Investigation of active ingredients within internet cognitive-behavioral therapy for depression: a randomized optimization trial

Edward Watkins, PhD[1]

Alexandra Newbold PhD [1]

Michelle Tester-Jones, PhD[1]

Linda M. Collins, PhD[2]

Mohammod Mostazir, MSc[1]

[1]Sir Henry Wellcome Building for Mood Disorders Research, College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4QG, UK.
[2] School of Global Public Health, New York University, 708 Broadway, New York, NY, 10003

Corresponding author: Edward Watkins, Sir Henry Wellcome Building for Mood Disorders Research, College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4QG, UK.
E: **e.r.watkins@exeter.ac.uk**; T: +00 1392724692

Date of revision: 21/4/2023
word count for text: 3000

Key points

Question Which specific components within internet-delivered cognitive-behavioral therapy (CBT) for depression are active ingredients that reduce symptoms?

Findings: In this optimization experiment that included 767 adults with depression, six treatment components (activity scheduling, thought challenging, relaxation, concreteness training, functional analysis, self-compassion) did not show a significant main effect on depression. However, the presence of the absorption component outperformed its absence in reducing depression.

Meaning: The majority of treatment benefit from internet-CBT is likely to accrue from either factors common to all CBT components and/or from generic factors common to all therapies, with the possible exception of absorption.

Abstract

Importance: There is limited understanding of how complex evidence-based psychological interventions such as Cognitive-behavioral therapy (CBT) for depression work. Identifying active ingredients can help to make therapy more potent, briefer, and scalable.

Objective: To test the individual main effects and interactions of seven treatment components within internet-CBT for depression to investigate its active ingredients.

Design: A 32-condition balanced fractional factorial optimization trial (IMPROVE-2).

Setting: Recruitment from internet advertising and UK National Health Service Improving Access to Psychological Therapies service.

Participants: 767 adults with depression (Patient Health Questionnaire-9 (PHQ-9) scores ≥10) were randomized between 7th July 2015 - 29th March 2017 and followed for 6 months post-treatment (until 29th December 2017).

Interventions: Participants were randomized to seven experimental factors within the internet CBT platform, each reflecting the presence versus absence of specific treatment components (activity scheduling, functional analysis, thought challenging, relaxation, concreteness training, absorption, self-compassion).

Main Outcomes and Measures: Primary outcome was depression symptoms (PHQ-9). Secondary outcomes include anxiety symptoms and work and social functioning.

Results: Among 767 participants (mean age 38.5 years, SD 11.62, range 18-76; 635 [83%] women), 506 (66%) completed the 6-month post-treatment follow-up. Baseline score adjusted Analysis of Covariance model using effect–coded intervention variables (-1,+1), found no main effect on depression symptoms for the presence versus absence of activity scheduling, functional analysis, thought challenging, relaxation, concreteness training or self-compassion training (post-treatment: largest $\Delta PHQ9 =$–0.09 [90%CI, -0.56-0.39]; 6-month follow-up: largest $\Delta PHQ9 =$-0.18 [90%CI, -0.61-0.25]). Only absorption

training had a significant main effect on depressive symptoms at follow-up (post-treatment: $\Delta PHQ9 =$ 0.21 [90%CI, -0.27-0.68]; 6-month follow-up: $\Delta PHQ9 =$ -0.54, [90%CI, -0.97 to -0.11]).

Conclusions and Relevance: Six of seven components within internet-CBT did not significantly reduce depression symptoms relative to their absence, despite an overall average reduction in symptoms (post-treatment: $\Delta PHQ9 =$ -7.79, p<0.001; 6-month follow-up: $\Delta PHQ9 =$-8.63, p<0.001). Treatment benefit from internet-CBT probably accrues from spontaneous remission, factors common to all CBT components (e.g., structure; making active plans) and from non-specific therapy factors (e.g., positive expectancy), bar the possible exception of absorption focused on enhancing direct contact with positive reinforcers.

Trial registration: Current Controlled Trials ISRCTN24117387. Registered 26 August 2014. https://www.isrctn.com/ISRCTN24117387

Introduction

Major depressive disorder is a common psychiatric disorder[1] ranked as the second leading contributor of years lived with disability.[2] Despite antidepressant medication and cognitive-behavioral therapy (CBT) being our best evidence-based treatments, they achieve remission rates of under 1/3 and limited sustained recovery (50–80% relapse/recurrence)[3].

Understanding the active mechanisms of how psychotherapy works is a priority to advance treatment efficacy[3-8] by enabling the content and delivery of interventions to be optimized to make them more potent and efficient. Psychological treatments are complex interventions, made up of multiple components, and varying in structure and modality of delivery, each of which potentially influences outcome. Whilst the parallel group comparative randomized controlled trial (RCT) is the gold standard for establishing the relative efficacy of one intervention versus another or a control, this experimental design is limited at investigating the specific mechanisms of how complex psychological interventions work because it can only compare the overall effects of each intervention package: All treatment components are aggregated and confounded together in the comparison of treatment packages, such that the main effects and interactions of individual treatment components cannot be disentangled. To address this, we need alternatives to the parallel group comparative RCT, that is, equally rigorous experimental designs that enable testing with strong causal inference of the effects of the presence or absence of individual therapeutic elements within a complex intervention[8,9,10].

One alternative is a factorial optimization trial framed within the Multiphase Optimization Strategy (MOST)[10-19]. MOST is a principled and comprehensive framework for optimizing and evaluating interventions[11-15], well-validated, and with proven value in health behavior research[11,20-26]. MOST uses efficient experimentation *before* a parallel group comparative RCT, typically via factorial designs,[13-15,27] to identify which of a set of candidate components is effective and should be retained in an optimized intervention to be evaluated subsequently. Factorial experiments allow one to explore main effects of

components and interactions among components within a treatment package[27, 28], necessary for developing a mechanistic understanding of therapy and for methodically enhancing and simplifying complex interventions[28].

