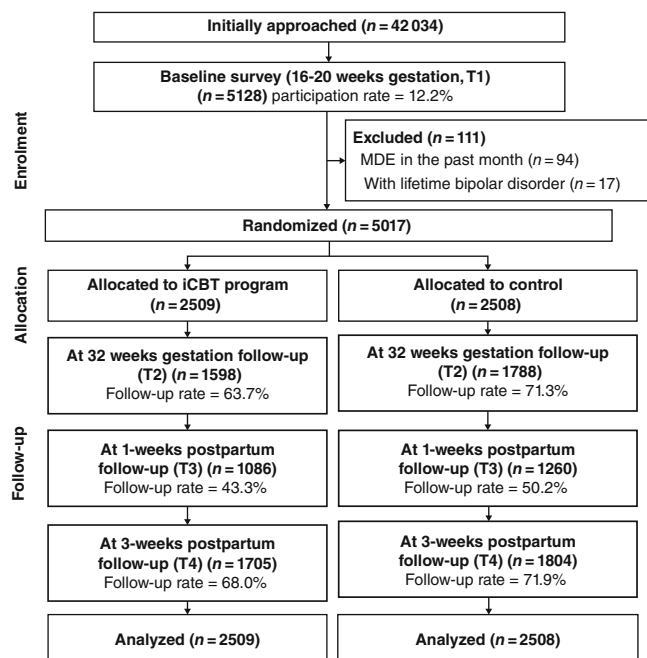


Fig. 1 Flow diagram of the study participants.

Table 1. Participant characteristics at baseline (*n* = 5017)

	Total (<i>n</i> = 5017)		iCBT program (<i>n</i> = 2509)		Control (<i>n</i> = 2508)	
	<i>N</i> (%)	Mean (SD)	<i>N</i> (%)	Mean (SD)	<i>N</i> (%)	Mean (SD)
Age (mean)		30.44 (4.6)		30.37 (4.6)		30.50 (4.6)
Pregnancy week (mean)		16.89 (1.3)		16.89 (1.3)		16.89 (1.3)
Educational status						
Junior high school	148	(2.9)	76	(3.0)	72	(2.9)
High school	1792	(35.7)	908	(36.2)	884	(35.2)
College	588	(11.7)	297	(11.8)	291	(11.6)
University	2283	(45.5)	1132	(45.1)	1151	(45.9)
Graduate school	206	(4.1)	96	(3.8)	110	(4.4)
Employment status						
Unemployed	1118	(22.3)	558	(22.2)	560	(22.3)
Student	31	(0.6)	15	(0.6)	16	(0.6)
Maternity leave	314	(6.3)	166	(6.6)	148	(5.9)
Part-time	795	(15.8)	404	(16.1)	391	(15.6)
Full-time	2759	(55.0)	1366	(54.4)	1393	(55.5)
Partner						
With partner	4977	(99.2)	2489	(99.2)	2488	(99.2)
Without partner	40	(0.8)	20	(0.8)	20	(0.8)
Planned pregnancy						
Planned	3581	(71.4)	1795	(71.5)	1786	(71.2)
Unplanned	1436	(28.6)	714	(28.5)	722	(28.8)
Number of children						
Primipara	3336	(66.5)	1682	(67.0)	1654	(65.9)
Multipara	1681	(33.5)	827	(33.0)	854	(34.1)
Past history of depression						
No	4312	(85.9)	2141	(85.3)	2171	(86.6)
Yes	705	(14.1)	368	(14.7)	337	(13.4)

depressive symptoms,^{17,30} and because a previous study has shown that stratum-specific likelihood ratios differ significantly between baseline K6 scores below 8 and above 9.²⁶ We considered the possibility that treatment effects could vary significantly between scores below 8 and above 9. The multiple imputation method for the main analysis was not adopted because the Cox discrete-time hazard model and the IPCW method took the missing due to censoring or dropout into account in the modeling.

Results

Participant flowchart

The flow of participants through the study is shown in Fig. 1. Out of 42 034 pregnant women who were sent invitation messages from November 2019 to March 2020, a total of 5128 (12.2%) pregnant women agreed to participate in the study. Of those, 111 were excluded:

94 met the diagnostic criteria of MDE in the past month, and 17 met the diagnostic criteria of lifetime bipolar disorder. The remaining 5017 participants were randomly allocated to the intervention (*n* = 2509) or control group (*n* = 2508).

