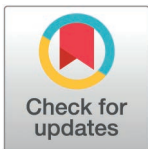


STUDY PROTOCOL

Developing a multivariable prediction model to support personalized selection among five major empirically-supported treatments for adult depression. Study protocol of a systematic review and individual participant data network meta-analysis

Ellen Driessen^{1,2*}, Orestis Efthimiou^{3,4,5}, Frederik J. Wienicke², Jasmijn Breunese², Pim Cuijpers^{6,7}, Thomas P. A. Debray^{8,9}, David J. Fisher¹⁰, Marjolein Fokkema¹¹, Toshiaki A. Furukawa¹², Steven D. Hollon¹³, Anuj H. P. Mehta¹⁴, Richard D. Riley¹⁵, Madison R. Schmidt¹⁶, Jos W. R. Twisk¹⁷, Zachary D. Cohen^{18*}



OPEN ACCESS

Citation: Driessen E, Efthimiou O, Wienicke FJ, Breunese J, Cuijpers P, Debray TPA, et al. (2025) Developing a multivariable prediction model to support personalized selection among five major empirically-supported treatments for adult depression. PLoS ONE 20(4): e0322124. <https://doi.org/10.1371/journal.pone.0322124>

Editor: Nishant Premnath Jaiswal, University of Glasgow School of Health and Wellbeing, INDIA

Received: February 2, 2024

Accepted: March 18, 2025

Published: April 23, 2025

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0322124>

Copyright: © 2025 Driessen et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

1 Department of Clinical Psychology, Behavioural Science Institute, Radboud University, Nijmegen, Netherlands, **2** Depression Expertise Center, Pro Persona Mental Health Care, Nijmegen, Netherlands, **3** Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland, **4** Institute of Primary Health Care, University of Bern, Bern, Switzerland, **5** Department of Psychiatry, University of Oxford, Oxford, United Kingdom, **6** Department of Clinical, Neuro and Developmental Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit Amsterdam, Amsterdam, Netherlands, **7** International Institute for Psychotherapy, Babeş-Bolyai University, Cluj-Napoca, Romania, **8** Smart Data Analysis and Statistics Besloten Vennootschap, Utrecht, Netherlands, **9** Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, Netherlands, **10** Medical Research Council Clinical Trials Unit, Institute of Clinical Trials and Methodology, University College London, London, United Kingdom, **11** Department of Methods and Statistics, Institute of Psychology, Leiden University, Leiden, Netherlands, **12** Department of Health Promotion and Human Behavior, Department of Clinical Epidemiology, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan, **13** Department of Psychology, Vanderbilt University, Nashville, United States of America, **14** Department of Psychological and Brain Sciences, University of Massachusetts Amherst, Amherst, United States of America, **15** Institute of Applied Health Research, University of Birmingham, Birmingham, United Kingdom, **16** Department of Clinical Psychology, Northwestern University Chicago, Chicago, United States of America, **17** Department of Epidemiology and Data Science, Amsterdam University Medical Centers, Amsterdam, Netherlands, **18** Department of Psychology, University of Arizona, Tucson, United States of America

* ellen.driessen@ru.nl (ED); zachary.d.cohen@gmail.com (ZDC)

Abstract

Background

Various treatments are recommended as first-line options in practice guidelines for depression, but it is unclear which is most efficacious for a given person. Accurate individualized predictions of relative treatment effects are needed to optimize treatment recommendations for depression and reduce this disorder's vast personal and societal costs.

Aims

We describe the protocol for a systematic review and individual participant data (IPD) network meta-analysis (NMA) to inform personalized treatment selection among five major empirically-supported depression treatments.

Data availability statement: Data availability is not applicable to this article as no data were analyzed for this study protocol paper.

Funding: ED received funding from the Dutch Research Council (<https://www.nwo.nl/en>) to supported this work (Grant number: NWO 016.Veni.195.215 6806). DF is supported by UK Research and Innovation Medical Research Council (<https://www.ukri.org/councils/mrc/>), grant number: MC_UU_00004/06. The funders had no role in the development of this study protocol, nor was there editorial direction or censorship from the sponsor in this manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: Dr. Driessen reports a grant from the Dutch Research Council during the conduct of the study. Dr. Debray owns a company that provides consulting services to pharmaceutical companies and contract research organizations. Dr. Debray is also editing a book on comparative effectiveness and personalized medicine research, for which he will receive royalties upon publication. Dr. Furukawa reports personal fees from Boehringer-Ingelheim, DT Axis, Kyoto University Original, Shionogi and SONY, and a grant from Shionogi, outside the submitted work. In addition, Dr. Furukawa has patents 2020-548587 and 2022-082495 pending, and intellectual properties for Kokoro-app licensed to Mitsubishi-Tanabe. Dr. Cohen previously held stock options in AbleTo, which he received as compensation for advice on the clinical content of Joyable's digital cognitive therapy for depression. The remaining authors have nothing to declare.

Method

We will use the METASPY database to identify randomized clinical trials that compare two or more of five treatments for adult depression: antidepressant medication, cognitive therapy, behavioral activation, interpersonal psychotherapy, and psychodynamic therapy. We will request IPD from identified studies. We will conduct an IPD-NMA and develop a multivariable prediction model that estimates individualized relative treatment effects from demographic, clinical, and psychological participant characteristics. Depressive symptom level at treatment completion will constitute the primary outcome. We will evaluate this model using a range of measures for discrimination and calibration, and examine its potential generalizability using internal-external cross-validation.

Conclusions

We describe a state-of-the-art method to predict personalized treatment effects based on IPD from multiple trials. The resulting prediction model will need prospective evaluation in mental health care for its potential to inform shared decision-making. This study will result in a unique database of IPD from randomized clinical trials around the world covering five widely used depression treatments, available for future research.

1. Introduction

People suffering from depression have a range of therapeutic options, including various antidepressant medications (ADMs) and psychological treatments such as cognitive therapy (CT), behavioral activation (BA), interpersonal psychotherapy (IPT), and short-term psychodynamic psychotherapy (STPP). Although no significant differences in average treatment effects have been found between these interventions [1,2], response is highly heterogeneous [3–5] warranting the need for more personalized treatment recommendations [6].

Current scientific methodologies are limited in their ability to discriminate individual response. What little information is available regarding patient characteristics associated with differential efficacy of depression treatments has been mostly obtained from univariable analyses [7]. These studies typically examine how an isolated patient characteristic (e.g., age) predicts differential outcomes across two treatment groups (e.g., ADM versus CT) [8–11]. Unfortunately, single predictors tend to have small effects [12], and their investigation is prone to false positive results and selective reporting. Moreover, clinical reality is much more complex, involving numerous patient characteristics and contextual factors, whose associations can vary across different treatment comparisons. Identifying and combining multiple prescriptive predictors is crucial for clinical decision-making [12–14]. Recent efforts in this regard for specific head-to-head treatment comparisons have shown promise, but resulting prediction models await validation [15,16].