Factorial designs have examined different types of support during psychological interventions[29,30] but none to our knowledge have investigated treatment components within CBT for adult depression. This trial was the first factorial experiment to examine the efficacy of components within internet-CBT for depression[31] (see also[32]). The component selection reflected well-established treatment elements within CBT as identified by the Delphi technique[33] (activity monitoring and scheduling; thought challenging; applied relaxation), functional analysis as a mainstay of behavioral activation[34-36], and recent innovations (concreteness, compassion, absorption)[37-40]. Internet-CBT was selected for treatment reach and scalability, to reduce "drift" from treatment protocols and ensure patients only received the relevant standardized treatment content, minimizing the risk of patients receiving unallocated components.

This factorial approach enables an enhanced test of the relative contribution of specific versus common treatment factors[10] within psychotherapy, without requiring a prohibitive number of participants for statistical power. An unresolved debate concerns the extent to which psychotherapies work through non-specific factors shared across all therapies ("pan-therapy common factors", e.g., hope, therapeutic alliance) versus elements specific to a particular therapy[41-43], whether specific strategies (e.g., thought challenging) or features common to all components in that therapy (in CBT: cognitive-behavioral model; homework). Disentangling specific from common treatment effects has been limited by a lack of well-powered experimental studies[43] and difficulties in creating appropriate placebo controls that genuinely match an active psychotherapy for credibility and structural equivalence[44]. Well-designed factorial trials can address this issue. In a balanced factorial experiment, such as the current design, for any component (e.g., relaxation), the aggregate of the 16 conditions where it is present (Table 1, conditions 17-32) are equivalent for treatment credibility, structure, delivery, rationale, therapist

contact, all other components, and therapist allegiance with the aggregate of the 16 conditions where it is absent (Table 1, conditions 1–16). Hence, any observed main effect would be conservative evidence for a specific factor driving therapeutic benefit, above and beyond pan-therapy common factors and factors common to all CBT components.

By comparing the presence versus absence of each component, this factorial design examines the main effect of each component on outcomes. Consistent with the Pareto principle and prior studies[21], we predicted that components and interactions will vary in effect size, with many insignificant (i.e., not all specific components are active ingredients in CBT).

Methods

Study design

This study was a stratified, block randomized, single-blind, optimization trial, randomized at the patient level. The design included seven experimental factors, each corresponding to a CBT component, and each factor had two levels (presence of component coded +1 versus absence coded -1; i.e., effect coded). We used a balanced fractional factorial design (2_{IV}^{7-2}; see Table 1), a special case of the factorial design in which logistics and expense are managed by including only a carefully selected subset of experimental conditions. Here only 25% of the conditions required by a full factorial were included. The experiment was designed without a no-treatment control condition making it suitable for implementation in a clinical service. Fractional factorial designs alias, or combine, effects: here each main effect is aliased with selected three-way interactions or higher-order interactions. To interpret any observed effects, we assumed that pre-specified two-way interactions and all three-way and higher-order interactions are negligible in size. For each main effect, half of the sample were randomized to presence and half randomized to absence of the factor (see Table 1). A factorial experiment should not be considered a multi-arm parallel group comparative group RCT – its logic is different. The full sample

size is used to determine each main effect and interaction making it highly efficient for power and sample size because each effect estimate involves all the conditions (½ cells aggregated for presence of component; ½ cells aggregated for absence of component). For full details of trial design and protocol see Watkins et al[31] and Supplement 1. Ethics approval was provided by the NHS Research Ethics Board. The trial was conducted between 7th July 2015 and 29th December 2017.

Recruitment and eligibility criteria

Participants were recruited via an NHS treatment service or by direct internet self-referral. Eligibility criteria were aged ≥18, registered with a general practice, meeting criteria for depression operationalized by Patient Health Questionnaire-9 (PHQ-9) scores ≥ 10[45]; not currently in psychotherapy, and if receiving antidepressant medication, the dose stable for ≥ previous four weeks (See eAppendix 1 in Supplement 2 for further details). All eligible participants were invited to a telephone-delivered screening and baseline interview including the Structured Clinical Interview for Diagnosis sections on current and past depressive episodes conducted by trained research assistants, and to complete self-report questionnaires via online survey, email, or post. Similar follow-up assessments took place at 3 months (post-treatment) and 6-months post-randomization.

Randomization

All eligible participants who provided written informed consent were given access to an introductory internet-CBT module, which included a mood diary and basic psychoeducation about depression. Only participants who completed this module were randomized to ensure participants tried out and were engaged with internet-CBT before committing. Participants were randomized with equal probability to all 32 conditions by a permuted block randomization program delivered independently by the Peninsula Clinical Trials Unit. Randomization was stratified by severity of current depression (moderate PHQ-9≥10< 20 vs. severe PHQ-9≥20), antidepressant use (receiving versus not receiving BNF-recommended

therapeutic dose) and by referral source (NHS vs internet self-referral). All outcome assessors and data analysts were blind to treatment allocation. Statistical analyses were carried out blinded for randomization.

Intervention

The intervention was hosted on an established, secure internet-treatment platform. Participants received between 1-7 specific components (average 3-4), organised into discrete modules, depending on the randomized condition. Therapists provided brief written online asynchronous feedback for each module to improve retention and adherence[46]. To ensure treatment integrity and fidelity, each patient could access only the specific modules allocated, constraining patient and therapist to the relevant treatment protocol. We counterbalanced the sequential order of the modules across all conditions to ensure that each component occurred equally early or late in therapy across participants.

The 7 component modules reflected core elements identified within CBT[47] for depression and CBT innovations[37-40], with each hypothesized to specifically target distinct mechanisms. Within a behavioral model of CBT, *activity scheduling* (*AS*) and *absorption training* (*AT*) were hypothesized to increase response-contingent positive reinforcement[35] by respectively increasing frequency of and direct contact with positive reinforcers[48], and *functional analysis* (*FA*) was proposed to target habitual avoidance and rumination by identifying antecedent cues, controlling exposure to these cues, and practising alternative responses to them[49]. Within a cognitive model, *thought challenging* (*TC*) and *concreteness training* (CT)[38] were hypothesized to respectively reduce the negative thinking and overgeneralization cognitive bias, both characteristic of depression[47,50]. Within an emotional regulation model, *relaxation (R)* was hypothesized to target physiological arousal and tension, whilst *self-compassion training* (*SC*) was hypothesized to activate the soothing and safeness emotional system and, thereby, reduce negative mood[51-54]. For further details see eAppendix 2.