Of the 5017 participants, 52 registered more than once and participated in the study multiple times from different smartphones or computers. Overall, 55 duplicate IDs were generated (107 research IDs were generated from 52 people). According to the allocated group, we analyzed all participants, including duplicate registration cases, based on the ITT principle. In other words, a participant who was allocated to the same group after registering for the second time was counted as two participants in the same group. After registering for the second time, a participant who was allocated to another group was counted as two participants, one in each group. Subsequently, two types of sensitivity analyses were conducted. First, the participants allocated to the same group for the second time were treated as

Table 2. Person-months observation, cases, and incident rate, and hazard ratio between groups on major depressive episode (MDE) onset

Group	<i>N</i>	Person-months		Incident rate (per months)	Survival time Mean (SE)	Hazard ratio	95% CI low	95% CI high	<i>P</i> -value
		Observed	Case (<i>N</i>)						
iCBT program	2509	15 712	59	0.376	13.70 (0.39)	0.85	0.61	1.20	0.362
Control	2508	16 718	73	0.437	13.65 (0.41)	1.00	–	–	–
Total	5017	32 430	132	0.407	13.67 (0.28)				

Note. Log rank test $\chi^2(1) = 0.830, P = 0.362$. CI, confidence interval.

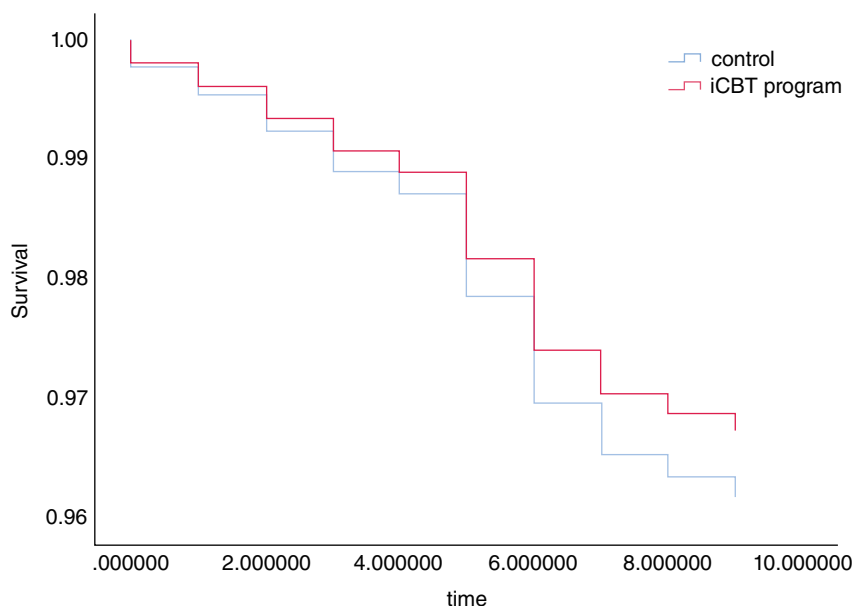


Fig. 2 Kaplan-Meier curve of the iCBT program and control group.

Table 3. Subgroup analysis: hazard ratio between groups on major depressive episode (MDE) onset stratified by K6 score (prespecified)

	N	Person-weeks observed	Case (N)	Incident rate (per week)	Survival time Mean (SE)	Hazard ratio	95% CI low	95% CI high
K6 low (≤ 4)								
iCBT program	1159	31 172	15	0.048	56.28 (0.18)	1.58	0.71	3.52
Control	1160	33 162	10	0.030	60.50 (0.16)	1.00	–	–
Total	2319	64 334	25	0.039	60.36 (0.13)			
K6 high (≥ 5)								
iCBT program	1350	34 457	95	0.276	56.66 (0.43)	1.04	0.78	1.37
Control	1348	37 111	99	0.267	57.58 (0.45)	1.00	–	–
Total	2698	71 568	194	0.271	57.58 (0.31)			

Table 4. Subgroup analysis: hazard ratio between groups on major depressive episode (MDE) onset stratified by K6 score (post hoc)

