Free choice of treatment content, support on demand and supervision in internet-delivered CBT for adults with depression: A randomized factorial design trial

Gerhard Andersson [a,b,c,*], Anton Käll [a,b], Simon Juhlin [a], Carl Wahlström [a],
Edvard de Fine Licht [a], Simon Färdeman [a], Anna Franck [a], Anna Tholcke [a], Karin Nachtweij [d],
Emma Fransson [d], Kristofer Vernmark [a], Mikael Ludvigsson [b,e,f], Matilda Berg [a]

[a] *Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden*
[b] *Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden*
[c] *Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden*
[d] *Department of Psychology, Faculty of Health and Life Sciences, Linnaeus University, Växjö, Sweden*
[e] *Department of Psychiatry, and Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden*
[f] *Department of Acute Internal Medicine and Geriatrics, and Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden*

ARTICLE INFO

ABSTRACT

Even if much is known regarding the effects of internet-delivered cognitive behaviour therapy (ICBT) for depression there are several topics that have not been studied. In this factorial design trial with 197 participants we investigated if clients in ICBT could select treatment modules themselves based on a selection of 15 tailored treatment modules developed for use in ICBT for depression. We contrasted this against clinician-tailored module selection. We also investigated if support on demand (initiated by the client) could work as well as scheduled support. Finally, we tested if clients that were mentioned in supervision would improve more than clients not mentioned (with the exception of acute cases). The treatment period lasted for 10 weeks, and we measured effects at post-treatment and two-year follow-up. Measures of depression and secondary outcomes were collected at pre-treatment, post-treatment and two-year follow-up. Overall, within-group effects were large across conditions (e.g., $d = 1.73$ on the BDI-II). We also found a small but significant difference in favour of self-tailored treatment over clinician-tailored ($d = 0.26$). Within-group effects for the secondary measures were all moderate to large including a test of knowledge about CBT. The other two contrasts "support on demand" and "supervision" yielded mostly non-significant differences, with the exception of a larger dropout rate in the support on demand condition. There were few negative effects (2.2%). Effects were largely maintained at a two-year follow-up. We conclude that clients can choose treatment modules and that support on demand may work. The role of supervision is not yet clear as advice can be transferred across clients.

1. Internet-delivered CBT for depression

Internet-delivered cognitive behaviour therapy (ICBT) has been around for more than 20 years and has resulted in a large number of controlled trials (Andersson, 2018). There are also indications that this treatment format has resulted in larger and better quality trials (Schuster et al., 2021), compared to regular psychotherapy studies which tend to be more expensive and time-consuming. Major depression and symptoms of depression are extremely common and even if there are several

evidence-based treatments (Malhi & Mann, 2018), there is still a treatment versus demand gap, as many people prefer psychological treatments over medication. ICBT is one way to reduce this gap and there are now several ICBT depression programs that have been tested in controlled efficacy trials, effectiveness trials, comparisons against face-to-face treatments and long-term follow-up studies (Andersson & Berger, 2021). Moreover, there are trials on adolescents, adults and older persons (Andersson & Berger, 2021).

It is well known that depression is associated with comorbid

problems and disorders such as anxiety disorders, insomnia, stress problems, just to name a few. One way to handle this comorbidity is to develop treatments that cover several problems by addressing underlying mechanisms. But it is also possible to tailor treatments based on client characteristics and preferences. In a number of controlled trials we have developed and tested this concept (Păsărelu, Andersson, Bergman Nordgren, & Dobrean, 2017), and findings suggest that tailoring treatment components can be effective in the treatment of mild to moderate depressive symptoms (including major depression) (Johansson et al., 2012). One example of how tailoring may help is when a client has mixed problems (which is very common in depression) and there needs to be a selection of treatment components as it is not feasible to deliver them all. It can also mean that a problem like insomnia is more feasible to work with than an alternative treatment module on for example anxiety.

1.1. The role of choice

While preference to some extent is inherent when tailoring ICBT as clinicians interview the patient and then recommend a selection of treatment modules (for example on psychoeducation, behavioural activation, stress and insomnia), we do not know if clients are able to select treatment components on their own independently. We tested this early in an open trial on anxiety (Andersson, Estling, Jakobsson, Cuijpers, & Carlbring, 2011). Treatment preference studies usually involve full treatment packages. Overall, there seems to be a preference effect in medical studies with an effect size = 0.18 for the preferred treatment (Delevry & Le, 2019), but the extent to which this applies to psychotherapy studies is not certain. Our approach in the present study was to investigate preference within treatment packages – in this case ICBT for depressive symptoms using a total of 15 different treatment modules. There are few previous controlled trials on tailoring treatment components based on client's own choice.

1.2. Effects of support format and supervision

Early studies on ICBT for depression clearly showed that minimal scheduled therapist guidance resulted in better outcomes than self-guided treatments without any therapist involvement (Andersson & Cuijpers, 2009). More recently researchers have begun to test if support can be delivered on demand with clients being given the option to request feedback and ask questions if they need to. This can reduce clinician time, but may decrease effects and increase dropout rates if clients are not provided scheduled support. Moreover, automated reminders and messages can be used to reduce clinician time. There are studies on optional support for other conditions than depression (Dahlin, Johansson, Romare, Carlbring, & Andersson, 2022; Hadjistavropoulos et al., 2017; Oromendia, Orrego, Bonillo, & Molinuevo, 2016), and on depression (Kleiboer et al., 2015) showing mixed findings. In one large trial (N = 1089) support on request was found to be as effective as individualized feedback (Zagorscak, Heinrich, Sommer, Wagner, & Knaevelsrud, 2018), but resulted in significantly larger dropout rates (25.8% versus 17.3% when guidance was provided). In one preference trial including clients with depressive symptoms (N = 401), clients reported a clear preference for guidance (78% vs. 22%) (Hadjistavropoulos et al., 2019). Both groups did however improve which motivates controlled studies on the effects of optional support.

Supervision of CBT clinicians is a topic that has been insufficiently studied in controlled research (Alfonsson, Parling, Spännargård, Andersson, & Lundgren, 2018). There is to our knowledge almost no systematic research on the importance of supervision in ICBT even if the topic has been mentioned (Drozd et al., 2016) - mostly in association with training of clinicians (Thew et al., 2019) and dissemination (Titov et al., 2019). Overall, there is a need to study if supervision has any effects for the clients. In the present study we randomized clients to be reviewed or not during supervision with the aim too investigate if

supervision would yield better outcomes compared to the ones who were not mentioned during supervision (with the exception of risk management). We decided not to randomize therapists to receive supervision or not which would have been a stronger test of the role of supervision. Thus all therapists received supervision but were only instructed to talk about the clients who were in the supervision condition.

1.3. Factorial design trials as a possible solution?

As with psychotherapy research and medicine in general, most trials on ICBT have been RCTs with two or sometimes three groups. In addition, the standard group design has been the way to test components in the form of dismantling studies, even if few have been sufficiently powered for detecting even moderate effects (Watkins & Newbold, 2020). Given the larger sample sizes possible in ICBT studies a factorial design approach may be a more effective way to answer more than one question relating to "what works for whom" and also allow for testing interactions. There are now factorial design trials on ICBT suggesting that the approach works (Hadjistavropoulos et al., 2022). Another example is a trial on depression in which participants (N = 239) were randomized into one of eight intervention arms, with each component being present in half of the intervention arms (Kelders, Bohlmeijer, Pots, & van Gemert-Pijnen, 2015). One important finding from that study was a small difference in favour of human over automated support. In a small factorial design pilot trial on generalized anxiety disorder (N = 85) we found that self-tailored treatment was largely as effective as a worry-specific program, with the other contrast being scheduled vs. support on demand (Dahlin et al., 2022).