The lack of scientific knowledge on treatment moderators is a barrier for efforts to develop evidence-based guidelines to help select the optimal depression treatment for an individual [17]. Current standard practice when selecting treatments for depression is largely guided by practical factors (cost, availability) or factors like clinical judgement and patient preference, which have been shown to be unrelated to symptom response [18]. All mental health stakeholders would benefit from more accurate and more personalized treatment selection strategies [19]. Research to generate precision treatment algorithms can help inform shared decision-making during the treatment selection process [18,20], and is one of many promising avenues for improving outcomes for individuals with depression [21].

In this article, we describe the protocol for a study that aims to improve personalized treatment selection for depression. We overcome the limitations of prior research by developing a model to predict relative treatment effects at the individual patient level. This model will cover five major empirically-supported depression treatments [1,22,23], will be based on multiple person and disorder characteristics, and will be evaluated for its generalizability and clinical performance.

To develop this model, we will conduct a systematic review and individual participant data (IPD) network meta-analysis (NMA) [24] comparing ADM, CT, BA, IPT, and STPP for adult depression on depressive symptom measures at treatment completion, including various available participant characteristics as potential effect modifiers. Estimation of effect modifiers becomes more reliable by combining data from multiple trials, and by focusing on IPD rather than on published aggregate data [24,25]. Furthermore, by adopting a network-based approach instead of focusing on pairwise comparisons, we can estimate treatment effects for multiple treatment comparisons (rather than only two). Moreover, IPD-NMA adequately accounts for the clustering of data within the studies and offers increased statistical power by incorporating both direct and indirect evidence [26–28]. We next describe the protocol for this study, outlining the methods that we will follow.

2. Methods

2.1. Design and pre-registration

This protocol builds on and extends prior work by our group [11,15,29–31]. This IPD-NMA will be pre-registered in the PROSPERO International prospective register of systematic reviews after acceptance of this protocol for publication in a peer-reviewed journal [29–31]. Additional important protocol amendments will be updated in this register [29–31]. Extraction of data from the primary datasets into the database for this project will start after PROSPERO registration. The Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) [32] and the Transparent Reporting of Multivariable Prediction Models Developed or Validated using Clustered Data (TRIPOD-Cluster) statements [33] guide the writing of this article (see S1 and S2 Tables), and a detailed analytic plan for predictive modelling will be pre-registered in the Open Science Framework.

2.2. Eligibility criteria

Eligible studies are randomized clinical trials for adult acute-phase depression comparing two or more of the following interventions: ADM, CT, BA, IPT, STPP. No restrictions will be placed concerning the years when the study was conducted, or with regard to publication language, date, or status [29–31].

2.2.1. Participants. Participants will be considered depressed if they meet specified criteria (e.g., Diagnostic and Statistical Manual of Mental Disorders) for major depressive disorder or another unipolar mood disorder assessed by means of a semi-structured interview or clinicians' assessment, or if they present a score at or above a validated cut-off indicating the likelihood of clinically significant depressive symptoms on an evaluator-assessed, clinician-assessed, or self-reported measure of depression (e.g., Hamilton Depression Rating Scale total score ≥ 10 ; Beck Depression Inventory total score ≥ 10) [29–31,34]. Comorbid mental and somatic disorders will be allowed [29–31,35]. Participants must be at least 18 years old, and we will place no restrictions concerning the maximum age of study participants [29–31,36].

We will assess eligibility criteria at study level rather than on the individual participant level [29–31]. Eligible participants from a study including a wider population (e.g., adults from a

study including both adolescents and adults) will not be included, because the integrity of randomization within the subgroup of eligible participants could be compromised [29–31]. We will include all participants randomized to treatment, and adopt an intention-to-treat analysis [29–31].

2.2.2. Interventions. We will focus on five major empirically-supported acute-phase depression treatments for which we expect enough trials including head-to-head comparisons with the other treatments are available to obtain relatively precise effect estimates. These five treatments are summarized below (for more extensive definitions we refer elsewhere [1]). We will not include control conditions because their inclusion may change the nature of comparisons and hence the relative effect sizes among the treatment conditions [37], whereas the comparisons of clinical interest are among the alternative treatments.

Concerning ADM, we will consider any type of standard oral antidepressant within the therapeutic dose range (e.g., selective serotonin [and noradrenaline] reuptake inhibitors, tricyclic antidepressants, monoamine oxidase inhibitors) [30].

CT aims to correct maladaptive thinking and beliefs to reduce depressive symptoms [31,38]. BA aims to increase a person's access to positively reinforcing stimuli [31,39,40]. A core technique in BA is activity scheduling, whereby individuals monitor their mood and daily activities to learn the connection between them, and then focus on increasing activities that are expected to result in a sense of pleasure, mastery, or accomplishment [31,39]. This behavioral intervention is also part of CT for depression [38] and the term cognitive-behavioral therapy (CBT) is often used in the literature to denote a single depression intervention that includes both a cognitive restructuring and a behavioral activation component [31].

We will consider an intervention to be CT if it is a manualized psychotherapy with cognitive restructuring as a main treatment component [11,31]. Behavioural techniques (e.g., activity scheduling) will be allowed, as long as they are part of an intervention protocol that is aimed at cognitive restructuring [31]. Beck's model [38] is considered the CT prototype, but other models are also eligible for inclusion [31]. We will consider an intervention to be BA if the core element of treatment is aimed at increasing positive reinforcement by means of activity scheduling [31]. Inclusion of cognitive restructuring techniques in BA will not be allowed [31].

IPT is a structured, time-limited intervention specifically developed for the treatment of major depression that focuses on current salient relational and interpersonal experiences [30]. We will consider an intervention to be IPT when it is a psychotherapy based on the manuals developed by Klerman and Weissman for IPT or for the briefer version called interpersonal counselling [30,41–45].

STPP is rooted in psychoanalytic theories, which consider the underlying personality structure to play an important role in the development and maintenance of depression. STPP aims to foster insight into (past) interpersonal relationships and unconscious feelings, desires, strivings, and thoughts to treat depression. We will consider an intervention to be STPP if it is based on psychoanalytic theories and practices, is time-limited from the onset (i.e., not a therapy that is brief only in retrospect) to distinguish it from long-term psychodynamic psychotherapy, and applies verbal techniques (e.g., therapies applying art as expression form are not considered STPP) [29].