	N	Person-months Observed	Case (N)	Incident rate (per months)	Survival time Mean (SE)	Hazard ratio	95% CI low	95% CI high	P-value
K6 ≤ 4									
iCBT program	1159	7365	10	0.136	12.91 (0.03)	1.76	0.64	4.85	0.273
Control	1160	7185	6	0.084	13.94 (0.03)	1.00	–	–	
Subtotal	2319	14 550	16	0.110	13.92 (0.02)				
5 \leq K6 \leq 8									
iCBT program	732	4560	10	0.219	13.82 (0.06)	0.38	0.19	0.79	0.009
Control	735	4929	28	0.568	13.54 (0.09)	1.00	–	–	
Subtotal	1467	9489	38	0.400	13.68 (0.05)				
9 \leq K6									
iCBT program	618	3787	39	1.030	13.17 (0.13)	1.04	0.67	1.63	0.853
Control	613	3974	39	0.981	12.31 (0.11)	1.00	–	–	
Subtotal	1231	7761	78	1.005	13.20 (0.09)				
Total	5017	32 430	132	0.407	13.67 (0.28)				

Values in bold indicates statistically significant difference.

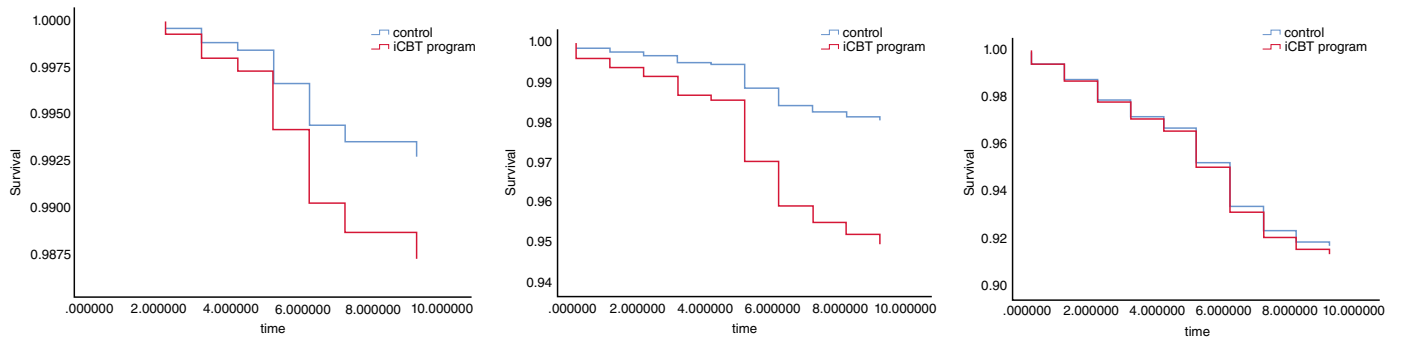


Fig. 3 Kaplan–Meier curve of the iCBT program and control group stratified by K6 at baseline.

one person in that group. Second, all participants who registered twice were excluded from the primary analysis.

At 32 weeks-of-gestation follow-up, 1598 (63.7%) participants in the intervention group and 1787 (71.3%) in the control group completed the follow-up survey. At the 3-month postpartum follow-up, 1705 (68.0%) participants in the intervention group and 1804 (71.9%) in the control group completed the follow-up survey. Attrition rates were significantly higher for the intervention group both at 32 weeks of gestation ($P < 0.01$) and at 3 months ($P < 0.01$).

Baseline characteristics

The demographic characteristics of the participants were similar for both groups (Table 1). The average age (standard deviation) was 30.37 (4.6) and 30.50 (4.6) years for the intervention and control groups, respectively. Approximately half of the participants were university or higher education graduates (48.9% and 50.3%, respectively) and employed as full-time workers (54.4% and 55.5%, respectively). Most participants were primipara (67.0% and 65.9%, respectively) and had planned pregnancies (71.5% and 71.2%,

Table 5. Change of Edinburgh Postnatal Depression Scale (EPDS) score between the groups