1.4. Purpose of the present study

In the present factorial design trial, we examined the three independent variables: self-tailored vs. clinician-tailored treatment modules delivered during 10 weeks; scheduled therapist support vs. support on demand; clients being mentioned in supervision vs. not mentioned in supervision. The main outcome was symptoms of depression measured at pre-treatment, post treatment and at a two-year follow-up. Given the fact that our independent variables may have differential effects on other measures than symptoms of depression we also measured secondary outcomes (e.g., anxiety, insomnia and quality of life), knowledge about depression and its treatment, and treatment satisfaction. In particular, we were interested in the role of knowledge and how well this was maintained at the two-year follow-up. We also measured treatment satisfaction as this outcome could reflect differences based on the independent variables.

2. Method

2.1. Trial design

The research protocol for a forthcoming updated trial was registered on ClinicalTrials.gov (registration number NCT04260750), and this was an initial test (originally intended as a pilot and done in 2017). The study was approved by the ethics committee at Linköping University in Sweden (2016/447-31). Informed consent was obtained through an online form that was mandatory in order to gain access to the screening. Since the study aimed to investigate the differences between treatment formats, support types and supervision, as well as possible interaction effects between the variables, a 2x2x2x3 factorial design was used (with three between group factors and one within-group factor with three measurement points: pre, post-treatment and two-year follow-up). An overview of the 8 conditions in the study is presented in Table 1.

Table 1
Description of the groups.

Group	Who chose the modules?	Scheduled guidance/Guidance on demand?	Supervision available for the participants in the group, yes/no
A	Tailored by the therapist and research team	Scheduled guidance	Yes
B	Tailored by the therapist and research team	Scheduled guidance	No
C	Tailored by the therapist and research team	Guidance on demand	Yes
D	Tailored by the therapist and research team	Guidance on demand	No
E	Self-tailored by the clients	Scheduled guidance	Yes
F	Self-tailored by the clients	Scheduled guidance	No
G	Self-tailored by the clients	Guidance on demand	Yes
H	Self-tailored by the clients	Guidance on demand	No

2.2. Recruitment, randomization and participants

Participants were recruited via social media, Google AdWords, postings at primary care centres and an article in a local newspaper. A site www.iterapi.se/sites/robin/was created on the treatment platform iterapi (Vlaescu, Alasjö, Miloff, Carlbring, & Andersson, 2016), with information about the study, the people behind and how to register.

Following an initial check of the results of the online screening some participants were directly excluded via a personalized email or via phone if they needed advice on where to seek help or an explanation why their problems were not suitable for the trial. Eligible participants were contacted for a diagnostic telephone interview using the M.I.N.I. version 7.0.1 (Sheehan et al., 1998). The interviews were conducted by six final year clinical psychology students under supervision.

A final decision regarding inclusion or exclusion was made at intake meetings. The principal investigator (GA) and the clinical psychology students involved in the study were present during these meetings with the possibility to contact the psychiatrist involved in the study (ML). We informed the excluded persons via e-mail or phone about the reasons for exclusion and if needed encouraged them to seek help in primary care. The screening and inclusion period lasted for three weeks. We phoned all participants for a post-treatment interview in which the clinical global impression (CGI) questions were asked (Busner & Targum, 2007) and feedback was received from the participants. They were also phoned at the two-year follow-up.

Inclusion criteria were: (a) being 18 years old or above, (b) screening positive for the diagnostic criteria of major depression or unspecified depressive disorder according to the DSM-5, (c) elevated scores on the PHQ-9 (at least 5 points) and the Beck Depression Inventory-II (10 points), (d) fluent in Swedish and being able to write and read Swedish text, (e) regular access to a computer/device and the internet, (f) no current substance or alcohol abuse, (g) no active suicidal ideation, (h) no ongoing psychological treatment, and (i) if using psychiatric medication, a stable dose (no dose adjustments during the previous six weeks or scheduled adjustments in the near future). Comorbidity was allowed with the exception of major medical or psychiatric problems that could interfere with the treatment. Other exclusion criteria were: (a) over 40 points on the BDI-II and (b) ongoing substance or drug abuse based on the M.I.N.I. interview.

After inclusion, the participants were randomly assigned to one of eight groups (see Table 1). An employee at Linköping University who was not involved in the research performed the randomization through

an online service.

A total of 513 persons reported interest in the trial out of which 403 completed the screening questionnaires (see flow chart Fig. 1). Following this, 224 persons took part in the structured telephone interview with the M.I.N.I. We excluded 27 following this interview. Excluded persons were informed via individual e-mail and were given the option to contact us if they wanted more information. A selection of persons were for ethical reasons phoned to handle reactions and immediate advice on where to seek other help.

A total of 197 participants were included with background characteristics presented in Table 2. Briefly, a majority were women (77.2%), ages ranged between 19 and 79 years (M = 34.64, SD = 13.15), and a majority had either a completed university education (43.7%) or an ongoing higher education (32.0%). A majority were either working (51.8%) or studying (33.0%), with few being on either sick leave, retired or unemployed.

2.3. Measures

Two questionnaires were included as primary outcomes to target depressive symptoms. First, we used the revised 21-item Beck Depression Inventory (BDI-II) which is designed to assess levels of depressive symptoms (Beck, Steer, & Brown, 1996). It is a widely used 21-item self-report measure of severity of depression during the last two weeks and is scored from zero to three, yielding a maximum score of 63. A score of >13 is said to indicate mild depression, a score >19 indicates moderate depression and >28 indicates severe depression (Beck et al., 1996). In the present study Cronbach's alpha was .80. Second, we used the 9-item Patient Health Questionnaire (PHQ-9) which also assesses levels of depressive symptoms (Kroenke, Spitzer, & Williams, 2001). It is scored on a four-point scale, from "Not at all" (0) to "Nearly every day" (4), with total scores ranging from 0 to 36 points. The PHQ-9 has good internal consistency, 0.89. In the present sample, Cronbach's alpha for the PHQ-9 was 0.76. This measure was also used on a weekly basis to monitor participants.

We also included a set of secondary outcomes with the first being the seven-item Generalized Anxiety Disorder (GAD-7) which measures the level of anxiety and worry and is scored on a four-point Likert-scale, "Not at all" (0) to "Nearly every day" (4) (Spitzer, Kroenke, Williams, & Lowe, 2006). The GAD-7 is often used as a screening instrument for anxiety symptoms and has an internal consistency of 0.92, and in the current study 0.84.

The 12-item Brunnsviken Brief Quality of Life Scale (BBQ; Lindner et al., 2016) was used to measure quality of life in six different domains (e.g., leisure and learning), and level of importance (e.g., "my leisure time is important to me"). The BBQ is scored on a four-point scale from "Strongly disagree" (1) to "Strongly agree" (4), with a mean score range of 0–96. The BBQ has an adequate internal consistency, 0.76 (Lindner et al., 2016), and in the current study 0.69.

The Insomnia Severity Index (ISI) measures severity of insomnia symptoms, and the impact of these symptoms on daytime functioning and distress. The ISI consists of 7 items, with a total score range of 0–28, with higher values indicating more severe insomnia (Bastien, Vallières, & Morin, 2001). The ISI has an internal consistency of 0.76, and in the present study 0.80.