Outcomes

Primary outcome

Primary outcome was depression symptoms (PHQ-9[45]) measured at 3 months post-randomisation and 6 months post-treatment follow-up.

Secondary outcomes

Secondary trial outcomes were anxiety severity (Generalized Anxiety Disorder-7; GAD-7,[55] and social, home and work functioning (Work and Social Adjustment Scale; WSAS)[56].

Other measures

Adherence for each component was defined a priori as completing the relevant online module. Questionnaires to examine potential mediators are described in eAppendix 3. Ethnicity was self-classified through drop-down menu and open text.

Sample size

We assumed the smallest Meaningful Clinical Important Difference would be a small effect size (Cohen's d or SMD = .2[43]) for the main effect of a component or interaction between components on pre-to-post change in depression. To detect d=0.20 with 80% power at α=0.10 (recommended to decrease the relative risk of Type II to Type I error and of prematurely ruling out potentially active components[11,12,18,]) required a sample size of N=632. Estimating 40% dropout attrition post-treatment, we required N≈1056 for ANOVA: with multiple measures on the primary outcome, a growth curve model required 30-50% fewer participants for the same power[57], giving a conservative estimate of N≈736.

Statistical methods

Participants were analyzed according to their randomization group, including all participants randomized regardless of intervention received or study withdrawal (Intention-To-Treat, ITT). Statistical reporting followed CONSORT standards[58]. The seven factors were effect coded (-1;+1) and modelled together to study the main effects and interactions independently.[12] Relevant analyses were adjusted for the stratification variables. Main effects and interactions were estimated based on aggregates across experimental conditions (see Table 1). For primary and secondary outcomes, we used (a) baseline-score adjusted analysis of covariance (ANCOVA), modelling outcomes at post-treatment (3 months post-randomisation) and 6-month post-treatment as dependent variables; (b) analyses of Maximum Likelihood (ML)-based mixed-effect growth curve models including 6 measurements (Baseline, end of introductory module, end of module 1, end of module 2, 3-month post-randomisation, 6-month post-treatment). Cohens' *d* effect size was calculated for the adjusted estimates using the samples (N) available for the regression models. Pooled standard deviation (SD) was calculated from the change scores of baseline and follow-up measures using the same samples.[59] Complier Average Causal Effect (CACE) analyses[60-61] were carried out using instrumental variable method implemented via structural equation modelling to estimate intervention effects accounting for good adherence with the interventions, whilst retaining the benefits of randomization. Results are presented with 90% CI as we powered at α=0.10. Analyses were conducted with statistical analytical software Stata (version-17)[62].

Results

Description of participants

Between July 7th 2015 and 29th March 2017, 6940 individuals visited the online screener, 1289 fulfilled the inclusion criteria as did 275 directly referred from clinical service, 987 completed the subsequent telephone screening, and 924 were eligible for the trial. After exclusion of individuals not willing/able to participate, 767 adults were randomized (see Figure 1). For each component, half of participants were

allocated (n=382-407) and half were not allocated to it (n=360-384). eAppendices 4 and 5 show baseline characteristics of participants by factor and recruitment source.

Follow-up attrition

The number of participants not completing follow-up assessment at 3 and 6 months was 286 (37%) and 261 (34%) respectively. Missing rates did not differ between the primary and secondary outcomes nor intervention conditions: participants with missing data ranged from 298-314 at 3-months and 267-273 at 6-month follow-up, respectively.

Adherence and fidelity to interventions

Between 42-49% of those allocated to a treatment module completed it (see eAppendix 6). Fidelity to each component was 100%.

Primary outcome – Depressive symptoms

On average, participants receiving internet-CBT had reduced depression (post-treatment: $\Delta PHQ9 =$ -7.79, P <.001, recovery rate (PHQ9-9<10) 61.7%; 6-month follow-up: $\Delta PHQ9 =$-8.63; P =.001, recovery rate 71.5%). ANCOVA and ML-based growth curve models gave comparable effect estimates: we report the ANCOVA results (see eAppendix 7 for ML-based growth curve models). Only absorption training had a significant main effect on depressive symptoms, i.e., its presence reduced depression more than its absence (post-treatment: $\Delta PHQ9 =$ 0.21 [90%CI, -0.27-0.68]; 6-month follow-up: $\Delta PHQ9 =$ -0.54, [90%CI, -0.97 to -0.11, P= 0.04]; Cohen's d = 0.18). No other components had a significant main effect on depression at either post-treatment or 6-month follow-up (see Tables 2 and 3). Of a priori 2 x 2 interactions tested, only 2 were significant: at 6-month follow-up, *SC* by *AS* had a positive synergistic

interaction ($\Delta PHQ9 =$ -0.83 [90%CI, -1.28 to -0.38], P<0.001) but *SC* by *CT* had an antagonistic interaction ($\Delta PHQ9 =$ 0.45 [95%CI, 0.01 to 0.90], P=0.09).

Secondary outcomes – anxiety, work and social functioning.

Results for secondary outcomes were similar to depression (see Table 4; see eAppendix 8): there was no main effect of any component for anxiety and only *AT* had a significant effect on work and social functioning (post-treatment: $\Delta WSAS =$ -0.26 [90%CI, -0.98-0.45]; 6-month follow-up: $\Delta WSAS =$ -0.61 [90%CI, -1.23 to -0.01], P<0.10).

Effect modification and sensitivity CACE analyses

eAppendix 9 shows the results of effect modification analyses by referral source, depression severity and prescription of antidepressant medication. The results of the CACE analyses were consistent with the ITT analyses (see eAppendix 10).

Adverse events and concealment

Five participants were hospitalized for suicide attempt or serious self-harm, no people died during the 6-month follow-up. These events were judged as unrelated to interventions.

Discussion

This optimization trial in 767 adults with depression symptoms found no significant main effect of six components within internet-CBT on depression, anxiety, or work and social functioning post-treatment and at 6-month follow-up. There was a significant but small main effect of absorption training on reducing depression.