EPDS	Estimated mean (SE)				P for interaction	Cohen's d		
	Baseline	32 weeks	Post 1-week	Post 3-month		32 weeks (95% CI)	Post 1-week (95% CI)	Post 3-month (95% CI)
Total (n = 5017)					0.608	0.01 (−0.04 to 0.07)	0.05 (−0.01 to 0.10)	0.03 (−0.02 to 0.09)
iCBT program (n = 2509)	5.26 (0.09)	5.55 (0.11)	6.47 (0.12)	5.21 (0.12)				
Control (n = 2508)	5.07 (0.09)	5.41 (0.11)	6.50 (0.12)	5.16 (0.12)				
K6 ≤ 4 (n = 2319)					0.400	−0.09 (−0.17 to −0.01)	0.04 (−0.04 to 0.12)	−0.05 (−0.13 to 0.03)
iCBT program (n = 1159)	3.16 (0.08)	3.69 (0.13)	5.15 (0.15)	3.53 (0.13)				
Control (n = 1160)	3.11 (0.08)	3.39 (0.13)	5.22 (0.14)	3.33 (0.13)				
5 ≤ K6 ≤ 8 (n = 1467)					0.619	−0.05 (−0.15 to 0.06)	−0.05 (−0.15 to 0.05)	0.07 (−0.03 to 0.17)
iCBT program (n = 732)	5.66 (0.14)	6.13 (0.21)	7.03 (0.24)	5.40 (0.21)				
Control (n = 735)	5.45 (0.14)	5.75 (0.19)	6.64 (0.23)	5.44 (0.20)				
9 ≤ K6 (n = 1231)					0.118	0.13 (0.02 to 0.25)	0.17 (0.05 to 0.28)	0.10 (−0.02 to 0.21)
iCBT program (n = 618)	8.73 (0.22)	8.41 (0.31)	8.04 (0.31)	8.05 (0.30)				
Control (n = 613)	8.35 (0.22)	8.77 (0.28)	8.57 (0.29)	8.20 (0.30)				

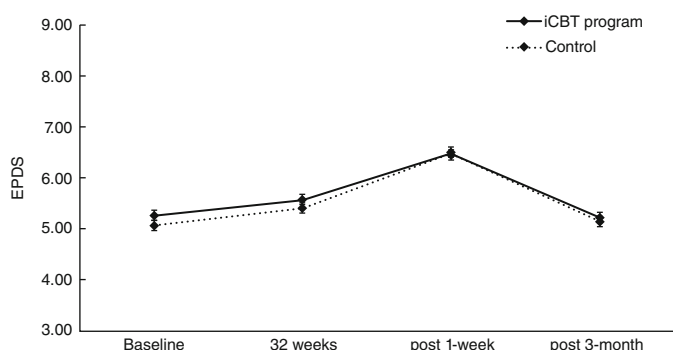


Fig. 4 Change in Edinburgh Postnatal Depression Scale (EPDS) score between groups.

respectively). Most participants had a partner (99.2% and 99.2%, respectively).

Effects of iCBT programs on preventing MDE

A total of 59 participants (2.35%) in the intervention group and 73 (2.91%) in the control group reported new onset of MDE during the follow-ups. Using the Kaplan–Meier method, the estimated event-free survival time was 13.70 (standard error [SE] = 0.39) weeks in the intervention group and 13.65 (SE = 0.41) weeks in the control group (Table 2 and Fig. 2). Compared with the control group, the HR of MDE in the intervention group was 0.85 (95% CI 0.61–1.20), and the adjusted HR using the IPCW method was 0.84 (95% CI 0.60–1.19) which was not significantly different. The results of sensitivity analyses corresponding to duplicate registration are shown in Tables S1–S3 and Figures S1, S2. The results were similar to those in the primary analyses.

Subgroup analysis

Subgroup analyses were conducted for both the prespecified subgroups (i.e., participants who scored four or less vs. five or more on K6 at baseline) and the post-hoc subgroup (i.e., participants who scored four or less, between five and eight, or nine or more on K6 at baseline). Of the participants who scored four or less at baseline, 10 (0.86%) in the intervention group and six (0.52%) in the control group reported new onset of MDE. Similarly, among the participants who scored five or more, 49 (3.63%) in the intervention group and 67 (4.97%) in the control group reported a new onset of MDE (Table 3). Among participants who scored between five and eight, 10 (1.37%) in the intervention group and 28 (3.81%) in the control group reported new onset of MDE. Among participants who scored nine or more, 39 (6.31%) in the intervention group and 39 (6.36%) in the control group reported new onset of MDE.

Compared with the control group, the HRs of MDE for the intervention group were 1.76 (95% CI 0.64–4.85) for participants who

scored four or less on K6 at baseline, 0.77 (95%CI 0.54–1.12) for those who scored five or more, 0.38 (95%CI 0.19–0.79) for those who scored between five and eight, and 1.04 (95%CI 0.67–1.63) for those who scored nine or more (Table 4, Fig. 3). The NNT to prevent one case of new MDE onset among participants who scored between five and eight on K6 at baseline was 41 (95% CI 25–121). In addition, we conducted post-hoc subgroup analyses of EPDS outcomes. The results showed that there was no significant effect on those with K6 scores between five and eight, while a small but statistically significant effect was observed on those with K6 scores of nine or higher (Table 5).