A knowledge test regarding depression and CBT was created and included (unpublished material), consisting of 20 items with three response options (with only one being correct). The following are examples of items: According to CBT which method is useful to handle negative thoughts in the long run? Which one of the following correspond with the first step in the ABC-model used in CBT? Which one of these could be regarded as a primary goal in CBT? Participants were also asked to rate each response (guessing, pretty sure, definitely sure). Scoring was made with 1 point given for the correct answer which was then weighted based on certainty (0, 1, 2). As a result of this, total scores could range between −40 and 40. Cronbach's alpha for the total score
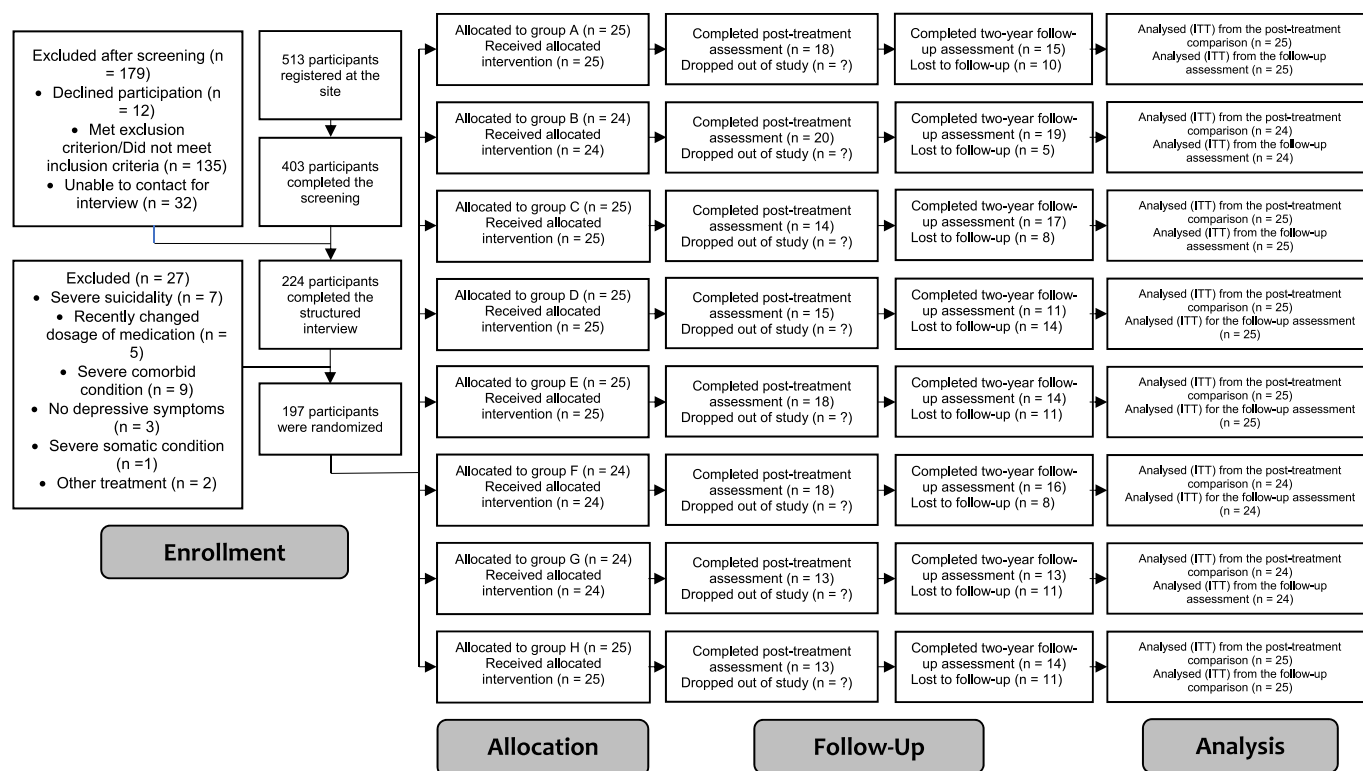
Fig. 1. Flowchart of the studies recruitment process and the outcome assessments.

was 0.89.

Finally, at post-treatment we administered the Client Satisfaction Questionnaire (CSQ-8) (Larsen, Attkisson, Hargreaves, & Nguyen, 1979), which measures satisfaction with a treatment received. It consists of 8 items scored 1–4 with higher score indicating more satisfaction (with total scores ranging from 8 to 32). The CSQ-8 has an internal consistency coefficient α = 0.91 (Attkisson & Zwick, 1982), and in the present study 0.94. We also asked our participants to rate the modules at post-treatment.

2.3.1. Treatment

All materials, measures, and text-based communication were accessed through the study's website (Iterapi.nu). Iterapi.nu is a secure platform that was developed to deliver internet-based questionnaires, treatments, and online communication with a two-factor authentication that has been used for several years in research on ICBT (Vlaescu et al., 2016). Once included participants were sent information explaining the treatment, the support type and the treatment format to which they had been randomly assigned. They were instructed to start treatment right away and were recommended to work with one module per week for 10 weeks.

In the self-tailored treatment condition participants were presented with 15 modules and advised to select between 6 and 13 modules that they thought would suit them best based on a brief description. Once they had decided which modules to include in their treatment, they could not change the selection (which they had been informed about). The modules were all structured in similar ways and included psychoeducation and exercises aimed at the problem that the module addressed. The following modules were used: Introduction, Behavioural activation I, Behavioural activation II, Cognitive restructuring, Acceptance, Emotion regulation, Anxiety and exposure, Social anxiety, Worry, Panic, Insomnia, Perfectionism, Stress management and Closure/relapse prevention. Each module consisted of the equivalent of 10-20 pages of text with illustrations, figures etc. Participants decided which order they would work with the modules. In the clinician-tailored treatment

condition modules had been set before the start based on the clinical interview and intake meeting together with the principal investigator (GA). As for the self-tailored condition 6–13 modules could be included. In this condition all were assigned the first introduction module and the last closure/relapse prevention modules.

Support

Six M.Sc. clinical psychology students, in their last term of a five-year clinical program, provided the support under the supervision of an experienced clinical psychologist. The team could also contact a psychiatrist if needed. Two support types were included in the study: scheduled weekly and support on demand. In the weekly support, participants were instructed to send a report of their work each week and received feedback within 24 h. They could ask questions at other times as well. The supporting psychology students were instructed to keep the work with each patient within 15 min per week and to contact the participant if no report was sent at the end of the week. The support guidelines for scheduled support condition also stated that the messages should be short, focused on problem-solving difficulties and questions about the treatment. The clinicians were also instructed to use validation and give positive feedback on the work, and when possible, to refer to the information in the treatment modules rather than to add extra information not covered in the modules.

The participants in the support on demand condition were instructed to go through the treatment on their own and to contact the support if they needed help or clarifications in any way. They received automated emails on a weekly basis as reminders and information that the next module was available. If they reported increased levels of depression on the weekly PHQ-9 measures, we could contact them for safety reasons.

2.3.2. Supervision

The therapists received weekly scheduled supervision with a licensed psychologist/expert on ICBT during the whole treatment period with a total of 8 sessions. Participants were randomized to being mentioned

during these sessions or not being mentioned. However, client security was not compromised and hence it was possible to intervene if a client, regardless of condition, would deteriorate. Moreover, therapists were not randomized and could thus use supervision advice for clients not being in the supervised condition. The supervision was in the form of process supervision with supervision questions being prepared in advance and a brief report of the clients that were in the supervised condition. As an additional security it was also possible to consult the psychiatrist in the team in case questions regarding medical issues would emerge.

2.4. Statistical analyses

Statistical analyses were conducted using R (R Core Team, 2019) and SPSS, version 27. The alpha level for all analyses was set at 0.05. All confidence intervals are reported at 95%. The data were analysed according to the Intention to Treat-principle (ITT), meaning that data from all included participants were used for estimation of model parameters assuming MAR (missing at random). For the fixed effects, significance testing relied on Wald's test, in which the unstandardized estimate is divided by the standard error and tested against a z-distribution. Inferences about the random effects (intercept, slope, and correlation between intercept and slope) were not made with a Wald test, but rather from the estimated confidence intervals where an interval not containing zero is interpreted in the same way as a significant *p*-value. Estimation of all parameters made use of restricted maximum likelihood estimation from the *lme4* package. Confidence intervals were calculated using the profile method in the *confint.Mermod* function from the *lme4* package.

To investigate change during the treatment and differences between the different conditions we built mixed effects models using the *lme4* package (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2018). For all models, we estimated the fixed effects of time and the interaction between time and the three main effects (self-tailored or clinician-tailored content, mode of therapist-support and supervision), as well as the three- and four-way interactions. Model fit for the random effects was investigated iteratively using a likelihood ratio test (returned using the *anova* function). All final models included a random intercept. For the PHQ outcome scores (for which we had weekly measurements), the final model included a linear rate of change, random slopes, and a correlation between intercept and slope. To ensure that the interpretations of the main effects and the interactions were independent, we used effect coding (Kugler, Dziak, & Trail, 2018) with the conditions coded as scheduled support $= -0.5$, support-on-demand $= 0.5$, clinician-tailored content $= -0.5$, self-tailored content $= 0.5$, and supervision available $= -0.5$, no supervision available $= 0.5$. Change during the follow-up phase (from post-treatment to the two-year follow-up) was investigated with a second timepiece (i.e. a piecewise model; Raudenbush & Bryk, 2002). Piecewise models allow estimation of distinct trajectories during different phases of a study (e.g., one during the active treatment, one during the follow-up period). Only fixed effects were estimated for the follow-up timepiece.