To our knowledge, this is the first randomized factorial trial to test the direct effects of components within internet-CBT in a large-scale sample of adults with depression. Internet-CBT had equivalent efficacy with the wider evidence-base for CBT for depression, exceeding the average recovery rates of 51% observed in IAPT services, and matching effects reported in recent meta-analyses[63], indicating that it was likely efficacious. There was no significant main effect of six from seven treatment components on symptoms. These preliminary results (see limitations) suggest that the active ingredients of internet-CBT are more likely to be common treatment factors than specific CBT strategies, consistent with recent arguments[42].

The observed reduction in depressive symptoms is likely due to some combination of spontaneous remission, regression to the mean, the specific components, and the constant component present in all conditions. The constant component includes the introductory module, asynchronous written support from a therapist, monitoring of symptoms, and factors common across all the CBT components, including cognitive-behavioral psychoeducation, learning and practicing a new behavioral or cognitive coping strategy, planning and review of homework. The constant component may engage both pan-therapy common factors (e.g., therapeutic alliance, hope, healthy actions) and factors only common to all CBT components, although we cannot distinguish their effects on outcome.

The main effect of *AT* at 6-month follow-up provides the first direct causal evidence for a specific treatment component influencing depression outcome, albeit a small, preliminary finding that needs replication. Relative to the average symptom reduction from baseline to 6-month follow-up, absorption contributed 6.4% (0.56/8.79). Absorption focused on increasing contact with positive reinforcers by using principles from flow theory[48] and encouraging patients to become immersed in their activities through changing their mindset, environment, attentional focus, and selection of activities. These positive effects are consistent with emergent evidence for the specific benefits of behavioral activation

and exposure to reward[64,65]. If a robust effect, enhancing absorption within CBT may increase treatment efficacy.

Study strengths are randomized factorial design, inclusion of participants recruited from both health services and the community, large sample size, high levels of treatment fidelity, 6-month follow-up, measurement of multiple outcomes, and adherence monitoring.

Limitations

First, as defined by our research question, we only examined components within CBT and thus cannot disentangle pan-therapy common factors from factors common to all CBT components. Second, the constant component may have been too strong, making it hard to detect additional effects of specific factors. Third, the dose of each component may have been insufficient since module completion was under 50% and their sequencing meant that patients only practiced each component for a few weeks. Nonetheless, we had comparable treatment effects to other CBT interventions and rates of treatment completion paralleled those typically found for internet-CBT[66]. Fourth, we don't know how generalizable internet-CBT is to face-to-face therapy. Fifth, because of the fractional design, the significant main effect of *AT* could be attributable to one or more of the interactions with which this main effect was aliased. Sixth, beyond the effects averaged across all patients, there may be individual differences in response to each component: individuals may only respond to certain components; a component may be positive for some but inactive or iatrogenic for others, producing an overall null effect. Developing reliable treatment rules to predict who responds optimally to which component may enable personalization within internet-CBT to improve outcomes[67,68].

Conclusions

Among adults with depression, of seven specific components, only absorption may have uniquely contributed to reduced depression at 6-month follow-up. Given this trial's novelty, limitations, and non-

significant findings, we cautiously suggest that internet-CBT may reduce depression through an as-yet-undetermined combination of spontaneous remission, pan-therapy common factors, and factors common to all CBT components, although further replication is needed (but see[69] for similar findings). These findings highlight the potential value of factorial designs in unpacking how therapies work.

Acknowledgements

Dr. Watkins and Mohammod Mostazir had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Conflicts of interest

No conflicts of interest for Mostazir, Newbold, Collins, Tester-Jones. Watkins has received royalties and fees from publishing a CBT training manual and providing therapy workshops.

Roles: Watkins was the lead author on the manuscript and Watkins and Mostazir analyzed the data. All authors interpreted the data. Watkins obtained funding for the project and designed the trial. Newbold coordinated the trial. Collins contributed to the design of the factorial trial. Watkins led the development and training of digital intervention. Newbold and Tester-Jones coordinated the recruitment, interventions and follow-ups. All authors contributed to the writing of the manuscript and approved the final version.

Role of funder

Funding for this trial was provided by grants from the Cornwall Partnership NHS Foundation Trust and South West Academic Health Sciences Network to EW. The sponsor and funder were not involved in the design and conduct of the study; collection, management, analysis, and interpretation of the data;

preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Additional Contributions

We thank Jenny Cadman, BA and Dr Carly Morriss (formerly Thornton) now at Affiliation Adult Eating Disorder Service, Cornwall (formerly College of Life and Environmental Sciences, University of Exeter) for their role as trial therapists and all the BeMe staff for their support in the delivery of the trial.

References

1. Kessler RC, Wang PS. The descriptive epidemiology of commonly occurring mental disorders in the United States. *Annu Rev Public Health*. 2008;29:115-129. doi:10.1146/annurev.publhealth.29.020907.090847.
2. Vos T, Flaxman AD, Naghavi M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: A systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380(9859):2163-2196. doi:10.1016/S0140-6736(12)61729-2.
3. Hollon SD, Munoz RF, Barlow DH, et al. Psychosocial intervention development for the prevention and treatment of depression: Promoting innovation and increasing access. *Biol Psychiatry*. 2002;52:610–30.
4. Barlow DH. Psychological treatments. *Am Psychol*. 2004;59(9):869–78.
5. Kazdin AE. Mediators and mechanisms of change in psychotherapy research. *Annu Rev Clin. Psychol*. 2007;3:1–27.
6. Holmes EA, Craske MG, Graybiel AM. Psychological treatments: a call for mental health science. *Nature.* 2014;511:287–9.
7. Holmes EA, Ghaderi A, Harmer CJ, et al. The Lancet Psychiatry Commission on psychological

treatments research in tomorrow's science. *Lancet Psychiatry.* 2018; 5(3):237–86. doi: 10.1016/S2215-0366(17)30513-8

8. Institute of Medicine. *Psychosocial interventions for mental and substance use disorders: A framework for establishing evidence-based standards*. Washington: The National Academies Press; 2015.