Effects of the iCBT program on EPDS and K6

The average EPDS and K6 at baseline in the intervention group were nearly the same as those in the control group. Based on the mixed-model analysis, the estimated effects of the iCBT program on EPDS and K6 were not significant. The effect sizes (Cohen’s d) on EPDS at 32 weeks of gestation, 1 week postpartum, and 3 months postpartum were 0.01, 0.05, and 0.03, respectively, and those on K6 were – 0.04, 0.02, and 0.01, respectively. The results of EPDS are shown in Table 5 and Figure 4, and the results of K6 are shown in the Tables S1–S3 and Figures S1, S2.

Process evaluation

Out of 2509 participants in the intervention group, 1995 (79.5%) completed module 1, 1800 (71.7%) completed module 2, 1734 (69.1%) completed module 3, 1,48 (65.7%) completed module 4, 1544 (61.5%) completed module 5, and 1402 (55.9%) completed module 6.

A total of 934 (37.2%) participants in the intervention group completed all six modules. There were no statistical differences between the two groups and among three groups (Table 6).

Discussion

This study is one of the largest RCTs globally aimed at preventing perinatal depression.¹³ The developed six-session iCBT program did not significantly prevent new onset of MDE or depressive symptoms, as assessed by the EPDS and K6. This was also true for subgroups scoring four or less, five or more, and nine or more on K6 at baseline. A limited number of therapy sessions might be insufficient to have an effect on prevention. However, a significant prevention effect was observed for subgroups that scored between five and eight on K6 at baseline. This study’s findings suggest that iCBT might prevent perinatal depression only among pregnant women with subthreshold depressive symptoms.

The reason for the lack of a significant effect might possibly be due to overestimating the effect of the iCBT program. The sample size in this study was calculated based on the previously estimated HR effect size of 0.65.²⁷ However, the HR in this study was 0.85. A meta-analysis reported that iCBT produces effects equivalent to face-to-face CBT.³¹ In contrast, the effect of unguided CBT has been reported to be a little less than that of guided CBT.³⁰ Accordingly, it is possible that the effect size should have been estimated more modestly. In the future, an AI-guided CBT program with additional

Table 6. Process evaluation

Completers	iCBT program (n = 2509)	K6 ≤ 4 (n = 1159)	5 ≤ K6 ≤ 8 (n = 732)	9 ≤ K6 (n = 618)	P-value
	N (%)	N (%)	N (%)	N (%)	
Module 1	1995 (79.5)	921 (79.5)	584 (79.8)	490 (79.3)	0.974
Module 2	1800 (71.7)	851 (73.4)	514 (70.2)	435 (70.4)	0.221
Module 3	1734 (69.1)	826 (71.3)	486 (66.4)	422 (68.3)	0.072
Module 4	1648 (65.7)	762 (65.7)	478 (65.3)	408 (66.0)	0.961
Module 5	1544 (61.5)	729 (62.9)	437 (59.7)	378 (61.2)	0.370
Module 6	1402 (55.9)	668 (57.6)	389 (53.1)	345 (55.8)	0.159
All modules	934 (37.2)	446 (38.5)	251 (34.3)	237 (38.3)	0.148

homework may be more effective. Since the intensity was low, this program might be better described as a “cognitive behavioral based program for self-care” rather than cognitive behavioral therapy.

Alternatively, it might be better to conservatively estimate the effects of iCBT on perinatal depression. Pregnant and postpartum women continue to experience significant changes, both biologically and in terms of their environment, with the progression of pregnancy and childbirth and child rearing, which is a major difference from other populations. Mental health is greatly affected by whether or not the pregnancy progression and child rearing are going well, which might exceed the effects of iCBT. Moreover, the four topics we chose were the main concerns of pregnant women, but the main concerns are not necessarily related to depression. It might have been better to interview patients with perinatal depression to determine the topics. Furthermore, compared with our previous smartphone based RCT,³² the completion rate in this study was relatively low. This might be explained by the fact that many pregnant women participated in the study in a casual manner and did not have a very high commitment to completion. Future research should be conducted in populations with higher motivation to engage in iCBT. It would be good to have instructions to encourage more motivated people to participate in the program.

