

BMJ Open Exploring the efficacy of psychological treatments for depression: a multiverse meta-analysis protocol

Constantin Yves Plessen ,^{1,2} Eirini Karyotaki ,^{1,3} Pim Cuijpers ¹

To cite: Plessen CY, Karyotaki E, Cuijpers P. Exploring the efficacy of psychological treatments for depression: a multiverse meta-analysis protocol. *BMJ Open* 2022;**12**:e050197. doi:10.1136/bmjopen-2021-050197

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-050197>).

Received 15 February 2021
Accepted 29 December 2021



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Department of Clinical, Neuro-, and Developmental Psychology, Vrije Universiteit Amsterdam, Amsterdam, Noord-Holland, Netherlands

²Department of Psychosomatic Medicine, Charité Universitätsmedizin Berlin, Berlin, Germany

³Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, Amsterdam, Noord-Holland, Netherlands

Correspondence to

Constantin Yves Plessen;
constantin-yves.plessen@charite.de

ABSTRACT

Introduction In the past four decades, over 700 randomised controlled trials (RCTs) and 80 meta-analyses have examined the efficacy of psychological treatments for depression. Overwhelming evidence suggests that all types of psychological treatments are effective. Yet, many aspects are still unexplored. Meta-analysts could perform hundreds of potential meta-analyses with the current literature, and a comprehensive bird's-eye view of all published studies is missing. This protocol outlines how a multiverse meta-analysis can evaluate the entire body of the literature on psychological treatments of depression in a single analysis. Thereby, gaps of evidence and areas of robustness are highlighted.

Methods and analysis We will conduct systematic literature searches in bibliographical databases (PubMed, Embase, PsycINFO and Cochrane Register of Controlled Trials) up until 1 January 2021. We will include all RCTs comparing a psychological treatment with a control condition. We will include studies published in English, German, Spanish or Dutch, and exclude trials on maintenance and relapse prevention as well as dissertations. Two independent researchers will check all records. All self-reported and clinician-rated instruments measuring depression are included. We will extract information on recruitment settings, target groups, age groups, comorbidity, intervention formats, psychotherapy types, number of sessions, control conditions and country. Two independent researchers will assess risk of bias using the Cochrane Risk of Bias assessment tool. As part of the multiverse meta-analysis, unweighted, fixed effect and random effects models will be calculated.

Ethics and dissemination As we will not collect any primary data, an ethical approval of this protocol is not required. We will publish the results in a peer-review journal and present them at international conferences. We will follow open science practices and provide our code and data.

INTRODUCTION

Depression is the most researched mental disorder worldwide as it is highly prevalent, costly and associated with many adverse outcomes such as reduced role functioning and quality of life as well as increased comorbidity and mortality.^{1–4} As of 10 July 2019, 81 meta-analyses have examined the efficacy of various psychological treatments on

Strengths and limitations of this study

- We will investigate the efficacy of psychological interventions for depression on a broad range of subgroups (ie, age groups, treatment formats, types of treatments, type of control group) with a multiverse meta-analysis—rather than having to conduct each of these meta-analyses in individual studies, this approach can investigate all of them in one single analysis.
- We will investigate how flexibility in data analysis might affect the emerging meta-analytical results and, in doing so, will be able to identify gaps in the literature and highlight areas of robust and reliable evidence.
- Thereby, this study can help to resolve conflicting meta-analyses and provide a bird's-eye perspective on the entire field of psychological depression research.
- Uncovering strong evidence for the efficacy of these psychological interventions for depression can help guide decision-making for the allocation of scarce healthcare resources and funding away from redundant research.
- Multiverse meta-analyses cannot end debates about which meta-analysis should be run or is the correct one but rather facilitate those debates by creating a space of all reasonable meta-analyses and visualising why these meta-analyses produced diverging results.