Standardized effect sizes (similar to Cohen's *d)* were estimated with the model parameters and the Satterwaite degrees of freedom according to the formula $d = 2t/\text{Sqrt}(df)$. These were interpreted according to the rule-of-thumb with 0.20, 0.50, and 0.80 corresponding to a small, moderate, and large effect size, respectively (Cohen, 1988).

Reliable clinical change and deterioration were investigated according to the formula in which the pre-treatment mean is subtracted from the post-treatment mean and divided by the pooled standard deviation adjusted for the instrument's test-retest reliability (0.93 for the BDI-II, 0.81 for the PHQ-9) (Jacobson & Truax, 1991). The critical values for both BDI-II and PHQ-9 were set at $\pm$ 6 points. To investigate the potential impact of the factors on the likelihood of achieving reliable clinical change we used a logistic regression model. The -0.5 coded conditions served as references. For inferences against norm scores, we

used the cut-off for minimal severity for both measures (equal to or below 4 for the PHQ-9, equal to or below 13 on the BDI-II).

3. Results

3.1. Baseline differences, dropout, adherence and therapist time

We found one significant baseline difference between the groups, namely that participants randomized to the self-tailored treatment were significantly more likely to have the equivalent of high school education as their highest completed level of education, $\chi2 (1) = 11.06$, $p = .026$. All other comparisons were not statistically significant (all $p > .148$). There were no significant differences in the pre-treatment data presented in Table 2. There was a small group effect on the GAD-7 ($p = .049$) (see Table 4).

The rate of explicitly discontinued participation before post treatment assessment was low (N = 18, 9.1%)). However, overall dropout from assessment at post treatment was 34.5% (N = 68). As expected, the dropout rate in the scheduled support condition 24/98 (24.5%) was lower than in the support on demand condition 44/99 (44.4%), $\chi2 (1) = 8.67$, $p = . 003$. There were no differences in dropout between the clinician-tailored (32.3%) and self-tailored conditions (36.7%). At the two-year follow-up the overall dropout rate was 39.6% (N = 78).

The average therapist time devoted to each client regardless of condition was 69.8 min (SD 7.57). As intended the therapist time per client in the scheduled support condition - 126.5 min (SD = 63.25) was about 11 times larger than in the support on demand condition - 13.8 min (SD = 26.75).

Overall, the number of modules prescribed and selected ranged between 192 (introduction) to 24 (panic) (see A to D of the Online Supplementary Materials for data on ratings of all modules including prescribed vs. self-tailored). As reported in the supplement, clinician-tailored and self-tailored module selection differed with more clinicians prescribing the modules acceptance (79% vs. 62%), behavioural activation I (97% vs. 81%) and II (96% vs. 78%). Clients on the other hand selected more of the modules social anxiety (41% vs. 21%), cognitive restructuring (68% vs. 49%), and worry (62% vs. 45%).

Participants were asked to rate the completed modules which is an indication of completion above opening and not necessarily working with the module. Overall, there were only a few differences between the clinician-tailored and self-tailored module completion of selected modules (see Supplement for ratings). Some modules were used more frequently (e.g., behavioural activation) and some more rarely (e.g., panic). Here we comment on differences in uptake and completion. Fewer completed the first introduction module in the self-tailored condition 48.9% vs. 73.7% in the clinician-tailored condition ($\chi2 (1) = 12.73$, $p = .0004$). There was a difference in behavioural activation I, with half as many completing and rating this module in the self-tailored 30.6% vs. 60.6% than in the clinician-tailored condition ($\chi2 (1) = 19.05$, $p = .00002$). This was also the case for behavioural activation II with 22.4% in the self-tailored vs. 51.5% in the clinician-tailored condition ($\chi2 (1) = 17.8$, $p = .00003$). Fewer also completed and rated the acceptance module in the self-tailored condition 14.3% vs. 28.3% in the clinician-tailored condition ($\chi2 (1) = 5.75$, $p = .02$). Thus in sum, the self-tailored condition differed from the clinician-tailored condition on 4/15 completed and rated modules but was largely similar for the remaining modules.

3.2. Symptoms of depression

We report the main effect of time for the sample as whole and the interaction between time and each factor. Model parameters for all two-, three- and four-way interactions can be seen in Appendix H of the Online Supplementary Materials. Means and standard deviations for the different factors are presented in Table 3. Means and standard deviations for each group are available in Appendix E and F of the Online

Table 2
Demographic characteristics (n = 197).

	Overall		Group A		Group B		Group C		Group D	
	M	SD	M	SD	M	SD	M	SD	M	SD
Age	34.59	13.19	31.96	10.59	34.96	13.44	35.28	13.73	30.84	9.92
	N	%	n	%	n	%	n	%	n	%
Gender identity										
Female	152	77.2	21	84	21	87.5	18	72	20	80
Male	44	22.3	4	16	3	12.5	7	28	5	20
Other	0	0	0	0	0	0	0	0	0	0
Educational level										
Primary school	3	1.5	1	4	0	0	0	0	0	0
High school	28	14.2	1	4	1	4.2	4	16	2	8
Vocational education	17	8.6	3	12	1	4.2	4	16	5	20
University (ongoing)	63	32	10	40	8	33.3	9	36	8	32
University (finished)	86	43.7	10	40	14	58.3	8	32	10	40
Current psychotropic medication: *% yes*	30	15.2	1	4	3	12.5	6	24	6	24
Prior psychological treatment: *% yes*	105	53.3	13	52	10	41.7	15	60	16	64
Somatic illness: *% yes*	10	5.1	2	8	1	4.2	1	4	0	0

	Group E		Group F		Group G		Group H		ANOVA	
	M	SD	M	SD	M	SD	M	SD	F	p
Age	34.24	13.53	37.67	17.80	36.17	12.92	35.84	12.83	0.71	.667
	n	%	n	%	n	%	n	%	χ^2	P
Gender										
Female	17	68	18	75	17	70.8	20	80	10.28	.700
Male	7	28	6	25	7	29.2	5	20		
Other	1	4	0	0	0	0	0	0		
Educational level										
Primary school	1	4	0	0	0	0	1	4	29.82	.372
High school	6	24	3	12.5	7	29.2	4	16		
Vocational education	1	4	2	8.3	0	0	1	4		
University (ongoing)	7	28	9	37.5	4	16.7	8	32		
University (finished)	10	40	10	41.7	13	54.2	11	44		
Current psychotropic medication: *% yes*	2	8	2	8.3	5	20.8	5	20	18.45	.187
Prior psychological treatment: *% yes*	10	40	13	54.2	15	62.5	2	8	14.46	.416
Somatic illness: *% yes*	1	4	2	8.3	2	8.3	1	4	3.06	.880

Note. Groups are as follows: A) clinican-selected treatment, scheduled support, supervised client; B) clinican-selected treatment, scheduled support, no supervision; C) clinican-selected treatment, on demand support, supervised client; D) clinican-selected treatment, on demand support, no supervision; E) client-selected treatment, scheduled support, supervised client; F) client-selected treatment, scheduled support, no supervision; G) client-selected treatment, on demand support, supervised client; H) client-selected treatment, on demand support, no supervision.

Table 3
Observed means for the outcome measures for the different levels of the factors.

