


REGISTERED REPORT

Open Access



# Psychotherapy or medication for depression? Using individual symptom meta-analyses to derive a Symptom-Oriented Therapy (SO<sub>r</sub>T) metric for a personalised psychiatry

Nils Kappelmann<sup>1,2\*</sup> , Martin Rein<sup>1</sup>, Julia Fietz<sup>1,2</sup>, Helen S. Mayberg<sup>3,4</sup>, W. Edward Craighead<sup>4,5</sup>, Boadie W. Dunlop<sup>4</sup>, Charles B. Nemeroff<sup>6</sup>, Martin Keller<sup>7</sup>, Daniel N. Klein<sup>8</sup>, Bruce A. Arnow<sup>9</sup>, Nusrat Husain<sup>10</sup>, Robin B. Jarrett<sup>11</sup>, Jeffrey R. Vittengl<sup>12</sup>, Marco Menchetti<sup>13</sup>, Gordon Parker<sup>14</sup>, Jacques P. Barber<sup>15</sup>, Andre G. Bastos<sup>16</sup>, Jack Dekker<sup>17</sup>, Jaap Peen<sup>17</sup>, Martin E. Keck<sup>1</sup> and Johannes Kopf-Beck<sup>1</sup>

## Abstract

**Background:** Antidepressant medication (ADM) and psychotherapy are effective treatments for major depressive disorder (MDD). It is unclear, however, if treatments differ in their effectiveness at the symptom level and whether symptom information can be utilised to inform treatment allocation. The present study synthesises comparative effectiveness information from randomised controlled trials (RCTs) of ADM versus psychotherapy for MDD at the symptom level and develops and tests the Symptom-Oriented Therapy (SO<sub>r</sub>T) metric for precision treatment allocation.

**Methods:** First, we conducted systematic review and meta-analyses of RCTs comparing ADM and psychotherapy at the individual symptom level. We searched PubMed Medline, PsycINFO, and the Cochrane Central Register of Controlled Trials databases, a database specific for psychotherapy RCTs, and looked for unpublished RCTs. Random-effects meta-analyses were applied on sum-scores and for individual symptoms for the Hamilton Rating Scale for Depression (HAM-D) and Beck Depression Inventory (BDI) measures.

Second, we computed the SO<sub>r</sub>T metric, which combines meta-analytic effect sizes with patients' symptom profiles. The SO<sub>r</sub>T metric was evaluated using data from the Munich Antidepressant Response Signature (MARS) study ( $n = 407$ ) and the Emory Predictors of Remission in Depression to Individual and Combined Treatments (PRE<sub>r</sub>DICT) study ( $n = 234$ ).

(Continued on next page)

\* Correspondence: [nils\\_kappelmann@psych.mpg.de](mailto:nils_kappelmann@psych.mpg.de);  
[n.kappelmann@gmail.com](mailto:n.kappelmann@gmail.com)

<sup>1</sup>Max Planck Institute of Psychiatry, Kraepelinstraße 2-10, 80804 Munich, Germany

<sup>2</sup>International Max Planck Research School for Translational Psychiatry (IMPRS-TP), Munich, Germany

Full list of author information is available at the end of the article



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

(Continued from previous page)

**Results:** The systematic review identified 38 RCTs for qualitative inclusion, 27 and 19 for quantitative inclusion at the sum-score level, and 9 and 4 for quantitative inclusion on individual symptom level for the HAM-D and BDI, respectively. Neither meta-analytic strategy revealed significant differences in the effectiveness of ADM and psychotherapy across the two depression measures. The SOrT metric did not show meaningful associations with other clinical variables in the MARS sample, and there was no indication of utility of the metric for better treatment allocation from PReDICT data.

**Conclusions:** This registered report showed no differences of ADM and psychotherapy for the treatment of MDD at sum-score and symptom levels. Symptom-based metrics such as the proposed SOrT metric do not inform allocation to these treatments, but predictive value of symptom information requires further testing for other treatment comparisons.