depression to identify which therapies are effective for which groups.⁵ Evidence from over 700 included randomised control trials suggests the efficacy of all common psychological treatments for depression,^{6–7} that is cognitive behavioral therapy,⁸ behavioural activation therapy,⁹ interpersonal psychotherapy,¹⁰ problem-solving therapy,¹¹ psychodynamic therapy,^{12,13} third-wave psychotherapies¹⁴ and non-directive counselling,¹⁵ with no significant differences in efficacy between these treatments.² These therapies are also effective in most target groups (older adults, college students, patients with general medical disorders), yet tend to be less effective in some—such as patients with comorbid



substance use disorders or chronic depression. Furthermore, these treatments are effective when delivered in individual, group and guided self-help format.

Due to the exponential growth of research on psychotherapy for adult depression, it is important to summarise, integrate and visualise the knowledge from these meta-analyses and primary studies on differences between therapies, target groups, treatment formats and control conditions. Although these meta-analyses exist, a bird's-eye view of the field is still missing. For this reason, we will apply the approach of a multiverse meta-analysis that allows to investigate all reasonable meta-analyses that could be conducted based on the available primary studies. Such a multiverse meta-analysis provides three important benefits over other research synthesis methods:

First, even though conventional meta-analyses also provide an overview of the published literature on a given research question, they do not consider different paths that could have been taken in selecting the data or the data could be analysed. Making sure that the conclusions of a meta-analysis are not disproportionately influenced by data analytical decisions, a multiverse meta-analysis can provide the entire picture and underpin the robustness of the findings—or lack thereof—by conducting all possible and reasonable meta-analyses at once.

Second, it allows us to (1) provide a research integration similar to umbrella reviews and (2) investigate the influence flexibility in data analysis could have on published meta-analyses.

Importantly, in contrast to umbrella reviews, which aim to narratively and visually synthesise multiple published meta-analyses, a multiverse analysis also includes not yet conducted meta-analyses. Conducting and including these possible, reasonable meta-analyses within the multiverse meta-analysis provides a complete picture that is less dependent on flexibility in study selection and available primary studies. For example, recent umbrella reviews evaluated the evidence for the effectiveness of physical activity on depression,^{16 17} the efficacy of psychosocial interventions for mental health disorders,¹⁸ the effectiveness of psychotherapy in general¹⁹ and biomarkers associated with mental health disorders.²⁰ A multiverse meta-analysis could also provide such integration of multiple meta-analyses but contain more information about diverging evidence by providing a robustness check of all potential meta-analyses.

Third, at times, multiple meta-analyses with overlapping research questions reach different conclusions due to differences in inclusion and exclusion criteria, data analytical decisions, differences in publication bias assessment and risk of bias assessment in general.^{21 22} It is therefore crucial to evaluate the influence such choices might have on the final result of each meta-analysis. Was the method, restriction of diagnostic criteria or other exclusion criteria decisive, or is the same result reached via multiple analytical strategies? A multiverse meta-analysis can provide the needed clarity to answer these questions by extending the idea of sensitivity analyses. All

Table 1 PICO's framework for explicit research question of this study

PICO's criteria	Determinants
Patient	Participants suffering from any depressive disorder, diagnosed either as major depression, subclinical depression, mood disorder, chronic depression or by a cut-off score.
Intervention	Any type of psychological intervention, such as cognitive behavioural therapy, problem-solving therapy, third-wave cognitive therapies, behavioural activation therapy, non-directive supportive therapy, psychodynamic therapy, interpersonal therapy, life review therapy or other types of therapy.
Comparison	Any type of control condition: care as usual, waiting list or other types of control conditions.
Outcome	All self-reported and clinician-rated instruments measuring depression.
Study design	Randomised controlled trials.
PICO's, Patient, intervention, comparison, outcome, study design.	

meta-analyses that can be considered as reasonable based on the included determinants can be calculated in a single analysis and the results can be visualised simultaneously.