	Factor											
Outcome measure	Self-/Therapist-chosen				Regular therapist support/Support-on-demand				Supervision available/not available			
	ST		CT		RS		SD		SA		NS	
BDI	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Pretreatment	26.40	7.47	26.25	7.45	26.27	7.12	26.38	7.78	26.68	7.37	25.97	7.53
Posttreatment	12.32	9.50	15.28	10.67	13.34	10.41	14.51	9.93	14.42	10.91	13.26	9.47
Two-year follow-up	11.98	9.60	12.20	10.18	10.72	9.08	13.75	10.58	12.27	10.11	11.92	9.69
PHQ-9												
Pretreatment	13.00	4.62	13.09	4.92	13.22	4.59	12.87	4.94	13.25	4.80	12.84	4.73
Posttreatment	5.83	4.70	7.31	5.55	6.44	5.63	6.77	4.54	6.97	5.68	6.18	4.64
Two-year follow-up	5.68	4.38	6.95	6.00	5.97	5.11	6.79	5.53	6.05	5.40	6.62	5.21
BBQ												
Pretreatment	34.50	16.51	34.89	14.98	34.65	15.52	34.74	15.99	35.15	16.47	34.23	15.00
Posttreatment	51.75	22.32	46.08	20.19	50.23	22.18	46.91	20.21	50.23	21.11	47.43	21.68
Two-year follow-up	51.93	24.32	55.81	22.10	57.37	21.90	49.79	24.21	56.28	20.22	51.60	25.75
GAD-7												
Pretreatment	9.12	4.76	9.54	4.85	9.27	4.61	9.39	5.00	10.13	4.50	8.52	4.98
Posttreatment	5.10	4.53	5.82	4.77	5.40	5.05	5.56	4.09	5.81	5.15	5.13	4.11
Two-year follow-up	4.43	4.95	4.78	4.77	4.59	4.82	4.64	4.92	4.97	5.41	4.26	4.22
ISI												
Pretreatment	12.69	5.39	13.21	5.37	12.89	5.57	13.01	5.20	13.39	5.50	12.50	5.23
Posttreatment	8.64	5.64	9.49	6.73	8.47	6.18	9.91	6.23	8.97	5.88	9.19	6.57

Note. ST = Self-tailored content; CT =Clinican-tailored content; RS = Regular therapist-support; SD; Support-on-demand; SA = Supervision available; NS = No support available; BDI-II = Becks Depression Inventory, Second version; PHQ-9 = Patient Health Questionnaire 9-item scale; BBQ = Brunnsviken Brief Quality of Life Questionnaire; GAD-7 = Generalized Anxiety Disorder 7-item Scale; ISI = Insomnia Severity Index.

Table 4
Observed means for the secondary outcome measures.

Group	BBQ								
	Pre-treatment			Post-treatment			Two-year follow-up		
	M	SD	N	M	SD	n	M	SD	n
A	36.56	36.56	25	49.58	24.95	19	61.75	19.95	16
B	34.83	34.83	24	49.23	17.30	22	58.37	24.41	19
C	36.72	36.72	25	47.00	17.28	15	51.35	18.64	17
D	31.44	31.44	25	36.13	18.72	15	49.64	25.67	11
E	32.44	32.44	25	51.30	20.56	20	60.07	15.36	15
F	34.79	34.79	24	50.89	27.18	19	49.76	25.30	17
G	34.87	34.87	24	52.87	21.70	15	51.62	26.31	13
H	35.92	35.92	25	52.38	20.25	13	46.53	28.71	15
GAD-7									
	Pre-treatment			Post-treatment			Two-year follow-up		
Group	M	SD	N	M	SD	n	M	SD	n
A	9.48	4.20	25	5.58	5.48	19	4.76	5.39	17
B	7.75	4.74	24	5.36	5.30	22	3.26	2.56	19
C	10.72	4.84	25	6.20	4.68	15	5.71	5.86	17
D	10.12	5.35	25	6.40	3.18	15	6.00	4.78	11
E	10.36	4.66	25	6.00	5.61	20	5.33	6.25	15
F	9.42	4.74	24	4.63	3.83	19	5.24	4.86	17
G	9.96	6.80	24	5.47	5.00	15	3.85	4.00	13
H	6.80	4.61	25	5.58	5.48	14	3.13	4.45	15
ISI									
	Pre-treatment			Post-treatment			Two-year follow-up		
Group	M	SD	N	M	SD	n	M	SD	n
A	13.50	5.90	25	7.79	6.32	19	6.88	5.58	16
B	13.83	5.42	24	11.23	6.84	22	10.05	5.76	19
C	13.68	5.60	25	10.13	6.29	15	6.94	4.74	17
D	11.83	4.56	24	8.47	7.42	15	9.00	6.81	11
E	12.64	5.02	25	7.65	5.48	20	6.13	5.22	15
F	11.63	5.99	24	6.84	5.28	19	8.47	5.27	17
G	13.75	5.74	24	11.07	5.09	15	8.77	5.22	13
H	12.76	4.90	25	10.00	6.27	13	9.13	7.56	15
Knowledge									
	Pre-treatment			Post-treatment			Two-year follow-up		
Group	M	SD	N	M	SD	n	M	SD	n
A	12.60	9.90	22.84	22.84	11.20	19	17.20	10.37	15
B	8.50	6.51	25.19	25.19	8.30	22	19.63	8.37	19
C	10.04	5.81	22.50	22.50	9.81	15	16.65	7.98	17
D	12.04	9.27	24.27	24.27	8.22	15	22.18	10.80	11
E	8.72	6.99	21.10	21.10	10.48	20	21.07	8.91	14
F	6.38	5.22	22.05	22.05	9.94	19	17.38	8.16	16
G	8.50	8.50	17.64	17.64	5.09	15	13.31	8.65	13
H	10.08	9.00	17.30	17.30	9.84	13	14.36	8.83	14

Note. BBQ = Brunnsviken Brief Quality of Life Questionnaire; GAD-7 = Generalized Anxiety Disorder 7-item Scale; ISI = Insomnia Severity Index.

Supplementary Materials. The results from the BDI-II model showed a significant heterogeneity in starting levels of depression symptoms, *SD* = 5.23 [95% CI 4.16, 6.13]. There was a significant effect of time during the treatment period with an estimated mean reduction of −12.70 [95% CI -14.32, −11.09], SE = 0.84, $p < .001$. The within-group effect size was $d = -1.73$ [95% CI -1.95, −1.51]. The interaction between time and tailoring condition showed a significantly steeper decline for the self-tailored group, compared to the clinician-tailored group, with an estimated mean difference of −3.85 [95% CI -7.09, −0.62], SE = 1.69, $p = .023$. The between-group effect size for this difference was small favouring the self-tailored group, $d = -0.26$ [95% CI -0.48, −0.04]. There were no other significant interaction effects.

The results from the PHQ-9 model showed a significant heterogeneity in intercept, *SD* = 4.81 [95% CI 4.18, 5.30], and slope, *SD* = 0.42 [95% CI 0.34, 0.48]. Additionally, there was a significant negative correlation between the two, indicating that participants with higher ratings at pre-treatment had a steeper decline during the treatment, $r = -0.59$ [95% CI -0.70, −0.44]. For the fixed effects, there was a significant decrease in symptom ratings per unit of time (one week), with an estimated mean difference of −0.48 [95% CI -0.56, −0.40], SE = 0.04, $p > .001$. The within-group effect size was large, $d = 1.64$ [95% CI -1.36,

−1.91]. No other interaction effects were statistically significant.

3.3. Generalized anxiety

Means and standard deviations for the secondary outcome measures are presented in Table 4. For the GAD-7, there was significant heterogeneity in initial ratings, *SD* = 2.87 [95% CI 2.28, 3.35]. For the fixed effects, there was a significant decrease during the treatment period, with an estimated mean difference of −4.05 [95% CI -4.89, −3.21], SE = 0.44, $p < .001$. The within-group effect size was large, $d = -1.10$ [95% CI -0.87, −1.32]. Interaction effects were not statistically significant.

3.4. Quality of life

For the BBQ results indicated a significant heterogeneity in starting levels, *SD* = 13.10 [95% CI 10.86, 14.89]. There was a significant increase for the whole sample during the treatment period, estimated mean difference = 14.33 [95% CI 11.08, 17.57], SE = 1.70, $p > .001$. The within-group effect size was large, $d = 0.99$ [95% CI 0.77, 1.21]. The interaction between time and therapist/self-tailored content was significant with an estimated mean difference = 6.68 [95% CI 0.18,

13.17], SE = 3.40, p = .05. The effect size for this comparison was small and favoured the group with self-tailored content, d = 0.23 [95% CI 0.00, 0.46]. We found no other interaction effects.