We will include psychotherapies in any delivery format (i.e., face-to-face, telephone, or videoconferencing), as long as a clinician delivers the therapy [30,31]. Bibliotherapy, internet therapy, or other self-help formats will be excluded, as will be blended treatment formats that combine clinician-delivered therapy with internet interventions. Treatment must not exceed 6 months with no restrictions on the number of sessions [30,31]. Inpatient settings, partial hospitalization programs, and intensive outpatient programs will be excluded, since by definition more

care is provided than psychotherapy or antidepressant monotherapy. We will place no other restrictions on the setting in which treatment is delivered (e.g., primary care, outpatient mental health care) [30,31]. We will only include studies of acute-phase depression treatment [30,31]. Thus, we will exclude, for example, studies that randomized participants to the interventions as maintenance treatments after successful acute treatment [30,31]. Only data from the relevant conditions and comparisons of eligible studies will be included [30]. For example, for studies involving augmentation of antidepressants or psychotherapy following non-response to monotherapy, only study data up until the first triage point for augmentation will be included [30,46]. Similarly, for studies involving other conditions (e.g., [placebo] control, combined treatment of ADM and psychotherapy), data from these conditions will not be included.

2.3. Outcome

Depressive symptom level at treatment completion will be the primary outcome of this study, as symptom reduction is considered to be the main aim of acute-phase depression treatments [29,31]. We chose a continuous measure over categorical outcomes (e.g., remission) as our primary target because dichotomizing continuous variables reduces statistical power for estimating interactions with treatment [25,29–31].

Depressive symptom level at treatment completion is operationalized as a participant's score on the Beck Depression Inventory-II (BDI-II) [47] at the primary post-treatment time point as defined by the study's authors [30,31]. We chose the BDI-II as the main outcome because it is a self-reported patient-centered instrument that was designed to measure depression severity [47], is frequently applied in depression treatment research, and is sensitive to change [48]. If collected study data does not include BDI-II, we will convert scores on the primary (secondary, third, fourth, etc. in this order) continuous depression scale, as defined by the study's authors, into BDI-II scores using existing conversion algorithms [49–51]. If none of the measures assessed in the study is included in conversion algorithms, we aim to develop one ourselves based on the collected IPD. When none of the used outcome scales can be justifiably converted into the BDI-II, we will exclude the study from analysis. If this results in three or more excluded studies, we will consider transforming primary continuous depression outcomes to a 0–100 scale for all studies.

2.4. Predictors

Moderators of treatment effect (elsewhere also referred to as “prescriptive predictors”, “effect modifiers” or “treatment-covariate interactions”) affect the direction or magnitude of differences in outcome between two treatments [52], and thus can help predict whether a patient will benefit more from one treatment than another [18]. Variables that moderate individuals' response to treatment can also have prognostic (or “main”) effects (see Cohen & DeRubeis [18] for additional discussion), which will also be included in the model.

Putative predictor variables include demographic (e.g., age, gender), clinical (e.g., depression severity, depressive episode duration, comorbid anxiety disorder), and psychological (e.g., personality, coping style) participant characteristics assessed before the start of treatment [29–31]. Candidate variables will be selected for inclusion in the model based on a recent systematic review on predictors of treatment response in depression and sufficient availability across the studies, following the approach described by Vale and colleagues [53].

2.5. Systematic literature search

To identify eligible studies, we will use a search strategy that has been described previously [30,31]. We will search a database of randomized clinical trials examining the efficacy and

effectiveness of psychological treatments for depression that has been used in a series of published meta-analyses (www.metapsy.org) [54]. This database was developed through comprehensive literature searches in the bibliographic databases PubMed, PsycINFO, Embase.com, and the Cochrane Library and is updated every four months. The search strings use a combination of index terms and free-text words indicative of depression and psychotherapies, with filters for randomized clinical trials. The exact search terms can be found at <https://osf.io/nv3ea/>. Two raters independently screen all records, assess full-text papers for METAPSY database eligibility, and categorize the treatment comparison(s) of included studies, with disagreements being resolved in consensus in each phase [30,31].

Two other raters will independently assess all full-text papers of studies marked as comparing a psychotherapy mono-treatment condition against another active (i.e., psychotherapy or ADM) mono-treatment condition for meeting the inclusion criteria for this work [30,31]. Disagreements will be resolved through consensus and if consensus cannot be reached, a third rater will be consulted [29–31]. To identify studies that might have been missed, we will check the references of prior reviews and meta-analyses [30,31]. We will also contact psychotherapy listservs to request ongoing or unpublished studies, and studies that were missed [29,30].

2.6. IPD collection

We will invite authors of eligible studies to participate in this project using a strategy that has been successful in soliciting participation in previous depression treatment IPD meta-analyses [11,29–31]. We will invite representatives of each team that shares IPD to join as authors on all publications resulting from the use of these data, inasmuch as they meet internationally accepted criteria for authoring scientific articles (www.icmje.org) [29–31]. In addition, we will make the combined database available to investigators sharing IPD to examine other research questions, provided that the authors of the original studies approve the use of their data for this purpose [11,29–31,55].

We will apply a multi-step protocol to contact authors. We have described this protocol in more detail previously [29–31] and it has proven to be successful in reaching authors for depression treatment IPD meta-analyses. In short, we will contact the authors (starting with the corresponding author and proceeding to the other co-authors) by email with three reminders. In case of no response to email, we will first mail a letter to the corresponding author (again with three attempts) and then contact the corresponding author by telephone. This is followed by contacting the other authors by letter and telephone, and other ways to contact one of the authors (e.g., via colleagues who might know them) [29–31]. A study's data will be considered unavailable only if all these attempts fail, or an author either indicates that the IPD were not retained or declines to share these data [29–31].

We will request the following anonymized participant-level data items: randomized treatment condition, all outcome variables assessed prior to, during, and after treatment (with item-level data for depression outcome measures), and all potential predictor variables assessed in the study [30,31]. We will also ask for the study protocol and the assessment batteries used in the study.

The following study-level characteristics are extracted from the publication by two independent raters upon inclusion in the METAPSY database: country, recruitment method (e.g., community, clinical), target group (e.g., adults in general, students), depression inclusion criteria, number of treatment sessions, treatment format (e.g., individual, group), and assessment time points [30,31]. We will examine whether the ADM treatment protocols meet minimum adequate clinical best practice guidelines by assessing studies for the use of a therapeutic dosage and titration schedule (i.e., therapeutic dose achieved within three weeks) [30]. Study ADM will be deemed adequate if both criteria are met [27,30]. We will examine psychotherapy

protocol adequacy with regard to use of a treatment manual, provision of therapy by trained therapists, and verification of treatment integrity [29–31]. Treatment quality characteristics will be extracted by two raters independently and disagreements will be resolved by consensus [29–31]. If consensus cannot be reached, a third rater will be consulted. Characteristics not reported in the publications will be requested from the authors [29–31].