To provide such a bird's-eye perspective on the entirety of depression research, we will conduct our multiverse meta-analysis using the MetaPsy database.^{7 23} The MetaPsy database is uniquely suited to help answer these outlined questions as it contains all randomised control trials evaluating the treatment efficacy for depression. Thus, the planned study will allow an examination of the robustness of all published randomised control trials on the efficacy of psychological treatments for depression by investigating all possible and reasonable meta-analyses in a single study—the multiverse of psychological treatments for depression. This will provide an exhaustive overview to guide future research as knowledge gaps are identified, and policy-makers can use this overview to inform evidence-based decision-making. Examining this multiverse of psychological depression research can help resolve conflicting meta-analyses and contested evidence, mitigate the associated adverse effects of these phenomena on research progress, and provide a bird's-eye perspective of the entire field.²⁴

Aim and objective

We aim to estimate the influence of different types of psychological treatments, control conditions, and participant groups have on the effect size estimates across all published randomised studies (see [table 1](#) for all determinants of the PICO framework).

METHODS AND ANALYSIS

Multiverse meta-analysis

A multiverse meta-analysis contains all reasonable meta-analyses on a research question that could be conducted—reasonable meaning that they have a

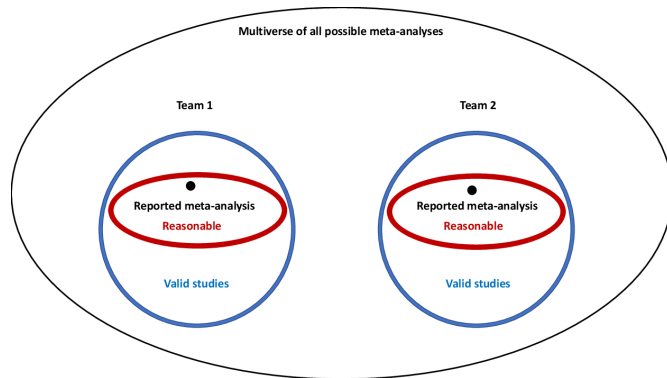


Figure 1 Sets of possible specifications as perceived by two research teams conducting meta-analyses.

theoretical foundation. Generally, the multiverse contains all combinations of reasonable specifications. For instance, different research teams would consider it to be reasonable to investigate psychological treatments for depression with different age groups, different types of therapies or control conditions. It would also be possible to investigate the efficacy of psychological treatments on depression for variables not guided by theory, for instance different first names, favourite French movies, or hair length, yet most experts would not consider these specifications to be valid or reasonable.

Figure 1 illustrates this difference between possible and reasonable specifications for multiverse meta-analyses. A large, possibly infinite, multiverse of meta-analyses could be conducted on a single research question (black circle). Different research teams consider only a subset of these to be valid (the blue circles) based on their inclusion and exclusion criteria. These questions can be investigated with different appropriate methods, resulting in a set of specifications that are considered reasonable (the red circles). Diverging meta-analyses (black dots) can emerge because different research teams draw different circles. Therefore, multiverse meta-analyses cannot end debates about what specifications are reasonable or which meta-analyses should be run but rather facilitate those debates by creating a space of all reasonable meta-analyses and visualising why these meta-analyses produced diverging results.²⁴ Even if two teams have non-overlapping sets of reasonable specifications, multiverse meta-analysis can help them understand why they may have reached different conclusions: Do these different conclusions arise due to differences in red circles (difference in which sets of specifications are deemed reasonable), or due to different black dots (differences in selectively reported meta-analyses).

To conduct such a multiverse meta-analysis, Voracek *et al*²⁴ suggest blending two approaches initially developed for the analysis of primary studies—the specification curve and the multiverse analysis approach.