**Keywords:** Major depressive disorder, Depression symptoms, Antidepressant medication, Psychotherapy, Systematic review, Meta-analysis, Precision psychiatry, Symptom-oriented therapy metric

## Background

“Major depressive disorder (MDD) is a debilitating disease” is one of the most frequent introductory phrases in psychiatric literature and rightfully so (e.g. [1–5]). Its lifetime prevalence varies across countries between 5.5 and 21% [6, 7], and it is estimated to be the second leading cause for years lived with disability [8], leaving no doubt about the importance and urgency of developing effective treatments. While research on potential new treatments such as esketamine and anti-inflammatory drugs is underway [9–11], antidepressant medication (ADM) and psychotherapy offer effective treatments for a majority of patients as shown in recent high-quality meta-analyses [12–14]. As evidence is mainly based on between-group comparisons (e.g. drug versus placebo), the specific patient-treatment match often remains unclear. Specifically, as patients react differently towards specific treatments in clinical practice, they often need to “cycle through” different types of ADM and/or psychotherapy to find the treatment that will eventually help them personally [15–17]. Consequently, there is an increasing awareness in psychiatric and psychotherapy research for the necessity of more personalised treatment approaches that align with Paul’s old but important question: “What works for whom?” [18].

Attempts have been made and are underway towards establishing such personalised psychiatric care. These cover a wide range of different approaches such as using big data sets and machine learning models to predict the MDD course from baseline self-reports [19, 20], explorative data-mining strategies in order to define decision trees for the treatment of depression [21], algorithm-based treatments associated with shorter treatment time [15], imaging-based functional connectivity indices for treatment selection [22], or statistical strategies to examine superiority between treatments depending on stratification variables [23, 24]. Among the latter is a promising attempt by DeRubeis and colleagues who developed the

Personalised Advantage Index (PAI) by re-analysing data from a randomised controlled trial (RCT) of cognitive behavioural therapy (CBT) versus ADM [25]. The PAI constitutes the predicted difference in how patients would have benefitted from the treatment they received to the treatment they did not receive. In order to estimate the PAI for each patient across treatments, the authors used regression-based models with prognostic and prescriptive (treatment-moderating) variables as predictor variables and depression sum-scores at study endpoint (8 weeks) as outcome.

Four more recent studies on the treatment of depression have also estimated the PAI in RCTs of sertraline versus placebo [26], cognitive therapy versus interpersonal therapy [27], CBT versus psychodynamic therapy [28], and continuation-phase cognitive therapy versus fluoxetine for relapse prevention [29]. Throughout these studies, findings indicate that patients randomised to their optimal treatment (as suggested by the PAI) had clinically significantly better improvements in depression symptoms. Variables moderating treatment effects differed between studies and ranged from sociodemographic factors over life events up to personality and specific problems, symptoms, temperament, and comorbid conditions [25–29]. Beyond the PAI and analyses of single RCTs, individual patient data (IPD) meta-analysis work is currently underway by Weitz and colleagues who are trying to identify treatment moderators that indicate a combination of ADM and psychotherapy as more effective than monotherapy [30]. These studies thus describe strategies of using pre-treatment variables to inform treatment allocation. Nevertheless, the range of different, hardly overlapping variables between different studies stresses the sample dependency of results. It further reveals the necessity of replication in a prospective design to determine the clinical utility of stratified treatment allocation.

Besides the lack of prospective studies, we argue that previous studies potentially miss out on making use of the heterogeneity present within the construct of MDD. Specifically, all aforementioned studies using the PAI defined reductions in sum-scores on depression scales as the singular outcome variable to indicate treatment efficacy. Yet, individual depressive symptoms load differently on overall psychosocial impairment [31], so that clinical severity might not be best expressed by summing over individual depression scores [32]. This phenomenon gets aggravated when considering how low the content overlap in symptoms is between prominent depression scales [33]. Additionally, symptom combinations in MDD, for instance as defined by the Diagnostic and Statistical Manual of Mental Disorders 5th edition (DSM-5), allow two patients to present with a completely diverging symptom profile when considering the diametrically opposed symptoms such as insomnia and hypersomnia [34] within that compound symptom. This is not only of concern theoretically but a recent descriptive study in 3703 patients has estimated that of 1030 unique symptom combinations the most frequent symptom combination was only present in 1.8% of patients [35].