Another possible reason for the program's non-significant effect was the relatively low incidence of MDE. The sample size was calculated based on a previous study showing that a prevalence of antenatal and postpartum depression in Japan at university hospitals of approximately 5.0%.³³ However, the incidence of MDE in the present study was lower than: 2.35% in the intervention group and 2.91% in the control group. Most pregnant women in Japan visit regional maternity hospitals instead of university hospitals, and the prevalence of antenatal depression was reported to be 1.1% at a regional obstetrics and gynecology hospital.³⁴ Thus, it might have been better to presume a lower incidence of MDE.

The HR for participants who scored between five and eight at baseline was 0.38, which was superior to the results of a meta-analysis showing that the incident rate ratio of perinatal depression in face-to-face CBT was 0.65.²⁷ Although this result needs to be replicated because only a post-hoc analysis was conducted with participants having subthreshold depressive symptoms at baseline, it should be noted that the difference between subgroups in HRs was relatively large. Unguided iCBT might be more effective for pregnant women with subthreshold symptoms than for those with severe symptoms, as reported in a recent systematic review.³⁰ In addition, the results for post-hoc subgroup analysis of EPDS showed that there was no significant effect on those with K6 scores between five and eight, while a small but statistically significant effect was observed on those with K6 scores of nine or higher. For those with baseline K6 scores of five to eight, survival time analysis results showed a more pronounced effect of prevention in the intervention group at 5–6 months after baseline. Thus, for depressive symptoms, the effect might be more seen at 3 months postpartum than at 32 weeks' gestation or 1 week postpartum. As for the group with a K6 score of nine or higher, it may be possible to consider that the intervention was effective in reducing subthreshold depressive symptoms in those with relatively less severe distress among them, although it did not prevent the onset of major depressive episode in those with the most severe distress among them.

The present study did not show significant effects of the iCBT program on improving depressive symptoms, as measured by EPDS and K6. These findings are consistent with our previous study, which showed that iCBT might be more effective in preventing the onset of depression than improving depressive symptoms.²¹ A possible reason for the discrepancy between group differences in the effects of preventing the onset of depression and improving depressive symptoms is that the MDE prevention effect of iCBT might be related to cumulative symptom levels during the follow-up period.²¹ This means that some people with high levels of baseline depressive symptoms (>9 points on the K6) might develop depression because of accumulated psychological distress, even if there is a small improvement in depressive symptoms. On the other hand, people with moderate levels of baseline depressive symptoms (5–8 points on the K6) might be able to prevent the onset of depression if

their depressive symptoms do not worsen, even if they do not show improvement in their depressive symptoms.

The strength of this study was the use of the WHO-CIDI to measure MDE, which allowed us to assess depressive episodes during the entire follow-up period. Fully automated CBT is cost-effective because it does not incur ongoing labor costs. The iCBT program can be disseminated to large numbers of pregnant women because Luna Luna Baby is the most widely used app of its kind in Japan.

This study has some limitations that should be noted. First, all outcomes were measured by self-report, which might have been affected by the participants' perceptions. Confirmation of whether the participant was pregnant was also self-reported, which would be a major limitation of the study. Moreover, gestational week calculation was also based on self-report, which is not optimal. However, many of the users of the app had moved from the ovulation date prediction app that they used before pregnancy to this app, which provides information for pregnant women. It is unlikely that most users would have continued to enter incorrect information from the pre-conception period. Second, the sample size might have been insufficient if the assumptions in the sample size calculation were not satisfied. Third, the follow-up rate was modest, although we used IPCW and a mixed model to account for attrition. Fourth, we did not exclude mental disorders other than depression and bipolar disorders, which might have affected the results. Fifth, there was no information about potential sources of bias such as previous experience with CBT and mindfulness practice, although the internal validity is considered to be high because this study was a RCT. Sixth, although this study is an RCT, the intervention effect might have been underestimated due to the app's functions of providing information about babies and pregnant women and the opportunity to talk about their problems with other pregnant women. Seventh, we did not have an external data monitoring committee because our intervention was not invasive, but in terms of maintaining the integrity of the trial, not having an external data monitoring committee could be considered inappropriate. Finally, users of the app cannot be considered representative of all pregnant women, although it is thought that approximately 1 in 4 pregnant women in Japan use the app. As mentioned above, many of the users of the app had used the ovulation date prediction app before they used the app. Accordingly, the findings of this study are generalizable to all pregnant women in Japan.

Acknowledgment

We would like to thank all the researchers and staff who contributed to this study, as well as all the study participants.