Specification curve analysis is comprised of four steps: (1) identifying all reasonable specifications for analysis, that is, deciding which data to analyse and how (see figure 1), (2) statistically analysing all of them, (3)

visualisation of the emerging results and (4) inferential statistical procedures to test if the overall results deviate from the null hypothesis.²⁵ Steegen *et al*²⁶ proposed a similar procedure called ‘multiverse analysis’. Both multiverse analysis and specification curve analysis are almost identical in first and second steps, but they deviate in the proposed graphical displays for the third step, and multiverse analysis avoids the inferential statistics of the fourth step.

Based on these considerations, our resulting multiverse meta-analysis will consist of three steps: (1) creating a list of all reasonable specifications, (2) conducting inferential statistical tests (a parametric bootstrap procedure) and (3) visualising the multiverse meta-analysis (descriptive and inferential statistical specification curve plots).

Step 1: creating a list of all reasonable specifications

We will consider seven Which factors (which data to meta-analyse) and one How factor (how to meta-analyse the data), as follows.

Which factors: which data to analyse?

The decisions of which groups to compare in a meta-analysis are manifold; we decided to specify the following seven relevant study features:

- ▶ *Target group.* Researchers may decide to compare specific populations only. Included are the categories: (1) adults (adults in general with no specific demographic characteristic), (2) general medical group, (3) perinatal depression, (4) student population, (5) older adults, (6) other groups or (7) all target groups.
- ▶ *Format.* Researchers might compare interventions for different types of therapy delivery: (1) group, (2) individual, (3) guided self-help, (4) couples, (5) telephone, (6) other formats or (7) all formats.
- ▶ *Diagnosis.* We compare interventions that based their inclusion of participants on different types of diagnoses. These specifications include the following diagnoses: (1) major depression according to criteria by the Diagnostic and Statistical Manual of Mental Disorders (DSM), DSM-V, DSM-IV, DSM-III-R, DSM-III, Research Diagnostic Criteria (RDC) for major depression, or Feighner criteria for depressive disorder. We will also include diagnoses of (2) mood disorder or other diagnosed disorders (eg, dysthymia; depression Not Otherwise Specified; minor depression according to RDC). (3) Chronic depression, i.e., participants meet criteria for chronic or treatment-resistant depression, according to any definition given by the authors of the study. We also include diagnoses defined as (4) scoring above a clinically relevant cut-off score or (5) diagnosed subclinical depression, that is, participants score above a cut-off on a self-rating scale, but do not meet criteria for a depressive disorder according to a diagnostic interview (such as the Composite International Diagnostic Interview or Structured Clinical Interview for DSM).

- ▶ **Type.** Researchers might compare interventions for different types of therapies: (1) cognitive behavioural therapy, (2) problem-solving therapy, (3) third-wave cognitive therapies, (4) behavioural activation therapy, (5) non-directive supportive therapy, (6) psychodynamic therapy, (7) interpersonal therapy, (8) life review therapy, (9) other types of therapy or (10) all types of therapy.
- ▶ **Control conditions.** Researchers might compare interventions for different types of control conditions. These specifications include: (1) care as usual, (2) waiting list, (3) other types of control conditions or (4) all types of control conditions.
- ▶ **Country.** Researchers might compare studies conducted in different regions: (1) EU, (2) the USA, (3) the UK, (4) Canada, (5) East Asia, (6) Australia, (7) other countries or (8) all countries.
- ▶ **Risk of bias.** Researchers might compare studies with different degrees of risk of bias assessed with the risk of bias assessment tool.²⁷ Based on the assessment of allocation concealment, blinding of assessors, intention-to-treat analyses and sequence generation, we will create a dichotomous variable indicating whether the study is graded as (1) low risk of bias (if each aspect is rated as low risk) or (2) some concern (if any or all aspects are rated as high risk). We will also include a specification including (3) all studies, regardless of bias assessment.

How factors: how to analyse the data?