Risk of bias in the included studies will be assessed using the Cochrane risk-of-bias tool for randomized trials version 2 [56], both before and after the IPD are obtained. Ratings will be primarily based on information reported in the publications, though risk of bias due to deviations from the intended intervention, due to missing outcome data, and in selection of the reported result will be assessed using the IPD (i.e., domain 3.1) or the IPD meta-analysis' abilities to correct for the relevant risk of bias (i.e., domains 2.6–2.7, 3.2, and 5.1–5.3). Two raters will independently assess the risk-of-bias tool at study level. Disagreements will be resolved by consensus [30,31]. If consensus cannot be reached a third rater will be consulted. As studies are expected to be included that were published before the universal adoption of reporting guidelines for randomized clinical trials [57], requisite information will be requested from the authors if not reported in the publications [31].

2.7. IPD integrity checks

We will perform three data integrity checks [29–31]. First, we will check whether the dataset includes the full intention-to-treat sample (defined as all participants randomized to treatment in the study) and otherwise matches the data reported in the published article [29–31]. To this end, means and standard deviations or raw counts for all baseline characteristics, and observed pre-treatment and post-treatment scores reported in the article will also be calculated from the dataset and both will be compared [29–31]. Second, we will check whether all outcome and potential predictor variables reported in the article are included in the dataset [29–31]. Third, the outcome and predictor variables will be checked for inconsistent, invalid, or out-of-range values [29–31], including values that conflict with the primary study's or this work's eligibility criteria (e.g., age < 18). Discrepancies resulting from these data integrity checks will be discussed with the authors [29–31]. In addition, authors will be contacted if concerns arise regarding duplicate participants across trials (e.g., when multiple trials are conducted by the same author group).

2.8. Building the IPD database

All datasets received that contain IPD for at least one continuous depressive symptom measure will be considered for quantitative synthesis and will be pooled in a single database [29–31]. Predictor variables are expected to be assessed differently across individual studies and will therefore be harmonized [29–31]. Categorical variables (e.g., marital status) will be recoded so that they contain similar categories and corresponding coding schemes across all studies that include that variable (e.g., 0=single, 1=married, 2=separated/divorced, 3=widowed, 4=not married, but living together, 5=other).

To improve rigor and reproducibility of our data processing and cleaning pipelines, we developed a spreadsheet- and code-based system that also increases efficiency and scalability, allowing for changes to be instantaneously implemented across the database, and facilitating addition of new data. We will describe this system in more detail on the project's Open Science Framework page.

2.9. Missing data imputation

Data will be imputed for variables that are 'systematically' missing for certain studies and values that are 'sporadically' missing for certain individual participants. We will use methods for

multiple imputation that account for the clustering of participants in different studies [58,59], utilizing both baseline variables as well as all intermediate and endpoint outcomes available in the datasets [60]. If there are no systematically missing predictors in the studies, we will perform multiple imputation of missing data for each study separately as our primary analysis, and use the procedure described above as sensitivity analysis. Single imputation will also be considered, for instance for imputation of baseline variables, if it follows recommended advice [61].

2.10. Modelling approach

First, we will estimate the average relative treatment effects (mean differences) using pairwise meta-analyses and a random effects NMA based on aggregate data. NMA enables us to obtain relative treatment effects for all treatment comparisons, even for those without direct evidence [62], as long as they are connected in the network. One prerequisite for NMA is transitivity, i.e., that the included studies do not differ with respect to the distribution of effect modifiers when grouped by comparison. Transitivity also means that a ‘mega-trial’ would be possible, i.e., a hypothetical trial comparing all treatments in the network, and that all patients would in principle be eligible to receive any of these treatments. We will assess transitivity conceptually and by examining candidate effect modifiers across comparisons. Lack of transitivity may affect the consistency of the network, or the extent to which the direct and indirect estimated effect of a particular comparison align. When direct and indirect estimated effects are inconsistent (i.e., do not align), NMA results may be invalid. Therefore, we will evaluate statistical consistency using the design-by-treatment interaction model [63] and the back calculation method [64]. We will assume a common heterogeneity parameter in the network, and we will assess its extent by looking at prediction intervals. For these analyses, we will use the netmeta package in R [65].

To model the relative individualized treatment effects amongst the five treatments, we will conduct a two-stage IPD-NMA in the rjags R package [66]. In the first stage, we will analyze each study separately using a suitable regression model (e.g., linear regression adjusting for baseline) to estimate the treatment effect at average values of the predictors as well as all treatment-predictor interactions. For estimating treatment-predictor interactions, we will consider using penalization (shrinkage). Such methods tend to shrink (decrease the values of) the estimated coefficients as compared to usual maximum likelihood approaches. A simulation has shown that penalization methods may lead to more accurate predictions (i.e., smaller mean squared error) regarding patient-specific treatment effects in IPD meta-analysis compared to unpenalized methods [67]. We will fit the model in each of the multiply imputed datasets for each study. Then, we will combine posterior samples of the model parameters across the imputed datasets, to estimate study-level parameters and their corresponding uncertainty [68].

In the second stage, we will synthesize all study-specific results (treatment effect estimates, and interactions, and their standard errors) using a (multivariate) NMA model, where, as above, we will assume common between-study heterogeneity across all treatment comparisons. The end product will be a set of estimates of the treatment effect at average values of the predictors as well as treatment-predictor interactions (i.e., effect modification). The final model will allow us, given a participant’s baseline values, to predict individualized treatment effects between any two treatments (T_A and T_B) in the network, where treatment effect means the predicted difference in outcome if the individual received T_A versus T_B .

2.11. Cross-validation of the model

We will assess the performance of the model with internal-external cross-validation [69,70]. More specifically, we will repeat the analytical process described above after taking one

study out of our sample and use the remaining studies to fit our model. Then, we will apply the model to the left-out study, and for each patient therein we will calculate the predicted treatment effects among all treatments in the study. Finally, using these predictions, we will evaluate model performance within and across studies using meta-analysis methods. Conceptually, this evaluation will follow the Personalized Advantage Index approach described by DeRubeis and colleagues [71] in which outcomes of participants who received the treatment recommended by the model are compared against the outcomes of participants who received a non-recommended treatment [71]. This has been also termed “population benefit” index [72]. This measure combines aspects of discrimination-for-benefit and calibration-for-benefit. A useful model for personalized treatment effects should have an index larger than the average treatment effect. We will also assess calibration-for-benefit purely by producing calibration plots and fitting a regression for benefit [72]. Analyses will be done with the *predieval* R package (v.0.1.1) [73].

2.12. Assessment of meta-biases

To assess the potential impact of risk-of-bias, we will conduct a random effect NMA excluding studies at high risk of bias. We will visually compare the 95% confidence intervals of the mean relative treatment effects for all treatment comparisons between this and the primary analysis.