9. Medical Research Council (2008). Complex Interventions Guidance https://www.mrc.ac.uk/documents/pdf/complex-interventions-guidance/. Accessed 11 Dec 2012

10. Watkins ER and Newbold A. Factorial Designs Help to Understand How Psychological Therapy Works. *Front Psychiatry. 2020;*11:429. doi: 10.3389/fpsyt.2020.00429

11. Collins LM. *Optimization of behavioral, biobehavioral, and biomedical interventions: The multiphase optimization strategy (MOST)*. New York: Springer; 2018.

12. Collins LM, Kugler, KC. O*ptimization of behavioral, biobehavioral, and biomedical interventions: Advanced topics.* New York: Springer; 2018.

13. Collins LM, Dziak JD, Kugler KC, Trail JB. Factorial experiments: Efficient tools for evaluation of intervention components. *Am J Prev Med*. 2014;47:498–504. doi: 10.1016/j.amepre.2014.06.021

14. Collins LM, Dziak JD, Li R. Design of experiments with multiple independent variables: A resource management perspective on complete and reduced factorial designs. *Psychol Method*s. 2009;14:202–24. doi:10.1037/a0015826

15. Baker TB, Smith SS, Bolt DM, et al. Implementing Clinical Research Using Factorial Designs: A Primer. *Behav Ther.* 2017;48(4):567–80. doi: 10.1016/j.beth.2016.12.005

16. Dziak JJ, Nahum-Shani I, Collins LM. Multilevel Factorial Experiments for Developing Behavioral Interventions: Power, Sample Size, and Resource Considerations. *Psychol Methods.* 2012; 17(2):153–75. doi: 10.1037/a0026972

17. Collins LM, Kugler KC, Gwadz MV. Optimization of multicomponent behavioral and biobehavioral

interventions for the prevention and treatment of HIV/AIDS. *AIDS Behav*. 2015;20:197–214.

18. Collins LM, Murphy SA, Nair VN, Strecher VJ. A strategy for optimizing and evaluating behavioral interventions. *Ann Behav Med.* 2005;30:65–73. doi:10.1207/s15324796abm3001_8

19. Collins LM, Nahum-Shani I, Almirall D. Optimization of behavioral dynamic treatment regimens based on the sequential, multiple assignment, randomized trial (SMART). *Clin Trials*. 2014;11:426–34.

20. McClure JB, Derry H, Riggs KR, et al. Questions about quitting (Q2): Design and methods of a Multiphase Optimization Strategy (MOST) randomized screening experiment for an online, motivational smoking cessation intervention. *Contemp Clin Trials*. 2012;33:1094–102.

21. Strecher VJ, McClure JB, Alexander G, et al. Web-based smoking cessation programs: results of a randomized trial. *Am J Prev Med*. 2008;34:373–81.

22. Schlam TR, Fiore MC, Smith SS, et al. Comparative effectiveness of intervention components for producing long-term abstinence from smoking: A factorial screening experiment. *Addiction*. 2016;111:142–55.

23. Piper ME, Fiore MC, Smith SS, et al. Identifying effective intervention components for smoking cessation: A factorial screening experiment. *Addiction*. 2016;111:129–41.

24. Bernstein SL, Carter PM, Meurer W, et al. Advances in clinical trials methodology: Intervention optimization approaches in emergency medicine. *Am J Emer Med. 2022;53*, 6-11.

25. Spring B, Pfamatter AF, Marchese SH, et al. A factorial experiment to optimize remotely delivered behavioral treatment for obesity: Results of the Opt-In study. *Obesity. 2020; 28*, 1652-1662.

26. Wyrick DL, Tanner AE, Milroy J, et al. itMatters: Optimization of an online intervention to prevent sexually transmitted infections in college students. *J Am Coll Health*. 2020; *70*, 1212-1222.

27. Wu CJ, Hamada MS. *Experiments: planning, analysis, and optimization*. 3rd ed. New York: John Wiley & Sons; 2011.

28. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ*. 2008;337:a1655. doi:10.1136/bmj.a1655.

29. Kelders SM, Bohlmeijer ET, Pots WTM, van Gemert-Pijnen JEWC. Comparing human and automated support for depression: Fractional factorial randomized controlled trial. *Behav Res Ther.* 2015; 72: 72-80.

30. Mohr DC, Schueller SM, Tomasino KN, et al. Comparison of the effects of coaching and receipt of app recommendations on depression, anxiety, and engagement in the intellicare platform: Factorial randomized controlled trial. *J Med Internet Research* 2019; 21(8): e13609.

31. Watkins E, Newbold A, Tester-Jones M, et al. Implementing multifactorial psychotherapy research in online virtual environments (IMPROVE-2): study protocol for a phase III trial of the MOST randomized component selection method for internet cognitive-behavioural therapy for depression. *BMC Psychiatry.* 2016; 16:13. doi: 10.1186/s12888-016-1054-8

32. Uwatoko T, Luo Y, Sakata M, et al. Healthy Campus Trial: a multiphase optimization strategy (MOST) fully factorial trial to optimize the smartphone cognitive behavioral therapy (CBT) app for mental health promotion among university students: study protocol for a randomized controlled trial. *Trials.* 2018; 19(1):353. doi: 10.1186/s13063-018-2719-z

33. Roth AD, Pilling S. Using an Evidence-Based Methodology to Identify the Competences Required to Deliver Effective Cognitive and Behavioural Therapy for Depression and Anxiety Disorders. *Behav Cogn Psychother*. 2008; 36(2):129–47. doi: 10.1017/S1352465808004141

34. Dimidjian S, Barrera M, Martell C, Munoz RF, Lewinsohn PM. The Origins and Current Status of Behavioral Activation Treatments for Depression. *Annu Rev Clin Psychol*. 2011;7: 1–38.