The How factor we analyse concerns the choice of the meta-analytic model—random effects model (REM), the fixed-effect model, three-level model or unweighted model (UWM). We will calculate standardised mean differences (Hedges g) based on post-treatment outcomes. We consider two REM variants, which differ in the way the between-study variance is estimated: (1) the DerSimonian-Laird estimator, which is the default estimator in the popular Comprehensive Meta-Analysis software²⁸ and the restricted maximum likelihood estimator (REML), which is the default in the popular metafor R package.²⁹ Multilevel models (3-LVL) are particularly well equipped to account for effect size dependency when multiple effect sizes are reported per study, and are therefore also included in our multiverse meta-analysis. Additionally, we follow the recommendation by Voracek *et al*²⁴ to calculate an unweighted meta-analytic model (UWM). Even though this is unusual for meta-analysis, we consider this approach necessary because of the similarities with the ‘cognitive algebra’ that is common in narrative, unsystematic reviews where empirical evidence is not weighted according to its information value (sample size). To include an effect size estimate accounting for the potential presence of publication bias, we include the p -uniform* estimate.³⁰ Together, the How factor makes up for five different ways to meta-analyse the same data. Theoretically, these Which and How factor combinations could produce up

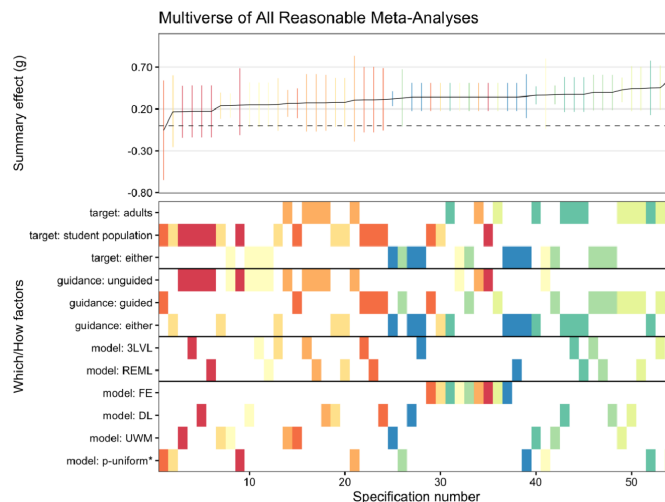


Figure 2 Descriptive meta-analytic specification plots with simulated data. DL, DerSimonian-Laird estimator; FE, fixed effect; REML, restricted maximum likelihood estimator; UWM, unweighted mode; 3-LVL, three-level model.

to $6 \times 8 \times 7 \times 6 \times 10 \times 4 \times 8 \times 6 \times 3 \times 5 = 58\,060\,800$ ways to meta-analyse different data subsets.

Step 2: specification curve analysis

We conduct an inferential statistical test with a parametric bootstrap approach to evaluate the descriptive meta-analytic specification curve plot against the null hypothesis of no psychological treatment effect on depression. We regard all study features for each sample from the literature as fixed but simulate random values as new effect sizes under the assumption that the null hypothesis is true. Means are randomly drawn from a normal distribution centred around zero, and the SD are set to the observed SD. Then, descriptive specification curve analysis is applied and repeated 1000 times.²⁴ The resulting 1000 bootstrapped specification curves are used to identify the lower and upper limits—the 2.5% and 97.5% quantiles. Exceeding these limits would indicate a deviation from the under-the-null scenario of no effect ($g=0$). We do not have any prior assumptions of direction, magnitude or significance of the deviations from an effect under-the-null scenario.

Step 3: visualisations of specification curve analysis

The first visualisation will depict descriptive meta-analytic specification plots that display the specification curve meta-analyses (for an example with simulated data, see figure 2). In particular, these plots visualise factor-level combinations of How and Which factors constituting a given specification, the number of included samples in a specification, and the resulting meta-analytic summary effects (g) for each specification, along with 95% CIs.