Utilising heterogeneity in symptom expression could offer additional, complementary insights for personalised treatment allocation of main treatments for depression. Empirically, two recent investigations have attempted to use information on symptom expression to differentiate effects of main treatments in re-analyses of RCTs comparing ADM and psychotherapy. In the first study, the authors compared treatment effects on distinct symptom clusters and results showed that cognitive therapy was better than both ADM and placebo in improving atypical vegetative symptoms (weight gain/increased appetite and hypersomnia) while there was no difference in clusters of (i) mood, (ii) cognitive/suicidal, (iii) anxiety, and (iv) typical vegetative symptoms (weight loss/decreased appetite and insomnia) [36]. These analysed symptom clusters still rely on assumptions of common factors underlying MDD, however, which has been criticised in recent research [37] and is suggestive of a move towards individual symptom analyses. Here, another re-analysis of individual symptom data indicated that ADM was better at reducing suicidality than CBT [38]. This latter study, however, was potentially underpowered for individual symptom analyses, and meta-analytic strategies could offer a more thorough approach. Nonetheless, these studies constitute promising attempts at delineating more symptom-specific treatment effects and are aided by other findings that show, for instance, that ADM efficacy as compared to placebo is most pronounced for the “depressed mood” symptom specifically [39].

In the current investigation, we wanted to develop and test what we term a *Symptom-Oriented Therapy* (SO<sub>r</sub>T)

metric that aspires to quantify potential preference (or not) of ADM or psychotherapy similar to the PAI. Contrary to the PAI, however, the SO<sub>r</sub>T metric is not based on pre-treatment (moderating) variables but instead makes use of patients’ heterogeneity in symptom expression. Additionally, the SO<sub>r</sub>T metric is not based on re-analyses of individual RCT data but instead on effect size estimates obtained in meta-analyses across a number of RCTs to avoid dependency of results on individual study samples. In particular, we wanted to conduct a systematic review and meta-analysis with the aim of synthesising existing evidence from RCTs comparing ADM and psychotherapy in the treatment of depression. While we also aimed to update existing meta-analytic evidence of these treatments on depression sum-scores, a previously used strategy [40–42], our primary aim was to analyse treatment efficacy on individual depressive symptoms. As prior RCTs have largely measured depressive symptoms using the Hamilton Depression Rating Scale (here, HAM-D; also commonly abbreviated as HRSD or HDRS) and/or Beck’s Depression Inventory (BDI) [40], we focussed our investigation on these two scales.

Following this overview of symptom-based treatment differences and similarities, we then aimed to compute the meta-analysis-based SO<sub>r</sub>T metric to indicate preference (or not) for a specific treatment type (i.e. ADM or psychotherapy) at the individual-patient level. This was evaluated using both data from an existing depressed in-patient sample, the Munich Antidepressant Response Signature study [43], and data from a previous RCT of ADM versus psychotherapy, the Emory Predictors of Remission in Depression to Individual and Combined Treatments (PReDICT) study [44]. Here, the primary test of our metric was to compare patients allocated to optimal versus non-optimal treatment in the PReDICT study as defined using the SO<sub>r</sub>T metric. Our primary hypothesis was that patients receiving their optimal treatment had significantly better outcomes on depressive symptoms than those receiving non-optimal treatment. We hoped that developing the SO<sub>r</sub>T metric based on our results would be particularly useful for future researchers (and ultimately clinicians) to test and individualise treatments for patients suffering from MDD. In this way, our results could provide a more refined view on heterogeneity of MDD and hopefully move the field closer towards personalised medicine.

## Methods

### Step 1: Systematic review and meta-analysis of RCTs of ADM versus psychotherapy

#### Protocol registration

The protocol for this systematic review was registered on PROSPERO (identifier: CRD42019123905) [45].