35. Ferster CB. Functional Analysis Of Depression. *Am Psychol.* 1973; 28 (10):857–70. doi: 10.1037/h0035605

36. Jacobson NS, Martell CR, Dimidjian S. Behavioral activation treatment for depression: Returning to contextual roots. *Clin Psychol.* 2001; 8 (3):255–70. doi: 10.1093/clipsy.8.3.255

37. Watkins ER, Mullan E, Wingrove J, et al. Rumination-focused cognitive-behavioural therapy for residual depression: phase II randomised controlled trial. *Br J Psychiatry*. 2011;199:317–22. doi: 10.1192/bjp.bp.110.090282

38. Watkins ER, Taylor RS, Byng R, et al. Guided self-help concreteness training as an intervention for major depression in primary care: a Phase II randomized controlled trial. *Psychol Med*. 2012;42:1359–71. doi: 10.1017/S0033291711002480

39. Watkins. Constructive and unconstructive repetitive thought. *Psychol Bull*. 2008; 134(2):163–206. doi: 10.1037/0033-2909.134.2.163

40. Watkins E, Moberly NJ, Moulds ML. Processing mode causally influences emotional reactivity: distinct effects of abstract versus concrete construal on emotional response. *Emotion.* 2008; 8(3):364–78. doi: 10.1037/1528-3542.8.3.364

41. Mulder R, Murray G, Rucklidge J. Common versus specific factors in psychotherapy: opening the black box. *Lancet Psychiatry.* 2017; 4(12):953–62. doi: 10.1016/S2215-0366(17)30100-1

42. Wampold BE. How important are the common factors in psychotherapy? An update. W*orld Psychiatry.* 2015; 14(3):270–7. doi: 10.1002/wps.20238

43. Cuijpers P, Reijners M, Huibers MJH. The role of common factors in psychotherapy outcomes. A*nnu Rev Clin Psychol.* 2018; 15(5):20. doi:10.1146/annurev-clinpsy-050718-095424

44. Baskin TW, Tierney SC, Minami T, Wampold BE. Establishing specificity in psychotherapy: A meta-analysis of structural equivalence of placebo controls. *J Cons Clin Psychol*. 2003;71:973–9. doi: 10.1037/0022-006X.71.6.973

45. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a brief depression severity measure. *J Gen Intern Med.* 2001;16:606–13. doi:10.1046/j.1525-1497.2001.016009606.x

46. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. *PLoS One*. 2010;5:e13196

47. Beck ATR AJ, Shaw BF, Emery G. *Cognitive Therapy of Depression*. New York: Guilford Press; 1979.

48. Csikszentmihalyi M. *Flow: The classic work on how to achieve happiness*. London: Rider; 2002.

49. Watkins ER, Nolen-Hoeksema S. A habit-goal framework of depressive rumination. *J Abnorm Psychol.* 2014;123:24–34. doi:10.1037/a0035540.

50. Carver CS. Generalization, adverse events, and development of depressive symptoms. *J Pers.* 1998; 66(4):607–19. doi: 10.1111/1467-6494.00026

51. Gilbert P. The origins and nature of compassion focused therapy. *Br J Clin Psychol*. 2014; 53(1):6–41. doi: 10.1111/bjc.12043

52. Matos M, Duarte C, Duarte J, et al. Psychological and Physiological Effects of Compassionate Mind Training: a Pilot Randomised Controlled Study. *Mindfulness*. 2017; 8(6):1699–712. doi:10.1007/s12671-017-0745-7

53. Gilbert P, Procter S. Compassionate mind training for people with high shame and self-criticism: Overview and pilot study of a group therapy approach. *Clin Psychol Psychother*. 2006; 13(6):353–79. doi: 10.1002/cpp.507

54. Leary MR, Tate EB, Adams CE, Allen AB, Hancock J. Self-compassion and reactions to unpleasant self-relevant events: the implications of treating oneself kindly. *J Pers Soc Psychol*. 2007; 92(5):887–904. doi: 10.1037/0022-3514.92.5.887

55. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: The GAD-7. *Arch Intern Med*. 2006;166(10):1092-1097. doi:10.1001/archinte.166.10.1092.

56. Mundt JC, Marks IM, Shear MK, Greist JM. The Work and Social Adjustment Scale: A simple

measure of Impairment in functioning. *Br J Psychiatry*. 2002;180:461–4. doi:10.1192/bjp.180.5.461.

57. Muthen BO, Curran PJ. General longitudinal modeling of individual differences in experimental designs: a latent variable framework for analysis and power estimation. *Psychol Methods.* 1997;2:371–402.

58. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC Med*. 2010;8:18.

59. Morris SB. Estimating effect sizes from pretest-posttest-control group designs. *Organ Res Methods*. 2008;11(2):364-386. doi:10.1177/1094428106291059.

60. Angrist JD, Imbens GW, Rubin DB. Identification of Causal Effects Using Instrumental Variables. *J Am Stat Assoc*. 1996;91(434):444-455. https://dash.harvard.edu/handle/1/3382969. Accessed March 8, 2016.

61. Emsley R, Dunn G, White IR. Mediation and moderation of treatment effects in randomised controlled trials of complex interventions. *Stat Methods Med Res*. 2010. doi:10.1177/0962280209105014.

62. Stata: StataCorp. 2021. Stata Statistical Software: Release 17. College Station, TX: StataCorp LLC

63. 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 Psychiatry.* 2021;78(4):361–371. doi:10.1001/jamapsychiatry.2020.4364

64. Alexopoulos GS, Raue PJ, Banerjee S, et al. Comparing the streamlined psychotherapy "Engage" with problem-solving therapy in late-life major depression. A randomized clinical trial. *Mol Psychiatry*. 2021;26(9):5180-5189. doi:10.1038/s41380-020-0832-3

65. Furukawa TA, Suganuma A, Ostinelli EG, et al. Dismantling, optimising, and personalising internet cognitive behavioural therapy for depression: a systematic review and component network meta-

analysis using individual participant data. *Lancet Psychiatry*. 2021;8(6):500-511. doi:10.1016/S2215-0366(21)00077-8

66. van Ballegooijen W, Cuijpers P, van Straten A, et al. Adherence to Internet-Based and Face-to-Face Cognitive Behavioural Therapy for Depression: A Meta-Analysis. *PLoS ON*E. 2014; 9(7): e100674. https://doi.org/10.1371/journal.pone.0100674

67. Cohen ZD, DeRubeis R. Treatment selection in depression. *Annu Rev Clin Psychol.* 2018; 14:209-236. doi.org/10.1146/annurev-clinpsy-050817-084746