We will assess potential data-availability bias by comparing effect sizes (i.e., mean relative treatment effects) between studies for which IPD were and were not available with a subgroup NMA. To facilitate comparability, effect sizes will be calculated based on the effect size data extracted from the publications upon inclusion in the METAPSY database for all studies (rather than being based on the IPD when available) [31].

Small study effects (potential publication bias) will be assessed by examining asymmetry in contour-enhanced funnel plots with Egger’s test of the intercept for pairwise comparisons including ten or more trials [30,31,74]. The analyses will include both studies for which IPD were versus were not available and will, therefore, also be based on the effect size data extracted from the publications [31]. We will assess the strength of the body of evidence based on the number of included studies and participants as well as the quality of the included studies [30,31].

2.13. Ethical standards

Institutional Review Board (IRB) approval is not required for this project, because we will work with anonymized data from treatment studies that have already been completed [29–31]. IRB approval might be required for the investigators to share their IPD depending on their institution’s policies [29–31]. It will be the responsibility of the investigators to obtain IRB approval if their institution’s policies require them to do so [29–31]. By signing the data sharing agreement, the authors who share their IPD will declare that those data were collected and will be transferred according to all applicable local and international laws and regulations [29–31]. Furthermore, they declare that all IPD will be anonymized, so that no personal data are transferred [31].

3. Discussion

In this article, we described the rationale and protocol for a study that aims to improve personalized treatment selection for depression. More specifically, we aim to build an IPD database from existing randomized clinical trials worldwide covering five major empirically-supported depression treatments, and to conduct an IPD-NMA to estimate individualized relative treatment effects for these five interventions.

This study has several strengths. First, we describe a data-analysis strategy that combines state-of-the-art methods to predict personalized treatment effects based on combined data from multiple clinical trials. Second, we will build a unique patient-level database of worldwide trials covering five of the most widely used depression treatments. We will make this database available for future research by ourselves and others, beyond the scope and duration of this project, thus prospectively reducing the burden that clinical research poses on patients. Moreover, we will build the lasting infrastructure that will allow this database to further expand (e.g., adding new studies), as well as systems that could be used to develop similar IPD databases in other contexts. Third, applying the data-analysis strategy in this database will result in a prediction model, which is based on multiple observed person and illness characteristics and results in directly usable clinical recommendations for optimizing treatment selection for individuals in the context of multiple treatments. Overcoming an important limitation of prior research in this field, we will evaluate this model and assess its potential generalizability. We will examine model performance using internal-external cross-validation to reduce the potential for biased estimates and overfitting.

This study also has several limitations. First, primary studies will likely show considerable heterogeneity in key features such as which and how predictors were assessed, number and frequency of treatment sessions, and assessment timing. Harmonization of variables across relevant studies will also be necessary to account for variation in how predictors were operationalized across different studies [29–31]. Recoding categorical variables into a smaller number of categories might be necessary, despite the potential resulting loss of information [29–31]. Second, this study does not cover all available depression treatments. We have focused on five widely used and studied interventions for which we expect a substantial amount of comparative IPD will be available. However, our methodology will be adaptable for future inclusion of low-intensity treatments (e.g., psychoeducation), other psychotherapies (e.g., problem-solving therapy), combined treatment of psychotherapy and ADM [46], or depression treatments in different contexts (e.g., digital therapy, inpatient treatment, or brain stimulation therapies). Third, although identification of studies is based on systematic literature searches, it is possible that we might miss studies. For instance, the searches underlying the METAPSY database do not include a formal grey literature search. In addition, we expect that IPD will not be obtained for all studies that we do identify. Although a recent study found few indications for data-availability bias in IPD meta-analyses [75], we will assess empirically the extent to which mean relative treatment effects differ between studies for which IPD are versus are not available. We also will examine empirically potential publication bias.

If accuracy and generalizability of the resulting statistical model's predictions are deemed adequate, it could be used to develop a depression treatment selection tool. This tool would need to be evaluated prospectively in clinical practice before it can be implemented in mental health care to facilitate knowledge-based decision making, for instance by prospective randomized comparisons of model-informed treatment allocation versus allocation-as-usual or randomized allocation [76]. Future work should also focus on developing and evaluating implementation strategies for precision mental health approaches [19]. We hope that the results of this study will be evaluated further for their potential to inform shared decision making and help depressed individuals receive the optimal treatment given their personal characteristics. In this way, we hope to contribute to personalizing evidence-based treatment selection for depression to reduce this disorder's tremendous personal and societal costs.

Supporting information

S1 Table. PRISMA-P (Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols) 2015 Checklist: Recommended Items to Address in a Systematic Review Protocol. (DOCX)

S2 Table. TRIPOD-Cluster Checklist of Items to Include When Reporting a Study Developing or Validating a Multivariable Prediction Model Using Clustered Data. (DOCX)

Author contributions

Conceptualization: Ellen Driessen, Pim Cuijpers, Marjolein Fokkema, Steven D. Hollon, Jos W. R. Twisk, Zachary D. Cohen.

Data curation: Zachary D. Cohen.

Funding acquisition: Ellen Driessen, Pim Cuijpers, Marjolein Fokkema, Steven D. Hollon, Jos W. R. Twisk, Zachary D. Cohen.

Investigation: Ellen Driessen, Frederik J. Wienicke, Jasmijn Breunese, Pim Cuijpers, Steven D. Hollon, Anuj H. P. Mehta, Madison R. Schmidt, Zachary D. Cohen.

Methodology: Ellen Driessen, Orestis Efthimiou, Frederik J. Wienicke, Pim Cuijpers, Thomas P. A. Debray, David J. Fisher, Marjolein Fokkema, Toshiaki A. Furukawa, Steven D. Hollon, Richard D. Riley, Jos W. R. Twisk, Zachary D. Cohen.

Project administration: Ellen Driessen, Jasmijn Breunese, Madison R. Schmidt.

Software: Zachary D. Cohen.

Supervision: Ellen Driessen, Zachary D. Cohen.

Writing – original draft: Ellen Driessen, Orestis Efthimiou, Frederik J. Wienicke, Jasmijn Breunese, Zachary D. Cohen.

Writing – review & editing: Ellen Driessen, Orestis Efthimiou, Frederik J. Wienicke, Jasmijn Breunese, Pim Cuijpers, Thomas P. A. Debray, David J. Fisher, Marjolein Fokkema, Toshiaki A. Furukawa, Steven D. Hollon, Anuj H. P. Mehta, Richard D. Riley, Madison R. Schmidt, Jos W. R. Twisk, Zachary D. Cohen.