The top panel in figure 2 shows the meta-analytic summary effects (g) for each specification with 95% CIs. The summary effects are sorted by their magnitude. Connecting the different summary effects results in the solid line, which is the specification curve. A horizontal

dotted line of no effect is shown at $g=0$. The vertical columns in the bottom panel represent factor combinations of How factors (in this example, different age groups and different types of guided or unguided therapy) and Which factors (different appropriate methods: fixed effect model, REM, REML estimator, 3-level meta-analysis) that constitute a given specification. Each vertical column is color coded, signifying the number of samples included in a specification (hot spectral colours for smaller numbers of included samples vs cool spectral colours code for larger number of included samples).

The second visualisation will display inferential meta-analytic specification plots that show the specification curve of the magnitude-sorted meta-analytic summary effects for all specifications. Included will be the corresponding pointwise 97.5% and 2.5% quantiles of 1000 specification curves, simulated under the null hypothesis for a given specification number using a parametric bootstrap procedure. Exceeding these limits would provide evidence against the null hypothesis ($g=0$).

In the third visualisation we follow the suggestion by Voracek *et al*²⁴ to additionally apply combinatorial meta-analysis which calculates the statistic of interest for all possible subsets of studies in the meta-analysis—for 2^{k-1} subsets when there are k studies. Although combinatorial meta-analysis is an exhaustive way to identify influential studies in a meta-analysis, it becomes computationally infeasible with an increasing number of primary studies. In the case of the dataset from the previous year, this would amount to at least $2^{363}=1.88 \times e^{109}$ meta-analyses—exceeding the computational feasibility of such a project. For this reason, Voracek *et al*²⁴ suggest only drawing a smaller, random subset due to feasibility and data visualisation considerations. We will visualise this reduced set of studies with a graphical display of study heterogeneity plot,³¹ which is particularly suited to visualise the cross-study effect heterogeneity of each subset in the combinatorial meta-analytic multiverse. As combinatorial meta-analysis is a brute-force method that automatically tests all possible study subsets in a single analysis, it will include by default many specifications that would not be regarded as reasonable. In that respect, the multiverse meta-analysis can be viewed as a theoretically and conceptually guided variant of combinatorial meta-analysis. A further important difference is that combinatorial meta-analysis analyses all study subsets with the same meta-analytic technique.²⁴ In contrast, the multiverse meta-analytic approach allows several methods (eg, fixed-effect, random effects and multilevel modelling). We will choose a random sample of 100 000 subsets for the combinatorial meta-analysis and use a stratified sampling approach based on the subsets' sizes. Thus, we can ensure the representativeness of the subset for the full set of combinations.

Data sources

We will update our existing database containing all randomised controlled trials (RCTs) on the efficacy of

psychological treatments on depression by systematic literature searches in bibliographical databases (PubMed, Embase, PsycINFO and Cochrane Register of Controlled Trials; see online supplemental file 1) for Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA) Protocols checklist³² and online supplemental file 2) for search strings). The update of the database will include studies published between 1 January 2020 and 1 January 2021.

Data extraction

After title and abstract screening, two independent researchers (PC and EK) will conduct full-text screening of all records. All eligible studies will be saved in an EndNote library and exported to R for further analyses. A PRISMA Flow Diagram will detail the entire literature search process. We will independently extract information on recruitment settings, target groups, comorbidity, intervention formats, psychotherapy types, number of sessions, control conditions and country (see online supplemental material 3 for our data extraction form). Inconsistencies will be extracted again by a different researcher.

Risk of bias assessment

Two independent researchers will assess risk of bias using the Cochrane Risk of Bias assessment tool.²⁷

Inclusion criteria

We will include all RCTs comparing a psychological treatment with a control condition. We will include studies published in English, German, Spanish or Dutch, and exclude trials on maintenance and relapse prevention as well as dissertations. Two independent researchers (PC and EK) will check all records. Eligible are all self-reported and clinician-rated instruments measuring depression. Therapies can be delivered by any person trained to deliver the therapy ranging from psychologist, psychiatrist, nurse, social worker to lay health counsellors and paraprofessionals (ie, lay people trained to deliver psychotherapy).