### Search strategy and study selection

We systematically searched PubMed Medline, PsycINFO, and Cochrane Central Register of Controlled Trials (CENTRAL) databases for randomised controlled trials (RCTs) of psychotherapy versus ADM in the treatment of depression. The search terms are presented in detail in Additional file 1: Table S1 in respective database grammar and align with the following search terms for PubMed Medline:

(depression[MeSH Terms] OR depressive disorder[-MeSH Terms] OR mood disorder[MeSH Terms] OR affective\* OR depress\*) AND (psychotherapy [mh] OR (Cogniti\* AND (technique\* OR therap\* OR restructur\* OR challeng\*)) AND (Antidepressive Agents [Pharmacological Action] OR agents, antidepressive[MeSH Terms] OR SSRI OR SNRI OR TCA OR “selective serotonin reuptake inhibitor” OR “selective norepinephrine inhibitor” OR “tricyclic antidepressant”) AND (randomized controlled trial[pt] OR controlled clinical trial[pt] OR randomized [tiab] OR randomly [tiab] OR trial [tiab] OR groups [tiab])

In addition to search of databases, we searched 352 RCTs mentioned on <http://www.evidencebasedpsychotherapies.org/>, a database created for psychotherapy RCTs and comparative trials [46]; hand-searched reference lists of retrieved RCTs and conference abstracts; searched for and, if necessary, contacted authors with published trial protocols; and wrote to prominent authors for unpublished data.

NK and JKB screened titles and abstracts of articles for the following inclusion and exclusion criteria: RCTs were required to (i) have at least one arm each for (individual and/or group) psychotherapy and ADM during part of the trial (e.g. crossover studies were allowed if data can be extracted before crossover), (ii) measure depressive symptoms using the Beck Depression Inventory (BDI) or Hamilton Depression Rating Scale (HAM-D), (iii) investigate adult depression, (iv) investigate major depression as a primary diagnosis without major medical comorbidity, (v) describe ethical approval and ascertainment of written informed consent, and (vi) include patients aged 18–75 years. For studies in languages other than English, we decided on a case-by-case basis whether we had the resources for translation of articles, which did not lead to further exclusions.

### Risk of bias assessment

In order to assess the quality of RCTs and risk of bias, we evaluated included studies using the gold standard Cochrane Risk of Bias tool [47] while making specific adaptations of the tool to the context of psychotherapy research as was recently suggested by Munder and Barth [48]. The Cochrane Risk of Bias tool includes assessments of (i) random sequence generation, (ii) allocation

concealment, (iii) blinding of participants and personnel, (iv) blinding of outcome assessment, (v) incomplete outcome data, (vi) selective reporting, and (vii) other biases. As blinding of participants and personnel is impossible in ADM versus psychotherapy comparisons, we assessed how differential expectations of patients and personnel about treatment were handled. Here, we focussed on bias being introduced by way of the study design, because wait-list control and combined treatment (i.e. ADM and psychotherapy) might bias participants towards expecting less or more treatment effects than single treatments (ADM or psychotherapy), respectively. Additionally, we assessed “bias due to deviations from intended interventions” (e.g. changes in treatment adherence or integrity), which has recently been added to the Cochrane Risk of Bias tool 2.0 and is particularly relevant for complex interventions like psychotherapy [48, 49]. First, NK and JKB independently assessed included studies for their risk of bias. In a second step, assessments were compared and inconsistencies discussed and resolved.

### Data acquisition and extraction

The main outcome for systematic review and meta-analyses were individual *symptom* data. Of note, individual *symptom* data—as required for our analyses—are not equal to individual *patient* data (IPD). While IPD have one row per participant, the format we required for our analyses is similar to frequency tables per symptom per treatment group; that is, one row per symptom, one column per symptom severity increment (e.g. 0, 1, 2, and 3 for all 21 BDI items) per treatment group, and cells indicating the number of patients indicating this symptom severity at end of treatment.

As depression outcome is usually not reported on the described individual *symptom* level, which was, however, required for our analyses, we only expected to be able to obtain a subset of data through extraction from manuscripts. The extracted variables from manuscripts were (i) study design and population, (ii) type and dosage of psychotherapy and ADM, (iii) study duration and follow-up, (iv) depression questionnaires used, (v) sample sizes of treatment arms at baseline and study endpoint, (vi) means and standard deviations of depression (sum-scores) at least at baseline and study endpoint (BDI and/or HAM-D), (vii) comorbidities (axis I and II disorders), (viii) age and sex of study participants, (ix) inclusion and exclusion criteria, (x) handling of missing data, dropouts, and use of intention-to-treat analysis, and (xi) researcher allegiance (i.e. main investigators’ training). NK and JKB independently extracted these data and subsequently discussed and resolved inconsistencies. If relevant data were only present in graph format and we did not get a response from authors, data were extracted using a reliable software tool [50].