References

1. Cuijpers P, Karyotaki E, de Wit L, Ebert DD. The effects of fifteen evidence-supported therapies for adult depression: A meta-analytic review. *Psychother Res*. 2020;30(3):279–93. <https://doi.org/10.1080/10503307.2019.1649732> PMID: 31394976
2. Cuijpers P, Quero S, Noma H, Ciharova M, Miguel C, Karyotaki E, et al. Psychotherapies for depression: a network meta-analysis covering efficacy, acceptability and long-term outcomes of all main treatment types. *World Psychiatry*. 2021; 20(2): 283–93.
3. Kaiser T, Volkmann C, Volkmann A, Karyotaki E, Cuijpers P, Brakemeier EL. Heterogeneity of treatment effects in trials on psychotherapy of depression. *Clin Psychol*. 2022; 29(3): 294–303.
4. Stone MB, Yaseen ZS, Miller BJ, Richardville K, Kalaria SN, Kirsch I. Response to acute monotherapy for major depressive disorder in randomized, placebo controlled trials submitted to the US Food and Drug Administration: individual participant data analysis. *BMJ*. 2022;378:e067606. <https://doi.org/10.1136/bmj-2021-067606> PMID: 35918097
5. Volkmann C, Volkmann A, Müller CA. On the treatment effect heterogeneity of antidepressants in major depression: A Bayesian meta-analysis and simulation study. *PLoS One*. 2020;15(11):e0241497. <https://doi.org/10.1371/journal.pone.0241497> PMID: 33175895
6. DeRubeis RJ. The history, current status, and possible future of precision mental health. *Behav Res Ther*. 2019;123:103506. <https://doi.org/10.1016/j.brat.2019.103506> PMID: 31706160
7. Cuijpers P, Ebert DD, Acarturk C, Andersson G, Cristea IA. Personalized Psychotherapy for Adult Depression: A Meta-Analytic Review. *Behav Ther*. 2016;47(6):966–80. <https://doi.org/10.1016/j.beth.2016.04.007> PMID: 27993344
8. Driessen E, Cuijpers P, Hollon SD, Dekker JJM. Does pretreatment severity moderate the efficacy of psychological treatment of adult outpatient depression? A meta-analysis. *J Consult Clin Psychol*. 2010;78(5):668–80. <https://doi.org/10.1037/a0020570> PMID: 20873902

9. Fournier JC, DeRubeis RJ, Hollon SD, Dimidjian S, Amsterdam JD, Shelton RC, et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA*. 2010;303(1):47–53. <https://doi.org/10.1001/jama.2009.1943> PMID: 20051569
10. Cuijpers P, Weitz E, Twisk J, Kuehner C, Cristea I, David D, et al. Gender as predictor and moderator of outcome in cognitive behavior therapy and pharmacotherapy for adult depression: an “individual patient data” meta-analysis. *Depress Anxiety*. 2014;31(11):941–51. <https://doi.org/10.1002/da.22328> PMID: 25407584
11. Weitz ES, Hollon SD, Twisk J, van Straten A, Huibers MJH, David D, et al. Baseline Depression Severity as Moderator of Depression Outcomes Between Cognitive Behavioral Therapy vs Pharmacotherapy: An Individual Patient Data Meta-analysis. *JAMA Psychiatry*. 2015;72(11):1102–9. <https://doi.org/10.1001/jamapsychiatry.2015.1516> PMID: 26397232
12. Furukawa TA, Weitz ES, Tanaka S, Hollon SD, Hofmann SG, Andersson G, et al. Initial severity of depression and efficacy of cognitive-behavioural therapy: individual-participant data meta-analysis of pill-placebo-controlled trials. *Br J Psychiatry*. 2017;210(3):190–6. <https://doi.org/10.1192/bjp.bp.116.187773> PMID: 28104735
13. Meehl PE. Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence. American Psychological Association; 1954.
14. Kessler RC, van Loo HM, Wardenaar KJ, Bossarte RM, Brenner LA, Ebert DD, et al. Using patient self-reports to study heterogeneity of treatment effects in major depressive disorder. *Epidemiol Psychiatr Sci*. 2017;26(1):22–36. <https://doi.org/10.1017/S2045796016000020> PMID: 26810628
15. Furukawa TA, Efthimiou O, Weitz ES, Cipriani A, Keller MB, Kocsis JH, et al. Cognitive-Behavioral Analysis System of Psychotherapy, Drug, or Their Combination for Persistent Depressive Disorder: Personalizing the Treatment Choice Using Individual Participant Data Network Metaregression. *Psychother Psychosom*. 2018;87(3):140–53. <https://doi.org/10.1159/000489227> PMID: 29847831
16. Furukawa TA, Suganuma A, Ostinelli EG, Andersson G, Beevers CG, Shumake J, 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):500–11. [https://doi.org/10.1016/S2215-0366\(21\)00077-8](https://doi.org/10.1016/S2215-0366(21)00077-8) PMID: 33957075
17. Cuijpers P. Psychotherapies for adult depression: recent developments. *Curr Opin Psychiatry*. 2015;28(1):24–9. <https://doi.org/10.1097/YCO.0000000000000121> PMID: 25415495
18. Cohen ZD, DeRubeis RJ. Treatment Selection in Depression. *Annu Rev Clin Psychol*. 2018;14:209–36. <https://doi.org/10.1146/annurev-clinpsy-050817-084746> PMID: 29494258
19. Deisenhofer A-K, Barkham M, Beierl ET, Schwartz B, Aafjes-van Doorn K, Beevers CG, et al. Implementing precision methods in personalizing psychological therapies: Barriers and possible ways forward. *Behav Res Ther*. 2024;172:104443. <https://doi.org/10.1016/j.brat.2023.104443> PMID: 38086157
20. Nye A, Delgadillo J, Barkham M. Efficacy of personalized psychological interventions: A systematic review and meta-analysis. *J Consult Clin Psychol*. 2023;91(7):389–97. <https://doi.org/10.1037/ccp0000820> PMID: 37166831
21. Holmes EA, Ghaderi A, Harmer CJ, Ramchandani PG, Cuijpers P, Morrison AP, et al. The Lancet Psychiatry Commission on psychological treatments research in tomorrow's science. *Lancet Psychiatry*. 2018;5(3):237–86. [https://doi.org/10.1016/S2215-0366\(17\)30513-8](https://doi.org/10.1016/S2215-0366(17)30513-8) PMID: 29482764
22. Marcus SC, Olfson M. National trends in the treatment for depression from 1998 to 2007. *Arch Gen Psychiatry*. 2010;67(12):1265–73. <https://doi.org/10.1001/archgenpsychiatry.2010.151> PMID: 21135326
23. Driessen E, Hegelmaier LM, Abbass AA, Barber JP, Dekker JJM, Van HL, et al. The efficacy of short-term psychodynamic psychotherapy for depression: A meta-analysis update. *Clin Psychol Rev*. 2015;42:1–15. <https://doi.org/10.1016/j.cpr.2015.07.004> PMID: 26281018
24. Riley RD, Tierney J, Stewart LA. Individual Participant Data Meta-Analysis: A Handbook for Healthcare Research. Wiley; 2021.
25. Riley RD, Debray TPA, Fisher D, Hattle M, Marlin N, Hoogland J, et al. Individual participant data meta-analysis to examine interactions between treatment effect and participant-level covariates: Statistical recommendations for conduct and planning. *Stat Med*. 2020;39(15):2115–37. <https://doi.org/10.1002/sim.8516> PMID: 32350891
26. Debray TP, Schuit E, Efthimiou O, Reitsma JB, Ioannidis JP, Salanti G, et al. An overview of methods for network meta-analysis using individual participant data: when do benefits arise? *Stat Methods Med Res*. 2018;27(5):1351–64. <https://doi.org/10.1177/0962280216660741> PMID: 27487843
27. Cuijpers P, Noma H, Karyotaki E, Vinkers CH, Cipriani A, Furukawa TA. A network meta-analysis of the effects of psychotherapies, pharmacotherapies and their combination in the treatment of adult depression. *World Psychiatry*. 2020;19(1):92–107. <https://doi.org/10.1002/wps.20701> PMID: 31922679