3.5. Sleep

There was significant heterogeneity in initial ratings on the ISI, SD = 3.82 [95% CI 3.15, 4.38]. For the fixed effects, there was a significant decrease on the ISI, b = −4.03 [95% CI -4.98, −3.07], SE = 0.50, p > .001. The within-group effect size was large, d = 0.96 [95% CI -0.73, −1.08]. There were no interaction effects.

3.6. Follow-up

Results from the second timepiece model did not show a further decrease or increase of symptoms of depression on the BDI-II, estimated mean difference = −1.44 [95% CI -3.23, 0.36], SE = 0.94, p = .127. There were no interaction effects.

The results for the PHQ-9 model indicated no significant change during this phase, estimated mean difference = 0.47 [95% CI -0.23, 1.16], SE = 0.36, p = .187. As for the BDI-II there were no significant interactions.

The GAD-7 model did not further decrease during the follow-up phase, and there were no interaction effects. For quality of life, the BBQ model showed no change during the follow-up phase. The interaction between time and therapist/self-tailored content was significant with participants in the clinician-tailored condition exhibiting a significant increase relative to the self-tailored condition, estimated mean difference = −9.89 [95% CI -17.10, −2.68], SE = 3.77, p = .009. The effect size for this comparison was small and in favour of the group with clinician-tailored content, d = −0.32 [95% CI -0.09, −0.55]. There were no other interaction effects.

The ISI model did not show a change during the follow-up phase, with no interaction effects.

3.7. Knowledge

On the weighted knowledge test result regarding depression and CBT scores there was significant heterogeneity in initial ratings, SD = 6.52 [95% CI 4.48, 6.65]. For the fixed effects, there was a significant increase during the treatment period, estimated mean difference = 11.81 [95% CI 10.54, 13.43], SE = 0.67, p < .001. The within-group effect size was large, d = 2.14 [95% CI 1.91, 2.38]. The interaction between time and mode of support was statistically significant, with an estimated mean difference of −3.29 [95% CI -6.05, −0.28], SE = 1.35, p = .015. The effect size for this comparison was small and favoured the group with scheduled support, d = −0.30 [95% CI -0.53, −0.07]. One of the three-way interactions was significant, with results indicating that participants with support on demand and not discussed during supervision had lower knowledge score at post-treatment compared to the participants in the opposite other conditions, estimated mean difference = −6.34 [95% CI -12.04, −0.50], SE = 2.70, p = .019. The effect size for this comparison was small, d = −0.29 [95% CI -0.52, −0.05].

In the follow-up phase, there was a significant reduction in knowledge scores, estimated mean difference = −4.29 [95% CI -12.04, −0.50], SE = 0.75, p < .001. This corresponds to an effect size of d = −0.72 [95% CI -0.48, −0.96]. There were no interaction effects.

3.8. Satisfaction with treatment and rating of modules

Ratings of satisfaction with the treatment according to CSQ-8 are presented in Appendix G of the Online Supplementary Materials. Scores ranged between 21.79 and 25.71, which indicates good overall satisfaction (with scores between 20 and 25) (Smith et al., 2014). There was no main effect of group on the CSQ ratings or significant interactions.

Rating of modules in terms of helpfulness, fit with needs, the percentage that would recommend the module to a friend with similar problems and an overall rating is presented in Appendix A to Dof the Online Supplementary Materials. Overall, ratings were high but there were also a few differences between the self-tailored and clinician-tailored groups. For all four the self-tailored group provided lower ratings (fit with need and overall rating of the acceptance module, and the same for emotional awareness).

3.9. Reliable and clinically significant change

Overall, 74.1% of the participants with complete post-treatment data (n = 139) met the criteria for reliable clinical change on the BDI-II outcome measure (with dropouts regarded as non-improved 52.3% showed reliable change). For PHQ-9, this number was 59.9% (n = 85 out of the 142 with available data; 43.1% including dropouts). Only three participants (2.2%) reported reliable clinical deterioration according to the BDI-II scores. This was also the case for PHQ-9, with three participants (2.2%) reporting reliable deteriorating. One participant (<1%) met the criterion for reliable deterioration for both outcome measures, while the other four participants met the criterion on one of them, but not the other.

Comparing against norm scores, 44.4% (n = 63) of the sample with post-treatment data scored within the category (a sum of 0–4) indicating minimal depression severity on the PHQ-9. For BDI-II, 55.4% (n = 77) of the sample scored within the minimal depressive severity category (a sum of 0–13 points).

The logistic regression model for the BDI-II outcome measure did not indicate a significant difference for any of the factors when investigating the odds of having undergone reliable clinical change. Likewise, the PHQ-9 model did not indicate any significant differences in the probability of achieving reliable clinical change.

As part of the protocol the CGI questions were asked during the post-treatment telephone interview (N = 131). Dividing the CGI into three categories: improved, no change and deterioration, 111 (84.73%) reported improvement, 15 (11.45%) no change, and 5 (3.87%) deterioration. Regarding the dropouts as either non-improved or deteriorated the proportion decreased to 56.3% showing improvement. There were no significant differences on this measure between the groups.

4. Discussion

The aims of this factorial design trial were to investigate if self-tailored treatment, support on demand and case-specific supervision would make a difference compared to clinician-tailored treatment, scheduled support and no supervision, respectively. Overall, we found large reductions of depressive symptoms across conditions and few differences or interactions. What stands out as surprising and unique is finding that the clients could select treatment content and that this not only yielded same results as when therapists selected the treatment, but also a small but statistically significant effect in favour of self-tailoring. We will start by discussing this finding and then move on to the other two contrasts with a focus on the primary outcomes. After this we will discuss the secondary outcomes and limitations.

4.1. Self-tailoring

In the meta-analysis by Delevry and Le (2019) the effect size in favour of the preferred treatment was very similar to what we found here (0.18 in the meta-analysis and 0.26 in our study). On the other hand it was only on the BDI-II and BBQ for which we could establish this effect and it is best to assume that self-tailored treatment selection can be as effective as clinician tailored, which is in line with our previous GAD study (Dahlin et al., 2022) and our early open pilot study on mixed anxiety (Andersson et al., 2011). We also did not find any difference on the PHQ-9 which could reflect that this measure is less sensitive to detect small differences, but also note the discrepancy that the BDI-II did not

yield the same outcome. Another aspect to consider is that we tested tailoring of content rather than treatment brand (e.g., CBT vs. psychodynamic treatment) or category (psychotherapy vs. medication), a topic that has been less studied in psychotherapy research overall. It is interesting to note that the results were similar given the potential risk that participants in the self-tailored condition could have picked ineffective modules or the same modules as the clinicians rendering choice less important. The modules selected most likely were sufficiently effective even if adherence was poorer in the self-tailored condition. With regards to own choice versus choice made by clinicians a few things can be noted. Given the evidence for behavioural activation it was striking that this module was selected less often by the clients than the clinicians (see supplement). However, there was no difference in the ratings of the behavioural activation modules. Given that half of the participants in the trial had some prior experience of psychological treatments – which in Sweden often is CBT and behavioural activation – we assume that at least some of our participants may have picked modules that were different from what they had previously experienced or read about. We did not find that clients chose the same modules as the clinicians. In addition to the difference in selecting behavioural activation clients more often selected social anxiety, worry and cognitive restructuring, and less often the introduction and acceptance modules. This indicates that self-tailoring is meaningful and does not result in the same treatment as when clinicians make the decision. Research on depression indicate that different psychotherapies can work equally good or bad (Cuijpers et al., 2021). Given this it can be hard to find any differences in a depression trial and equal effects is a likely finding. We are also aware of the fact that the difference between self-tailored and clinician-tailored treatment may not be that great as the clinicians made choices based on client interviews and pre-treatment questionnaire data.