68. Kessler RC. The potential of predictive analytics to provide clinical decision support in depression treatment planning. *Curr Opin Psychiatry.* 2018; 31: 32–39. doi:10.1097/yco.0000000000000377

69. Sakata M, Toyomoto R, Yoshida K, et al. Components of smartphone cognitive-behavioural therapy for subthreshold depression among 1093 university students: a factorial trial. *Evid Based Ment Health*. 2022;25(e1):e18-e25. doi:10.1136/ebmental-2022-300455

Table 1. Experimental groups of the IMPROVE-2 fractional factorial design

Condition	Functional Analysis	Concrete training	Compassion	Absorption	Relaxation	Activity Scheduling	Thought Challenging
1	no	no	no	no	no	yes	yes
2	yes	no	no	no	no	no	no
3	no	no	yes	no	no	no	no
4	yes	no	yes	no	no	yes	yes
5	no	no	no	yes	no	yes	no
6	yes	no	no	yes	no	no	yes
7	no	no	yes	yes	no	no	yes
8	yes	no	yes	yes	no	yes	no
9	no	yes	no	no	no	no	no
10	yes	yes	no	no	no	yes	yes
11	no	yes	yes	no	no	yes	yes
12	yes	yes	yes	no	no	no	no
13	no	yes	no	yes	no	no	yes
14	yes	yes	no	yes	no	yes	no
15	no	yes	yes	yes	no	yes	no
16	yes	yes	yes	yes	no	no	yes
17	no	no	no	no	yes	no	yes
18	yes	no	no	no	yes	yes	no
19	no	no	yes	no	yes	yes	no
20	yes	no	yes	no	yes	no	yes
21	no	no	no	yes	yes	no	no
22	yes	no	no	yes	yes	yes	yes
23	no	no	yes	yes	yes	yes	yes
24	yes	no	yes	yes	yes	no	no
25	no	yes	no	no	yes	yes	no
26	yes	yes	no	no	yes	no	yes
27	no	yes	yes	no	yes	no	yes
28	yes	yes	yes	no	yes	yes	no
29	no	yes	no	yes	yes	yes	yes
30	yes	yes	no	yes	yes	no	no
31	no	yes	yes	yes	yes	no	no
32	yes	yes	yes	yes	yes	yes	yes

Note. Every factor occurs an equal number of times at high and low levels (i.e. balanced) and all factors are orthogonal to each other. Each effect estimate involves all 32 of the conditions in Table 1, thereby maintaining the power associated with all participants. This Resolution IV design means that all main effects are aliased with 3-way and higher interactions, and all 2-way interactions are aliased with 2-way and higher interactions, on assumption that non-negligible 3-way and higher interactions are unlikely. In contrast, a standard RCT is aliased for **all** main effects and interactions of treatment component

Table 2: Baseline score adjusted ANCOVA model for change in PHQ-9 at post-treatment (3 months post-randomisation) and 6-month follow-up

Variable	ΔPHQ-9 baseline to post-treatment (90% CI)	P	Cohen's d	Δ PHQ-9 baseline to 6 month follow-up (90% CI)	P	Cohen's d
Baseline PHQ-9	0.49 (0.38 to 0.61)***	0.00	–	0.42 (0.32 to 0.53)***	0.00	–
Recruit source	0.49 (-0.12 to 1.11)	0.19	0.16	0.54 (-0.02 to 1.10)	0.11	0.18
Use of antidepressant	0.31 (-0.16 to 0.78)	0.27	0.10	0.68 (0.25 to 1.11)**	0.01	0.22
Absorption training (AT)	0.21 (-0.27 to 0.68)	0.47	0.07	-0.54 (-0.97 to -0.11)*	0.04	-0.18
Activity scheduling (AS)	-0.08 (-0.56 to 0.40)	0.79	-0.03	0.14 (-0.29 to 0.58)	0.59	0.05
Concreteness training (CT)	-0.04 (-0.51 to 0.43)	0.88	-0.01	-0.15 (-0.58 to 0.28)	0.57	-0.05
Functional analysis (FA)	-0.09 (-0.56 to 0.39)	0.76	-0.03	0.18 (-0.25 to 0.61)	0.49	0.06
Relaxation (R)	0.15 (-0.33 to 0.63)	0.62	0.05	-0.18 (-0.61 to 0.25)	0.50	-0.06
Self-compassion training (SC)	0.05 (-0.44 to 0.53)	0.87	0.02	0.05 (-0.39 to 0.50)	0.84	0.02
Thought challenging (TC)	0.09 (-0.39 to 0.56)	0.76	0.03	-0.05 (-0.48 to 0.38)	0.85	-0.02
FA X AT	0.09 (-0.39 to 0.57)	0.75	–	-0.21 (-0.64 to 0.22)	0.42	–
FA X AS	-0.09 (-0.57 to 0.40)	0.76	–	0.06 (-0.38 to 0.49)	0.83	–
FA X CT	0.18 (-0.29 to 0.65)	0.52	–	-0.22 (-0.65 to 0.21)	0.39	–
FA X R	-0.28 (-0.77 to 0.20)	0.34	–	-0.16 (-0.59 to 0.28)	0.55	–
FA X SC	-0.06 (-0.55 to 0.43)	0.85	–	0.14 (-0.30 to 0.58)	0.59	–
FA X TC	-0.01 (-0.49 to 0.46)	0.96	–	-0.14 (-0.57 to 0.29)	0.60	–
SC X AT	-0.17 (-0.66 to 0.33)	0.57	–	0.15 (-0.30 to 0.60)	0.58	–
SC X AS	-0.36 (-0.85 to 0.14)	0.24	–	-0.83 (-1.28 to -0.38)***	0.00	–
SC X CT	0.10 (-0.39 to 0.59)	0.73	–	0.45 (0.01 to 0.90)+	0.09	–
SC X R	0.08 (-0.42 to 0.57)	0.80	–	0.13 (-0.32 to 0.58)	0.64	–
SC X TC	-0.43 (-0.92 to 0.06)	0.15	–	-0.05 (-0.50 to 0.39)	0.85	–
AT X CT	-0.15 (-0.63 to 0.32)	0.60	–	0.25 (-0.18 to 0.68)	0.33	–
CT X TC	-0.21 (-0.69 to 0.26)	0.46	–	0.06 (-0.37 to 0.49)	0.81	–
CT X R	-0.42 (-0.90 to 0.06)	0.15	–	-0.41 (-0.84 to 0.02)	0.12	–
CT X AS	-0.50 (-0.98 to -0.02)+	0.09	–	0.31 (-0.12 to 0.74)	0.24	–

+p<0.10, *p<0.05, **p<0.01, ***p<0.001. Models are adjusted for baseline score, source of referral and use of antidepressant medication.