28. Riley RD, Dias S, Donegan S, Tierney JF, Stewart LA, Efthimiou O, et al. Using individual participant data to improve network meta-analysis projects. *BMJ Evid Based Med*. 2023;28(3):197–203. <https://doi.org/10.1136/bmjebm-2022-111931> PMID: 35948411
29. Driessen E, Abbass AA, Barber JP, Connolly Gibbons MB, Dekker JJM, Fokkema M, et al. Which patients benefit specifically from short-term psychodynamic psychotherapy (STPP) for depression? Study protocol of a systematic review and meta-analysis of individual participant data. *BMJ Open*. 2018;8(2):e018900. <https://doi.org/10.1136/bmjopen-2017-018900> PMID: 29463590
30. Driessen E, Cohen ZD, Weissman MM, Markowitz JC, Weitz ES, Hollon SD, et al. The efficacy of antidepressant medication and interpersonal psychotherapy for adult acute-phase depression: study protocol of a systematic review and meta-analysis of individual participant data. *BJPsych Open*. 2021;7(2):e56. <https://doi.org/10.1192/bjo.2021.4> PMID: 33602371
31. Driessen E, Cohen ZD, Lorenzo-Luaces L, Hollon SD, Richards DA, Dobson KS, et al. Efficacy and moderators of cognitive therapy versus behavioural activation for adults with depression: study protocol of a systematic review and meta-analysis of individual participant data. *BJPsych Open*. 2022;8(5):e154. <https://doi.org/10.1192/bjo.2022.560> PMID: 35946068
32. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ*. 2015;350:g7647. <https://doi.org/10.1136/bmj.g7647> PMID: 25555855
33. Debray TPA, Collins GS, Riley RD, Snell KIE, Van Calster B, Reitsma JB, et al. Transparent reporting of multivariable prediction models developed or validated using clustered data (TRIPOD-Cluster): explanation and elaboration. *BMJ*. 2023;380:e071058. <https://doi.org/10.1136/bmj-2022-071058> PMID: 36750236
34. Levis B, Benedetti A, Thombs BD, DEPRESSION Screening Data (DEPRESSD) Collaboration. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. *BMJ*. 2019;365:l1476. <https://doi.org/10.1136/bmj.l1476> PMID: 30967483
35. Miguel C, Karyotaki E, Ciharova M, Cristea I, Penninx B, Cuijpers P. Psychotherapy for comorbid depression and somatic disorders: a systematic review and meta-analysis. *Psychol Med*. 2023;53(6):2503–13.
36. Cuijpers P, Karyotaki E, Eckshtain D, Ng MY, Corteselli KA, Noma H, et al. Psychotherapy for Depression Across Different Age Groups: A Systematic Review and Meta-analysis. *JAMA Psychiatry*. 2020;77(7):694–702. <https://doi.org/10.1001/jamapsychiatry.2020.0164> PMID: 32186668
37. Salanti G, Chaimani A, Furukawa TA, Higgins JPT, Ogawa Y, Cipriani A, et al. Impact of placebo arms on outcomes in antidepressant trials: systematic review and meta-regression analysis. *Int J Epidemiol*. 2018;47(5):1454–64. <https://doi.org/10.1093/ije/dyy076> PMID: 29878123
38. Beck AT, Rush AJ, Shaw BF, Emery G. *Cognitive Therapy of Depression*. Guilford Press; 1979.
39. Lewinsohn PM. A behavioral approach to depression. In: Friedman RJ, Katz MM, editor. *The Psychology of Depression: Contemporary Theory and Research*. John Wiley & Sons; 1974. p.157–85.
40. Jacobson NS, Martell CR, Dimidjian S. Behavioral activation treatment for depression: returning to contextual roots. *Clin Psychol*. 2001;8(3):255–70.
41. Klerman GL, Weissman MM, Rounsaville BJ, Chevron E. *Interpersonal Psychotherapy of Depression*. Basic Books; 1984.
42. Weissman MM, Markowitz JC, Klerman GL. *Comprehensive Guide to Interpersonal Psychotherapy*. Basic Books; 2000.
43. Weissman MM, Markowitz JC, L. KG. *Clinician's Quick Guide to Interpersonal Psychotherapy*. Oxford University Press; 2007.
44. Weissman M, Hankerson S, Scorza P, Olfson M, Verdelli H, Shea S. Interpersonal counselling (IPC) for depression in primary care. *Am J Psychother*. 2014;68(4).
45. Weissman MM, Markowitz JC, L. KG. *The Guide to Interpersonal Psychotherapy*. Oxford University Press; 2018.
46. Driessen E, Fokkema M, Dekker JJM, Peen J, Van HL, Maina G, et al. Which patients benefit from adding short-term psychodynamic psychotherapy to antidepressants in the treatment of depression? A systematic review and meta-analysis of individual participant data. *Psychol Med*. 2023;53(13):6090–101. <https://doi.org/10.1017/S0033291722003270> PMID: 36404677
47. Beck AT, Steer RA, Brown GK. *Manual for the Beck Depression Inventory-II*. Psychological Corporation; 1996.
48. Schneibel R, Brakemeier E-L, Wilbertz G, Dykieriek P, Zobel I, Schramm E. Sensitivity to detect change and the correlation of clinical factors with the Hamilton Depression Rating Scale and the