Exclusion criteria

Studies were excluded if (1) the study did not explicitly state it was randomised; (2) depression was not an inclusion criterion; (3) the studies investigated patients in treatment intending to prevent relapse or maintain outcomes over time; (4) the studies investigated children and adolescents; (5) the study was a dissertation; (6) the specific effects of psychological treatment could not be discerned; (7) the psychological treatment was not aimed at depression (eg, depression scores were assessed for insomnia treatment); (8) insufficient data were reported to calculate effect sizes (even if another meta-analysis reported an effect size for a specific RCT, we will not include that effect size if the RCT does not provide enough information to calculate a standardised mean difference) and if (9) studies were reported in another language than German, Dutch, English or Spanish.

Patient and public involvement

Patients or the public were not involved in the design or development of this manuscript.

ETHICS AND DISSEMINATION

Because we will not collect any primary data, the study does not require additional formal ethical assessment and informed consent. We will present the results from this study at relevant conferences and publish the results in a peer-reviewed journal. All statistical manipulations (the How factors) and study inclusion criteria (the Which factors) are addressed in detail in the present protocol and will be addressed in the final analysis as well. All components necessary for reproducible data analysis (open data, open materials, and open code) will be made accessible and will comply with the findable, accessible, interoperable and reusable guiding principles for scientific data.³³

Twitter Constantin Yves Plessen @CYPlessen and Eirini Karyotaki @KaryotakiEirini

Contributors CYP conceived of the presented idea. PC and EK prepared the data curation. CYP wrote the initial draft, which was reviewed and edited by PC and EK. PC and EK supervised the findings of this work. All authors discussed the results and contributed to the final manuscript.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