Contacting authors for data (esp. individual *symptom* data) was also done in a standardised manner to maximise available data for this investigation (Additional file 2).

### Statistical analysis

The effect size for meta-analysis of depression is usually the standardised mean difference (SMD) when using sum-scores of depression scales. Although we wanted to provide an updated quantitative synthesis of depression sum-scores in the comparison of psychotherapy and pharmacotherapy using SMDs, our primary aim was to evaluate treatment differences at the individual symptom level. The SMD, however, does not address the potential non-normality of individual symptom items, which are on an ordinal but not necessarily interval scale. To that end, we have chosen to calculate proportional odds ratios (pORs) as effect sizes of first choice; these are appropriate for ordinal data if the proportionality assumption is met (i.e. if steps in symptom severity are similar) [51, 52]. pORs might be thought of as average ORs of all possible item severity steps (e.g. from 0 to 1, from 1 to 2, etc.).

In addition to pORs, we also presented analyses based on SMDs and "normal" ORs (after median split of symptom items) as sensitivity analysis. While SMDs and normal ORs are not ideal for this type of meta-analysis due to potentially violated assumptions and loss of power, respectively, they are more commonly used than pORs, so might offer additional insights to readers and highlight potential method dependency and robustness of the results. To allow comparability between metrics, we converted between (p)ORs and SMDs (referred to in results as converted (p)ORs and converted SMDs) according to the formula  $SMD = \ln(OR)/1.81$  [53]. As BDI and HAM-D have partially overlapping symptoms, we also performed a "spill over" sensitivity analysis by comparing individual symptom effect sizes between these scales for all comparable items. For instance, meta-analyses of "feelings of guilt" items in BDI and HAM-D should show comparable effect sizes towards ADM, psychotherapy, or neither. Additional file 3: Tables S2-S3 highlight scale differences and similarities on the item level, which authors NK and JKB have evaluated with reference to a prior content analysis of depression scales [33]. In addition to BDI and HAM-D comparability, we also expected differences in the BDI, depending on whether BDI-I or BDI-II was used. Comparable items of BDI-I and BDI-II were aggregated if possible but separated if necessary (Additional file 3: Table S5).

Meta-analyses were conducted using the *metafor* package [54] in *R* [55]. For meta-analyses, effects were weighted based on study sample size (i.e. inverse variance method) and we used a random-effects approach. Of note, we only analysed available data at the study

endpoint (i.e. completer data) as intention-to-treat approaches for individual symptom items would have made analyses too complex and we were not aware of any specific procedures for individual symptom items. We hoped, however, that missing data at study endpoint (e.g. through dropout) was less of a problem for our comparison of two active treatments, which should have had relatively more similar levels of dropout as compared to, for instance, studies with wait-list control arm.

We investigated heterogeneity between studies using Cochrane's *Q* and the  $I^2$  statistic. Publication bias was tested for meta-analysis of the sum-score depression outcome by visual funnel plot inspection and using Egger's test [56].

## Step 2: Development and validation of the Symptom-Oriented Therapy (SO<sub>r</sub>T) metric

### Computation of the SO<sub>r</sub>T metric

If our meta-analyses of individual depressive symptoms demonstrated treatment differences between ADM and psychotherapy, this would have potential benefit for the development of individualised treatment. To that end, we wished to provide researchers with a meta-analysis-based tool to compute what we term *Symptom-Oriented Therapy* (SO<sub>r</sub>T) metric. This metric is based on the quantitative results from meta-analyses, which served as weightings for patients' individual depressive symptoms (on either HAM-D or BDI). Specifically, computation of the SO<sub>r</sub>T metric followed the formula  $SO_{rT} = \sum_i m_i s_i$ , where *m* equals the meta-analytic effect size (favouring ADM or psychotherapy) as (converted) SMD, *s* equals the symptom score on BDI or HAM-D, and *i* equals a specific symptom item. Of note, *m* is defined as converted SMD and not pOR because SMDs follow a linear scale centred at zero, which allows more feasible computation and interpretation. See Additional file 4 for a detailed rationale and discussion of the SO<sub>r</sub>T metric, potential extensions of the formula, and hypothetical computation examples (Additional file 4: Table S5).