4.2. Support on demand

As expected, clinicians spent much less time in the support on demand condition (which in theory is not obvious as clients were free to contact us). Even if there were no differences in depression outcomes it is clear the support on demand was associated with more dropout (44.4% vs. 24.5% in the scheduled support condition). The findings in the literature are mixed with some ICBT studies finding no differences in dropout rates from assessment (Hadjistavropoulos et al., 2019) and some showing more dropout when guidance was offered on demand (Zagorscak et al., 2018). Studies differ in how much participants are encouraged to complete outcome measures as post treatment and in this trial very few explicitly stated that they wanted to drop out, but several did not complete outcomes anyway. It is interesting to note that some did return for the two-year follow-up and overall dropout at this stage was not markedly larger than at post treatment (39.6% vs. 34.5%). Another aspect to consider is the role of automated reminders as this can be interpreted as guidance even if it is not personalized and also make use of persuasive technology. Previous studies suggest that this can be a way to boost unguided treatments (Kelders et al., 2015), which may partly explain why unguided treatments appear to yield better results now than in earlier research (Andersson & Cuijpers, 2009).

4.3. Supervision

There were basically no effects of the supervision condition. It is likely that this was not sufficiently manipulated. Another approach would have been to randomize clinicians instead of participants. The likelihood of advice for one client being useful for another client is obvious. We encourage more experimental research on the role of supervision in CBT in general as there are very few controlled studies (Alfonsson et al., 2018).

4.4. Results on secondary measures

For the GAD-7 there was a large within-group effect ($d = -1.10$) but no differential effects of the conditions. Quality of life as measured by the BBQ also showed a large effect ($d = -0.90$), and for this measure we also found a small effect in favour of the self-tailored condition ($d = 0.20$). Somewhat surprisingly given that the insomnia was prescribed/self-tailored by about half of the participants (53.2%) there was also a large effect on the ISI ($d = 0.96$). This is an argument in favour of tailoring in general as many CBT depression protocols do not include insomnia management.

In line with our previous research (Berg et al., 2020), there were large improvements in knowledge scores ($d = 2.14$), and a small effect in favour of the scheduled support condition ($d = 0.30$). The lack of difference between self-vs. clinician-tailored treatment could indicate that self-tailoring does not lead to less knowledge acquisition. It is interesting to note that knowledge scores decreased at two-year follow-up ($d = -0.72$). This calls for more research on the stability of knowledge gains following treatment. We acknowledge these results as preliminary as the test was developed for this study.

Treatment satisfaction on the CSQ-8 and ratings of modules suggested that the treatment and the modules were appreciated overall.

We note that the follow-up results at two year follow-up overall suggest that treatment effects were maintained, which is in line with previous ICBT research (Andersson, Rozental, Shafran, & Carlbring, 2018), but also note that there were no differential effects based on our independent variables with the exception of a small effect on the BBQ in favour of the clinician-tailored group.

Clinical and reliable change was observed for a majority when it comes to reliable change and about half with regards to norms on the depression outcomes, with no differences between the independent variables. The clinical interview with the CGI largely confirmed the self-report findings. Finally, few participants deteriorated.

4.5. Limitations and strengths

A first limitation is statistical power. In spite of being a large trial compared to how psychotherapy research has been done previously the trial would have benefitted from a larger sample. Power calculations for factorial design trials are complicated by the discrepancy between power for main effects and the power for interaction effects. We did not conduct a power analysis before the trial was conducted and in that sense the trial can be regarded as a large pilot investigation only powered to detect moderate group differences ($d = 0.40$ or larger). Moreover, dropout rate was larger in one of the conditions which of course has consequences for the trial even if we used an ITT analytic approach. Another limitation is the recruitment of participants as we recruited from the general public but reached a more educated and somewhat younger group than would have been more clinically representative. Yet another limitation is that we did not report or focus on moderators and mediators of change. Instead, and given that this in some respects was a pilot trial preparing for a second trial, we added measures to capture different outcomes. Other possible predictors of change such as therapeutic alliance, credibility and expectations were also not measured due to measurement overload. We also did not conduct any open ended qualitative interviews which would give us information about how the process of self-selecting treatment modules is experienced by the clients. A final limitation is the obvious lack of a control group receiving either no treatment or a placebo treatment. As much is known regarding the effects of ICBT for depression we do not regard this a major limitation but it cannot be excluded that regression to the mean play a role, in particular at the two-year follow-up.

In spite of these limitations, we believe the trial has some strengths. First, we reported large within-group effects, which given the need for treatment cannot be ignored as ICBT can be a cost-effective complement and alternative to standard ways of delivering CBT. Second, we added a

long-term follow-up and asked our participants about their view of the treatment modules and their knowledge. Third, the factorial design is also a strength and a way to help us to understand in an experimental manner what works and how.

5. Conclusions

We conclude that treatment clients most likely can be more involved in tailoring treatment content in ICBT for depression, perhaps even deciding themselves. Tailoring treatment can be a way to handle co-morbid problems and is not the same as a transdiagnostic treatment that may work for different problems. We hesitate to conclude that support on demand is as effective as scheduled guidance, but it is likely that there are clients for whom optional support is enough. Future research will hopefully tell us who these are in advance to reduce dropout and increase effects.

Author note

CRediT author statement

Gerhard Andersson – PI, clinical supervisor, author of MS, initiator of the study, Anton Käll – statistical analyses, write-up of results, Simon Juhlin, Carl Wahlström, Edvard de Fine Licht, Simon Färdeman, Anna Franck, Anna Tholcke, Karin Nachtweij, Emma Fransson, - MSC students working in the trial with interviews, treatment, module reconstruction and also commenting on ms and analyses, Kristofer Vernmark – super-vision of therapists, planning of study, comment on ms. Mikael Lud-vigsson, medical doctor, clinical supervision, safety management, comment on ms, Matilda Berg, responsible for the knowledge test, su-pervision, clinical follow-up interviews, comment on ms, All authors have read and approved the submitted ms.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.brat.2023.104265.

References

Alfonsson, S., Parling, T., Spännargård, Å., Andersson, G., & Lundgren, T. (2018). The effects of clinical supervision on supervisees and patients in cognitive behavioral therapy: A systematic review. *Cognitive Behaviour Therapy, 47*, 206–228. https://doi.org/10.1080/16506073.2017.1369559

Andersson, G. (2018). Internet interventions: Past, present and future. *Internet Interventions, 12*, 181–188. https://doi.org/10.1016/j.invent.2018.03.008

Andersson, G., & Berger, T. (2021). Internet approaches to psychotherapy: Empirical findings and future directions. In M. Barkham, W. Lutz, & L. G. Castonguay (Eds.), *Bergin and garfield's Handbook of Psychotherapy and behavior change (50th anniversary* ed., pp. 749–772). Wiley.

Andersson, G., & Cuijpers, P. (2009). Internet-based and other computerized psychological treatments for adult depression: A meta-analysis. *Cognitive Behaviour Therapy, 38*, 196–205. https://doi.org/10.1080/16506070903318960

Andersson, G., Estling, F., Jakobsson, E., Cuijpers, P., & Carlbring, P. (2011). Can the patient decide which modules to endorse? An open trial of tailored internet treatment of anxiety disorders. *Cognitive Behaviour Therapy, 40*, 57–64. https://doi.org/10.1080/16506073.2010.529457

Andersson, G., Rozental, A., Shafran, R., & Carlbring, P. (2018). Long-term effects of Internet-supported cognitive behavior therapy. *Expert Review of Neurotherapeutics, 18*, 21–28. https://doi.org/10.1080/14737175.2018.1400381

Attkisson, C. C., & Zwick, R. (1982). The client satisfaction questionnaire: Psychometric properties and correlations with service utilization and psychotherapy outcome. *Evaluation and Program Planning, 5*, 233–237. https://doi.org/10.1016/0149-7189(82)90074-X

Bastien, C. H., Vallières, A., & Morin, C. M. (2001). Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Medicine, 2*, 297–307.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck depression inventory-II*. Psychological Corporation.