Funding statement

This work was supported by a Grant-in-Aid for Scientific Research (A) from The Japan Society for the Promotion of Science (19H01073 to DN). The funder had no role in the study design, data collection and analysis, decision to publish, or manuscript preparation.

Disclosure statement

MTI Ltd. was involved in this study, as mentioned in the manuscript. Specifically, MTI Ltd. provided the author with an opportunity to recruit research participants, but no financial support was provided. The iCBT program was developed by the authors and the copyright belongs to the first author. MTI Ltd. wrote the programming code to implement the program in the app. DN reports personal fees from Startia, Inc., empower, Inc., MD.net, AIG General Insurance Company, Ltd., outside the submitted work. NK reports grants from Infocom Corp, Fujitsu Ltd., Fujitsu Software Technologies, and TAK Ltd., as well as personal fees from Occupational Health Foundation, Japan Dental Association, Sekisui Chemicals, Junpukai Health Care Center, Osaka Chamber of Commerce and Industry, outside the submitted work.

Author contributions

D.N. conceived and designed the study. D.N., K.I., E.O., N.S., and Y.S. contributed to creating programs. N.Y. and N.K. contributed to

the development of study design. K.W. calculated sample size. K.W. and Y.M. developed an analysis plan. D.N. wrote the first draft of the manuscript, and all other authors revised the manuscript critically. All authors approved the final version of the manuscript.