CI = confidence interval.

d = Cohen's d calculated only for main effects i.e., the between-condition difference from the regression model was divided by the pooled standard deviation (SD), where pooled SD was calculated from the change score from baseline to relevant follow-up measure for the samples used in the regression.

The two-way interactions examined are those pre-specified in factorial design of particular interest where the 2-way interaction is only aliased /confounded with 3-way and higher interactions, which are assumed to be negligible in size.

Table 3: Primary and secondary outcomes by absence versus presence of treatment components at baseline, post-treatment, and 6-month post-treatment follow-up.

Outcomes	Absorption Training		Activity Scheduling		Concreteness Training		Functional Analysis		Relaxation		Self-Compassion		Thought Challenging	
	Yes (+1)	No (-1)	Yes (+1)	No (-1)	Yes (+1)	No (-1)	Yes (+1)	No (-1)	Yes (+1)	No (-1)	Yes (+1)	No (-1)	Yes (+1)	No (-1)
PHQ-9, mean (SD), N														
Baseline	16.46 (4.07), 383	16.51 (4.04), 384	16.67 (3.97), 385	16.30 (4.13), 382	16.44 (4.17), 383	16.53 (3.94), 384	16.31 (4.15), 387	16.66 (3.95), 380	16.74 (4.07), 384	16.22 (4.03), 383	16.60 (4.01), 407	16.36 (4.10), 360	16.52 (4.08), 382	16.44 (4.04), 385
Post-treatment	8.68 (6.59), 228	8.51 (6.00), 253	8.64 (6.21), 244	8.54 (6.36), 237	8.58 (6.44), 240	8.60 (6.12), 241	8.47 (6.21), 244	8.71 (6.36), 237	8.85 (6.30), 241	8.33 (6.26), 240	8.68 (6.43), 258	8.48 (6.12), 223	8.75 (6.31), 229	8.44 (6.26), 252
6-month follow-up	7.19 (5.49), 254	8.20 (6.42), 252	7.92 (5.90), 246	7.48 (6.08), 260	7.68 (6.10), 257	7.71 (5.88), 249	7.81 (6.21), 255	7.58 (5.77), 251	7.52 (5.87), 250	7.86 (6.11), 256	7.91 (6.07), 274	7.44 (5.90), 232	7.81 (6.31), 248	7.58 (5.68), 258
GAD-7, mean (SD), N														
Baseline	13.15 (4.64), 383	13.46 (4.60), 384	13.50 (4.54), 385	13.10 (4.70), 382	12.95 (4.81), 383	13.66 (4.40), 384	13.41 (4.55), 387	13.20 (4.55), 383	13.41 (4.69), 384	13.20 (4.55), 383	13.41 (4.36), 407	13.18 (4.90), 360	13.32 (4.61), 382	13.29 (4.64), 385
Post-treatment	6.85 (5.41), 225	6.97 (5.17), 244	6.93 (5.23), 241	6.89 (5.35), 228	6.80 (5.48), 232	7.03 (5.10), 237	6.83 (5.39), 239	7.00 (5.18), 230	7.02 (5.39), 237	6.81 (5.18), 232	7.30 (5.59), 256	6.45 (4.87), 213	6.98 (5.35), 223	6.85 (5.23), 246
6-month follow-up	6.68 (5.30), 250	7.17 (5.48), 250	7.04 (5.38), 246	6.81 (5.42), 254	6.58 (5.40), 253	7.28 (5.37), 247	7.17 (5.57), 253	6.67 (5.20), 247	6.68 (5.29), 246	7.16 (5.49), 254	7.06 (5.37), 272	6.76 (5.43), 228	7.03 (5.50), 244	6.82 (5.30), 256
WSAS, mean (SD), N														
Baseline	21.65 (7.18), 382	21.89 (7.17), 383	22.23 (7.36), 385	21.31 (6.96), 380	21.66 (7.09), 382	21.88 (7.26), 383	21.70 (7.28), 386	21.84 (7.06), 379	21.96 (7.14), 382	21.58 (7.20), 383	21.84 (7.04), 407	21.69 (7.32), 358	21.94 (7.41), 381	21.60 (6.93), 384
Post-treatment	12.55 (9.68), 214	13.40 (9.03), 239	13.58 (9.61), 231	12.39 (9.04), 222	12.93 (9.44), 224	13.06 (9.26), 229	12.97 (9.32), 231	13.02 (9.39), 222	13.38 (9.43), 229	12.61 (9.25), 224	13.21 (9.95), 246	12.75 (8.58), 207	13.15 (8.97), 218	12.86 (9.69), 235
6-month follow-up	9.12 (8.11), 248	10.32 (8.69), 246	10.42 (8.76), 242	9.05 (8.04), 252	9.76 (8.43), 252	9.68 (8.42), 242	9.77 (8.47), 249	9.67 (8.38), 245	9.81 (8.40), 243	9.63 (8.45), 251	10.28 (8.95), 269	9.05 (7.70), 225	10.16 (8.66), 241	9.30 (8.18), 253

Yes (+1), presence of respective component (i.e., aggregate of 16 conditions that include the component). No (-1), absence of respective component (i.e., aggregate of 16 conditions that omit the component). PHQ-9, Patient Health Questionnaire-9; GAD-7, Generalised Anxiety Disorder scale-7; WSAS, Work and Social Adjustment Scale

Figure titles and legends

Figure 1. Consort Diagram showing flow of participants through IMPROVE-2 study

Figure 2. Change in depression over time for presence vs absence of each component within internet CBT (absorption, activity scheduling, concreteness training, functional analysis, relaxation, self-compassion, thought challenging). Errors bars represent standard error.