Berg, M., Rozental, A., de Brun Mangs, J., Näsman, M., Strömberg, K., Viberg, L., et al. (2020). The role of learning support and chat-sessions in guided internet-based cognitive behavioural therapy for adolescents with anxiety: A factorial design study. *Frontiers in Psychiatry, 11*, 503. https://doi.org/10.3389/fpsyt.2020.00503

Busner, J., & Targum, S. D. (2007). The clinical global impressions scale: Applying a research tool in clinical practice. *Psychiatry, 4*, 28–37.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2 ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Cuijpers, P., Karyotaki, E., Ciharova, M., Miguel, C., Noma, H., & Furukawa, T. A. (2021). The effects of psychotherapies for depression on response, remission, reliable change, and deterioration: A meta-analysis. *Acta Psychiatrica Scandinavica, 144*, 288–299. https://doi.org/10.1111/acps.13335

Dahlin, M., Johansson, A., Romare, H., Carlbring, P., & Andersson, G. (2022). Worry-specific versus self-tailored internet-based treatments for generalized anxiety disorder, with scheduled support or support on demand: A pilot factorial design trial. *Internet Interventions, 28*, Article 100531. https://doi.org/10.1016/j.invent.2022.100531

Delevry, D., & Le, Q. A. (2019). Effect of treatment preference in randomized controlled trials: Systematic review of the literature and meta-analysis. *Patient, 12*, 593–609. https://doi.org/10.1007/s40271-019-00379-6

Drozd, F., Vaskinn, L., Bergsund, H. B., Haga, S. M., Slinning, K., & Bjorkli, C. A. (2016). The implementation of internet interventions for depression: A scoping review. *Journal of Medical Internet Research, 18*, e236. https://doi.org/10.2196/jmir.5670

Hadjistavropoulos, H. D., Peynenburg, V., Thiessen, D. L., Nugent, M., Karin, E., Dear, B. F., et al. (2022). A randomized factorial trial of internet-delivered cognitive behavioural therapy: An 8-week program with or without extended support and booster lesson. *Internet Interventions, 27*, Article 100499. https://doi.org/10.1016/j.invent.2022.100499

Hadjistavropoulos, H. D., Schneider, L. H., Edmonds, M., Karin, E., Nugent, M. N., Dirkse, D., et al. (2017). Randomized controlled trial of internet-delivered cognitive behaviour therapy comparing standard weekly versus optional weekly therapist support. *'Journal of Anxiety Disorders, 52*, 15–24. https://doi.org/10.1016/j.janxdis.2017.09.006

Hadjistavropoulos, H. D., Schneider, L. H., Mehta, S., Karin, E., Dear, B. F., & Titov, N. (2019). Preference trial of internet-delivered cognitive behaviour therapy comparing standard weekly versus optional weekly therapist support. *Journal of Anxiety Disorders, 63*, 51–60. https://doi.org/10.1016/j.janxdis.2019.02.002

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12–19. https://doi.org/10.1037/0022-006X.59.1.12

Johansson, R., Sjöberg, E., Sjögren, M., Johnsson, E., Carlbring, P., Andersson, G., et al. (2012). Tailored vs. standardized internet-based cognitive behavior therapy for depression and comorbid symptoms: A randomized controlled trial. *PLoS One, 7*(5), Article e36905. https://doi.org/10.1371/journal.pone.0036905

Kelders, S. M., Bohlmeijer, E. T., Pots, W. T., & van Gemert-Pijnen, J. E. (2015). Comparing human and automated support for depression: Fractional factorial randomized controlled trial. *Behaviour Research and Therapy, 72*, 72–80. https://doi.org/10.1016/j.brat.2015.06.014

Kleiboer, A., Donker, T., Seekles, W., van Straten, A., Riper, H., & Cuijpers, P. (2015). A randomized controlled trial on the role of support in Internet-based problem solving therapy for depression and anxiety. *Behaviour Research and Therapy, 72*, 63–71. https://doi.org/10.1016/j.brat.2015.06.013

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*, 606–613. jgi01114.

Kugler, K. C., Dziak, J. J., & Trail, J. (2018). Coding and interpretation of effects in analysis of data from a factorial experiment. In LM. C, & K. C. Kugler (Eds.), *Optimization of behavioral, biobehavioral, and biomedical interventions: Advanced topics* (pp. 175–205). Springer International Publishing.

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

Lindner, P., Frykheden, O., Forsström, D., Andersson, E., Ljótsson, B., Hedman, E., et al. (2016). The Brunnsviken Brief Quality of life scale (BBQ): Development and psychometric evaluation. *Cognitive Behaviour Therapy, 45*, 182–195. https://doi.org/10.1080/16506073.2016.1143526

Malhi, G. S., & Mann, J. J. (2018). Depression. *Lancet, 392*, 2299–2312. https://doi.org/10.1016/S0140-6736(18)31948-2

Oromendia, P., Orrego, J., Bonillo, A., & Molinuevo, B. (2016). Internet-based self-help treatment for panic disorder: A randomized controlled trial comparing mandatory versus optional complementary psychological support. *Cognitive Behaviour Therapy, 45*, 270–286. https://doi.org/10.1080/16506073.2016.1163615

Păsărelu, C., Andersson, G., Bergman Nordgren, L., & Dobrean, A. (2017). Internet-delivered transdiagnostic and tailored cognitive behavioral therapy for anxiety and

depression: A systematic review and meta-analysis. *Cognitive Behaviour Therapy, 46*, 1–28. https://doi.org/10.1080/16506073.2016.1231219

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team. (2018). nlme: Linear and nonlinear mixed effects models. https://CRAN.R-project.org/package=nlme.

R Core Team. (2019). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models : Applications and data analysis methods.* Thousand Oaks, CA: Sage Publications.

Schuster, R., Kaiser, T., Terhorst, Y., Messner, E. M., Strohmeier, L.-M., & Laireiter, A.-R. (2021). Sample size, sample size planning, and the impact of study context: Systematic review and recommendations by the example of psychological depression treatment. *Psychological Medicine, 51*, 902–908. https://doi.org/10.1017/S003329172100129X

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., et al. (1998). The mini-international neuropsychiatric interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry, 59*(Suppl 20), 22–33.

Smith, D., Roche, E., O'Loughlin, K., Brennan, D., Madigan, K., Lyne, J., et al. (2014). Satisfaction with services following voluntary and involuntary admission. *Journal of Mental Health, 23*, 38–45. https://doi.org/10.3109/09638237.2013.841864

Spitzer, R. L., Kroenke, K., Williams, J. B., & Lowe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine, 166*, 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Thew, G. R., Powell, C. L., Kwok, A. P., Lissillour Chan, M. H., Wild, J., Warnock-Parkes, E., et al. (2019). Internet-based cognitive therapy for social anxiety disorder in Hong Kong: Therapist training and dissemination case series. *JMIR formative research, 3*, Article e13446. https://doi.org/10.2196/13446

Titov, N., Hadjistavropoulos, H., Nielssen, O., Mohr, D., Andersson, G., & Dear, B. (2019). From research to practice: Ten lessons in delivering digital mental health services. *Journal of Clinical Medicine, 8*, 1239. https://doi.org/10.3390/jcm8081239

Vlaescu, G., Alasjö, A., Miloff, A., Carlbring, P., & Andersson, G. (2016). Features and functionality of the Iterapi platform for internet-based psychological treatment. *Internet Interventions, 6*, 107–114. https://doi.org/10.1016/j.invent.2016.09.006

Watkins, E. R., & Newbold, A. (2020). Factorial designs help to understand how psychological therapy works. *Frontiers in Psychiatry, 11*, 429. https://doi.org/10.3389/fpsyt.2020.00429

Zagorscak, P., Heinrich, M., Sommer, D., Wagner, B., & Knaevelsrud, C. (2018). Benefits of individualized feedback in internet-based interventions for depression: A randomized controlled trial. *Psychotherapy and Psychosomatics, 87*, 32–45. https://doi.org/10.1159/000481515