- Beck Depression Inventory in depressed inpatients. *Psychiatry Res.* 2012;198(1):62–7. <https://doi.org/10.1016/j.psychres.2011.11.014> PMID: 22445070
49. Wahl I, Löwe B, Bjorner J, Fischer F, Langs G, Voderholzer U, et al. Standardization of depression measurement: a common metric was developed for 11 self-report depression measures. *J Clin Epidemiol.* 2014;67(1):73–86.
 50. Leucht S, Fennema H, Engel RR, Kaspers-Janssen M, Szegedi A. Translating the HAM-D into the MADRS and vice versa with equipercenile linking. *J Affect Disord.* 2018;226:326–31. <https://doi.org/10.1016/j.jad.2017.09.042> PMID: 29031182
 51. McCabe-Beane JE, Segre LS, Perkhounkova Y, Stuart S, O'Hara MW. The identification of severity ranges for the Edinburgh Postnatal Depression Scale. *J Reprod Infant Psychol.* 2016;34(3):293–303. <https://doi.org/10.1080/02646838.2016.1141346>
 52. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. *J Pers Soc Psychol.* 1986;51(6):1173–82. <https://doi.org/10.1037//0022-3514.51.6.1173> PMID: 3806354
 53. Vale CL, Fisher DJ, Godolphin PJ, Rydzewska LH, Boher J-M, Burdett S, et al. Which patients with metastatic hormone-sensitive prostate cancer benefit from docetaxel: a systematic review and meta-analysis of individual participant data from randomised trials. *Lancet Oncol.* 2023;24(7):783–97. [https://doi.org/10.1016/S1470-2045\(23\)00230-9](https://doi.org/10.1016/S1470-2045(23)00230-9) PMID: 37414011
 54. Cuijpers P, Miguel C, Harrer M, Plessen CY, Ciharova M, Papola D, et al. Psychological treatment of depression: A systematic overview of a “Meta-Analytic Research Domain”. *J Affect Disord.* 2023;335:141–51. <https://doi.org/10.1016/j.jad.2023.05.011> PMID: 37178828
 55. Rydzewska LHM, Stewart LA, Tierney JF. Sharing individual participant data: through a systematic reviewer lens. *Trials.* 2022;23(1):167. <https://doi.org/10.1186/s13063-021-05787-4> PMID: 35189931
 56. Higgins JPT, Savović J, Page MJ, Elbers RG, Sterne JAC. Chapter 8: Assessing risk of bias in a randomized trial. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editor. *Cochrane Handbook for Systematic Reviews of Interventions*, Version 6.3 (Updated February 2022). Cochrane; 2020.
 57. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA.* 1996;276(8):637–9. <https://doi.org/10.1001/jama.276.8.637> PMID: 8773637
 58. Resche-Rigon M, White IR. Multiple imputation by chained equations for systematically and sporadically missing multilevel data. *Stat Methods Med Res.* 2018;27(6):1634–49. <https://doi.org/10.1177/0962280216666564> PMID: 27647809
 59. Jolani S. Hierarchical imputation of systematically and sporadically missing data: An approximate Bayesian approach using chained equations. *Biom J.* 2018;60(2):333–51. <https://doi.org/10.1002/bimj.201600220> PMID: 28990686
 60. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ.* 2009;338:b2393. <https://doi.org/10.1136/bmj.b2393> PMID: 19564179
 61. White IR, Thompson SG. Adjusting for partially missing baseline measurements in randomized trials. *Stat Med.* 2005;24(7):993–1007. <https://doi.org/10.1002/sim.1981> PMID: 15570623
 62. Efthimiou O, Debray TPA, van Valkenhoef G, Trelle S, Panayidou K, Moons KGM, et al. GetReal in network meta-analysis: a review of the methodology. *Res Synth Methods.* 2016;7(3):236–63. <https://doi.org/10.1002/jrsm.1195> PMID: 26754852
 63. Higgins JPT, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies. *Res Synth Methods.* 2012;3(2):98–110. <https://doi.org/10.1002/jrsm.1044> PMID: 26062084
 64. König J, Krahn U, Binder H. Visualizing the flow of evidence in network meta-analysis and characterizing mixed treatment comparisons. *Stat Med.* 2013;32(30):5414–29. <https://doi.org/10.1002/sim.6001> PMID: 24123165
 65. Balduzzi S, Rücker G, Nikolakopoulou A, Papakonstantinou T, Salanti G, Efthimiou O. Netmeta: an R package for network meta-analysis using frequentist methods. *J Stat Softw.* 2023;106(2):1–40.
 66. Plummer M, Stukalov A, Denwood M. Rjags: Bayesian graphical models using MCMC. <https://mcmc-jags.sourceforge.io/>
 67. Seo M, White IR, Furukawa TA, Imai H, Valgimigli M, Egger M, et al. Comparing methods for estimating patient-specific treatment effects in individual patient data meta-analysis. *Stat Med.* 2021;40(6):1553–73. <https://doi.org/10.1002/sim.8859> PMID: 33368415
 68. Zhou X, Reiter J. A note on Bayesian inference after multiple imputation. *Am Stat.* 2012;64(2):159–63.

69. Royston P, Parmar M, Sylvester R. Construction and validation of a prognostic model across several studies, with an application in superficial bladder cancer. *Stat Med*. 2004;23(6):907–26.
70. Debray TPA, Moons KGM, Ahmed I, Koffijberg H, Riley RD. A framework for developing, implementing, and evaluating clinical prediction models in an individual participant data meta-analysis. *Stat Med*. 2013;32(18):3158–80. <https://doi.org/10.1002/sim.5732> PMID: 23307585
71. DeRubeis RJ, Cohen ZD, Forand NR, Fournier JC, Gelfand LA, Lorenzo-Luaces L. The Personalized Advantage Index: translating research on prediction into individualized treatment recommendations. A demonstration. *PLoS One*. 2014;9(1):e83875. <https://doi.org/10.1371/journal.pone.0083875> PMID: 24416178
72. Efthimiou O, Hoogland J, Debray T, Seo M, Furukawa T, Egger M. Measuring the performance of prediction models to personalize treatment choice. *Stat Med*. 2023;42(8):1188–206.
73. Efthimiou O. Predieval: Assessing Performance of Prediction Models for Predicting Patient-Level Treatment Benefit. R package version 0.1.1. <https://CRAN.R-project.org/package=predieval> 2022.
74. Sterne JAC, Sutton AJ, Ioannidis JPA, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*. 2011;343:d4002. <https://doi.org/10.1136/bmj.d4002> PMID: 21784880
75. Tsujimoto Y, Fujii T, Onishi A, Omae K, Luo Y, Imai H, et al. No consistent evidence of data availability bias existed in recent individual participant data meta-analyses: a meta-epidemiological study. *J Clin Epidemiol*. 2020;118:107–114.e5. <https://doi.org/10.1016/j.jclinepi.2019.10.004> PMID: 31654789
76. Chekroud AM, Bondar J, Delgadillo J, Doherty G, Wasil A, Fokkema M, et al. The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry*. 2021;20(2):154–70. <https://doi.org/10.1002/wps.20882> PMID: 34002503