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

ORCID iDs

Constantin Yves Plessen <http://orcid.org/0000-0002-4907-3505>

Eirini Karyotaki <http://orcid.org/0000-0002-0071-2599>

Pim Cuijpers <http://orcid.org/0000-0001-5497-2743>

REFERENCES

- GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the global burden of disease study 2017. *Lancet* 2018;392:1789-858.
- Whiteford HA, Degenhardt L, Rehm J, et al. Global burden of disease attributable to mental and substance use disorders: findings from the global burden of disease study 2010. *Lancet* 2013;382:1575-86.
- Saarni SI, Suvisaari J, Sintonen H, et al. Impact of psychiatric disorders on health-related quality of life: general population survey. *Br J Psychiatry* 2007;190:326-32.
- Cuijpers P, Vogelzangs N, Twisk J, et al. Comprehensive meta-analysis of excess mortality in depression in the general community versus patients with specific illnesses. *Am J Psychiatry* 2014;171:453-62.
- Cuijpers P, van Straten A, Warmerdam L, et al. Psychological treatment of depression: a meta-analytic database of randomized studies. *BMC Psychiatry* 2008;8:36.
- Cuijpers P. Four decades of outcome research on psychotherapies for adult depression: an overview of a series of meta-analyses. *Can Psychol* 2017;58:7-19.
- Cuijpers P, Karyotaki E, de Wit L, et al. The effects of fifteen evidence-supported therapies for adult depression: a meta-analytic review. *Psychother Res* 2020;30:279-93.
- Cuijpers P, Cristea IA, Karyotaki E, et al. How effective are cognitive behavior therapies for major depression and anxiety disorders? A meta-analytic update of the evidence. *World Psychiatry* 2016;15:245-58.
- Ekers D, Webster L, Van Straten A, et al. Behavioural activation for depression; an update of meta-analysis of effectiveness and subgroup analysis. *PLoS One* 2014;9:e100100.
- Cuijpers P, Donker T, Weissman MM, et al. Interpersonal psychotherapy for mental health problems: a comprehensive meta-analysis. *Am J Psychiatry* 2016;173:680-7.
- Cuijpers P, de Wit L, Kleiboer A, et al. Problem-solving therapy for adult depression: an updated meta-analysis. *Eur Psychiatry* 2018;48:27-37.
- Driessen E, Hegelmaier LM, Abbass AA, et al. The efficacy of short-term psychodynamic psychotherapy for depression: a meta-analysis update. *Clin Psychol Rev* 2015;42:1-15.
- Leichsenring F, Rabung S. Effectiveness of long-term psychodynamic psychotherapy: a meta-analysis. *JAMA* 2008;300:1551.
- Churchill R, Moore THM, Furukawa TA, et al. 'Third wave' cognitive and behavioural therapies versus treatment as usual for depression. *Cochrane Database Syst Rev* 2013;198.
- Cuijpers P, Driessen E, Hollon SD, et al. The efficacy of non-directive supportive therapy for adult depression: a meta-analysis. *Clin Psychol Rev* 2012;32:280-91.
- Dale LP, Vanderloo L, Moore S, et al. Physical activity and depression, anxiety, and self-esteem in children and youth: an umbrella systematic review. *Ment Health Phys Act* 2019;16:66-79.
- Rebar AL, Stanton R, Geard D, et al. A meta-meta-analysis of the effect of physical activity on depression and anxiety in non-clinical adult populations. *Health Psychol Rev* 2015;9:366-78.
- Barbui C, Purgato M, Abdulmalik J, et al. Efficacy of psychosocial interventions for mental health outcomes in low-income and middle-income countries: an umbrella review. *Lancet Psychiatry* 2020;7:162-72.
- Dragioti E, Karathanos V, Gerdle B, et al. Does psychotherapy work? an umbrella review of meta-analyses of randomized controlled trials. *Acta Psychiatr Scand* 2017;136:236-46.
- Carvalho AF, Solmi M, Sanches M, et al. Evidence-Based umbrella review of 162 peripheral biomarkers for major mental disorders. *Transl Psychiatry* 2020;10:152.
- Cristea IA, Stefan S, Karyotaki E, et al. The effects of cognitive behavioral therapy are not systematically falling: a revision of Johnsen and Friborg (2015). *Psychol Bull* 2017;143:326-40.
- Johnsen TJ, Friborg O. The effects of cognitive behavioral therapy as an anti-depressive treatment is falling: a meta-analysis. *Psychol Bull* 2015;141:747-68.
- Harrer M, Karyotaki E, Cuijpers P. MetaPsyR: access the meta-analytic psychotherapy database, 2020. Available: evidencebase.dpsychotherapies.shinyapps.io/metapsy/
- Voracek M, Kossmeier M, Tran US. Which data to meta-analyze, and how?: a specification-curve and multiverse-analysis approach to meta-analysis. *Z Psychol* 2019;227:64-82.
- Simonsohn U, Simmons JP, Nelson LD. Specification curve: descriptive and inferential statistics on all reasonable specifications. *SSRN* 2015:1-18.
- Steege S, Tuerlinckx F, Gelman A, et al. Increasing transparency through a Multiverse analysis. *Perspect Psychol Sci* 2016;11:702-12.
- Higgins JPT, Altman DG, Gotzsche PC, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ* 2011;343:d5928.
- Borenstein M, Hedges L, Higgins JPT. Comprehensive meta-analysis. NJ.; 2013. Available: <https://www.meta-analysis.com/>
- Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. *J Stat Softw* 2010;36.
- van Aert RCM, MALM vanA. P-uniform* [Preprint]. *MetaArXiv* 2020.
- Olkin I, Dahabreh IJ, Trikalinos TA. GOSH - a graphical display of study heterogeneity: All-subsets meta-analysis for heterogeneity. *Res Synth Methods* 2012;3:214-23.

- 32 Shamseer L, Moher D, Clarke M, *et al*. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ* 2015;349:g7647.
- 33 Wilkinson MD, Dumontier M, Aalbersberg IJJ, IjJ A, *et al*. The fair guiding principles for scientific data management and stewardship. *Sci Data* 2016;3:160018.