Ethical standards

The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

References

1. Wisner KL, Miller ES, Tandon D. Attention to prevention-can we stop perinatal depression before it starts? *JAMA Psychiat*. 2019; **76**: 355–356.
2. Bennett HA, Einarson A, Taddio A, Koren G, Einarson TR. Prevalence of depression during pregnancy: Systematic review. *Obstet. Gynecol*. 2004; **103**: 698–709.
3. Hahn-Holbrook J, Cornwell-Hinrichs T, Anaya I. Economic and health predictors of National Postpartum Depression Prevalence: A systematic review, meta-analysis, and meta-regression of 291 studies from 56 countries. *Front. Psych*. 2017; **8**: 248.
4. Chaffin M, Kelleher K, Hollenberg J. Onset of physical abuse and neglect: Psychiatric, substance abuse, and social risk factors from prospective community data. *Child Abuse Negl*. 1996; **20**: 191–203.
5. Takehara K, Suto M, Kakee N, Tachibana Y, Mori R. Prenatal and early postnatal depression and child maltreatment among Japanese fathers. *Child Abuse Negl*. 2017; **70**: 231–239.
6. Davis EP, Glynn LM, Schetter CD, Hobel C, Chicz-Demet A, Sandman CA. Prenatal exposure to maternal depression and cortisol influences infant temperament. *J. Am. Acad. Child Adolesc. Psychiatry* 2007; **46**: 737–746.
7. Leech SL, Larkby CA, Day R, Day NL. Predictors and correlates of high levels of depression and anxiety symptoms among children at age 10. *J. Am. Acad. Child Adolesc. Psychiatry* 2006; **45**: 223–230.
8. Field T, Diego M, Hernandez-Reif M. Prenatal depression effects on the fetus and newborn: A review. *Infant Behav. Dev*. 2006; **29**: 445–455.
9. Diego MA, Field T, Hernandez-Reif M, Schanberg S, Kuhn C, Gonzalez-Quintero VH. Prenatal depression restricts fetal growth. *Early Hum. Dev*. 2009; **85**: 65–70.
10. Stein A, Pearson RM, Goodman SH *et al*. Effects of perinatal mental disorders on the fetus and child. *Lancet* 2014; **384**: 1800–1819.
11. Underwood L, Waldie KE, Peterson E *et al*. Paternal depression symptoms during pregnancy and after childbirth among participants in the growing up in New Zealand study. *JAMA Psychiat*. 2017; **74**: 360–369.
12. Force USPST, Curry SJ, Krist AH *et al*. Interventions to prevent perinatal depression: US preventive services task Force recommendation statement. *JAMA* 2019; **321**: 580–587.
13. Yasuma N, Narita Z, Sasaki N *et al*. Antenatal psychological intervention for universal prevention of antenatal and postnatal depression: A systematic review and meta-analysis. *J. Affect. Disord*. 2020; **273**: 231–239.
14. Fairburn CG, Patel V. The impact of digital technology on psychological treatments and their dissemination. *Behav. Res. Ther*. 2017; **88**: 19–25.
15. Barrera AZ, Wickham RE, Munoz RF. Online prevention of postpartum depression for Spanish- and English-speaking pregnant women: A pilot randomized controlled trial. *Internet Interv*. 2015; **2**: 257–265.
16. Haga SM, Drozd F, Lisoy C, Wentzel-Larsen T, Slinning K. Mamma Mia - a randomized controlled trial of an internet-based intervention for perinatal depression. *Psychol. Med*. 2019; **49**: 1850–1858.
17. 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.
18. Buntrock C, Ebert DD, Lehr D *et al*. Effect of a web-based guided self-help intervention for prevention of major depression in adults with subthreshold depression: A randomized clinical trial. *JAMA* 2016; **315**: 1854–1863.
19. Nishi D, Imamura K, Watanabe K *et al*. Internet-based cognitive-behavioural therapy for prevention of depression during pregnancy and in the post partum (iPDP): A protocol for a large-scale randomised controlled trial. *BMJ Open* 2020; **10**: e036482.
20. Campbell MK, Piaggio G, Elbourne DR, Altman DG, Group C. Consort 2010 statement: Extension to cluster randomised trials. *BMJ* 2012; **345**: e5661.
21. Imamura K, Kawakami N, Furukawa TA *et al*. Does internet-based cognitive behavioral therapy (iCBT) prevent major depressive episode for workers? A 12-month follow-up of a randomized controlled trial. *Psychol. Med*. 2015; **45**: 1907–1917.
22. Peters L, Clark D, Carroll F. Are computerized interviews equivalent to human interviewers? CIDI-auto versus CIDI in anxiety and depressive disorders. *Psychol. Med*. 1998; **28**: 893–901.
23. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression. Development of the 10-item Edinburgh postnatal depression scale. *Br. J. Psychiatry* 1987; **150**: 782–786.
24. Okano T, Murata M, Masuji F, Tamaki R, Nomura J, Miyaoko H. Validation and reliability of Japanese version of the EPDS. *Arch Psychiatr Diag Clin Evaluat*. 1996; **7**: 525–533.
25. Kessler RC, Andrews G, Colpe LJ *et al*. Short screening scales to monitor population prevalences and trends in non-specific psychological distress. *Psychol. Med*. 2002; **32**: 959–976.
26. Furukawa TA, Kawakami N, Saitoh M *et al*. The performance of the Japanese version of the K6 and K10 in the world mental health survey Japan. *Int. J. Methods Psychiatr. Res*. 2008; **17**: 152–158.
27. Cuijpers P, van Straten A, Smit F, Mihalopoulos C, Beekman A. Preventing the onset of depressive disorders: A meta-analytic review of psychological interventions. *Am. J. Psychiatry* 2008; **165**: 1272–1280.
28. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* 2000; **56**: 779–788.
29. Richards DA, Ekers D, McMillan D *et al*. Cost and outcome of Behavioural activation versus cognitive Behavioural therapy for depression (COBRA): A randomised, controlled, non-inferiority trial. *Lancet* 2016; **388**: 871–880.
30. Karyotaki E, Efthimiou O, Miguel C *et al*. Internet-based cognitive behavioral therapy for depression: A systematic review and individual patient data network meta-analysis. *JAMA Psychiat*. 2021; **78**: 361–371.
31. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlof E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: An updated systematic review and meta-analysis. *Cogn. Behav. Ther*. 2018; **47**: 1–18.
32. Imamura K, Tran TTT, Nguyen HT *et al*. Effect of smartphone-based stress management programs on depression and anxiety of hospital nurses in Vietnam: A three-arm randomized controlled trial. *Sci. Rep*. 2021; **11**: 11353.
33. Kitamura T, Yoshida K, Okano T *et al*. Multicentre prospective study of perinatal depression in Japan: Incidence and correlates of antenatal and postnatal depression. *Arch. Womens Ment. Health* 2006; **9**: 121–130.
34. Usuda K, Nishi D, Makino M *et al*. Prevalence and related factors of common mental disorders during pregnancy in Japan: A cross-sectional study. *Biopsychosoc Med*. 2016; **10**: 17.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Figure S1 Kaplan–Meier curve of the iCBT program and control group corresponding to duplicate registration

Figure S2 Change of K6 score between the groups

Table S1 Descriptive characteristics after deleting duplication

Table S2. sensitivity analysis: hazard ratio between groups on MDE onset after deleting duplication

Table S3. Change of secondary outcomes score between the groups

Table S4. Participant characteristics at baseline by post-hoc subgroup