### Validation in MARS study

We computed SO<sub>r</sub>T scores for patients who took part in the MARS study, a subset of which has data available on both the BDI and HAM-D measures and was targeted in this validation step (*n* = 407). The MARS project was a naturalistic inpatient study of MDD patients conducted between years 2000 and 2015 at the Max Planck Institute of Psychiatry (Munich, Germany), the Bezirkskrankenhaus Augsburg (Germany), and the Klinikum Ingolstadt (Germany) [43]. Participating patients received various psychological assessments, and genetic and neuroendocrine measures were obtained with the goal of identifying (i) drug response predictors and (ii) subgroups of patients with similar biological pathophysiology.

We (i) provided descriptive statistics of patients' SOrT scores for BDI and HAM-D based on the MARS psychiatric inpatient sample, (ii) compared these on important clinical and demographic characteristics, and, most importantly, (iii) cross-validated the SOrT metric for BDI and HAM-D. This cross-validation was performed by computing (i) correlations between SOrT scores of scales and (ii) differences of BDI and HAM-D SOrT scores on an individual person level, respectively. Based on this, we were able to determine potential scale dependency or independence of the SOrT metric, which, in turn, provided an indication of its value as a tool for further research.

#### **Validation in PReDICT study**

We computed SOrT scores for patients who took part and were randomised in the PReDICT study ( $n = 344$ ) and performed analyses on all patients who completed the RCT ( $n = 234$ ). PReDICT was an RCT comparing CBT versus escitalopram versus duloxetine for the treatment of depression [44, 57]. For validation and in line with the setup of this investigation, ADM groups of escitalopram and duloxetine were pooled. SOrT score computation was done using meta-analytic effect estimates based on all retrieved studies *except* the PReDICT study, so that validation of the clinical utility of our metric was done independently from the meta-analytic "training" samples. Similar to the approach for the PAI used by DeRubeis et al. [25], we divided patients into those who received their optimal treatment (i.e. the SOrT score valence matches the treatment group) versus those who received their non-optimal treatment (i.e. the SOrT score valence does not match the treatment group). Doing this allowed us to compare outcome at end of treatment as suggested by HAM-D sum-scores of the optimal versus non-optimal groups and test for significant differences using a simple independent samples  $t$  test. In a second step, we followed the same procedure with a more restrictive subsample that was more extreme in their SOrT scores. Specifically, we used the two-thirds of patients with more extreme SOrT scores and compared optimal versus non-optimal treatment allocation groups. These procedures were applied both for SOrT scores created from BDI and for SOrT scores created from HAM-D. In sum, this allowed us to see whether the SOrT metric might pose a clinically significant benefit for treatment allocation.

#### **Timeline (steps 1 and 2)**

We describe the planned timeline for conductance of this registered report as well as timeline adherence in Additional file 5.

## **Results**

### **Step 1: Systematic review and meta-analysis of RCTs of ADM versus psychotherapy**

#### **Literature search**

The literature search was conducted on 31 January 2019 and revealed 4567 reports in total. Following duplicate removal, screening, and full-text assessments, we included 38 studies in our qualitative synthesis [57–94]. For our quantitative synthesis on the sum-score level, this was reduced to 27 studies with information on the HAM-D and 19 studies with information on the BDI. After corresponding with original authors to obtain individual symptom-level data, we were able to use 9 and 4 studies with HAM-D and BDI information, respectively, for our symptom-level meta-analyses. Of note, studies that reported data on both HAM-D and BDI were included in sum-score and symptom-specific analyses of both scales. Figure 1 shows the PRISMA flow diagram, and Additional file 6: Table S6 provides details of the included studies.

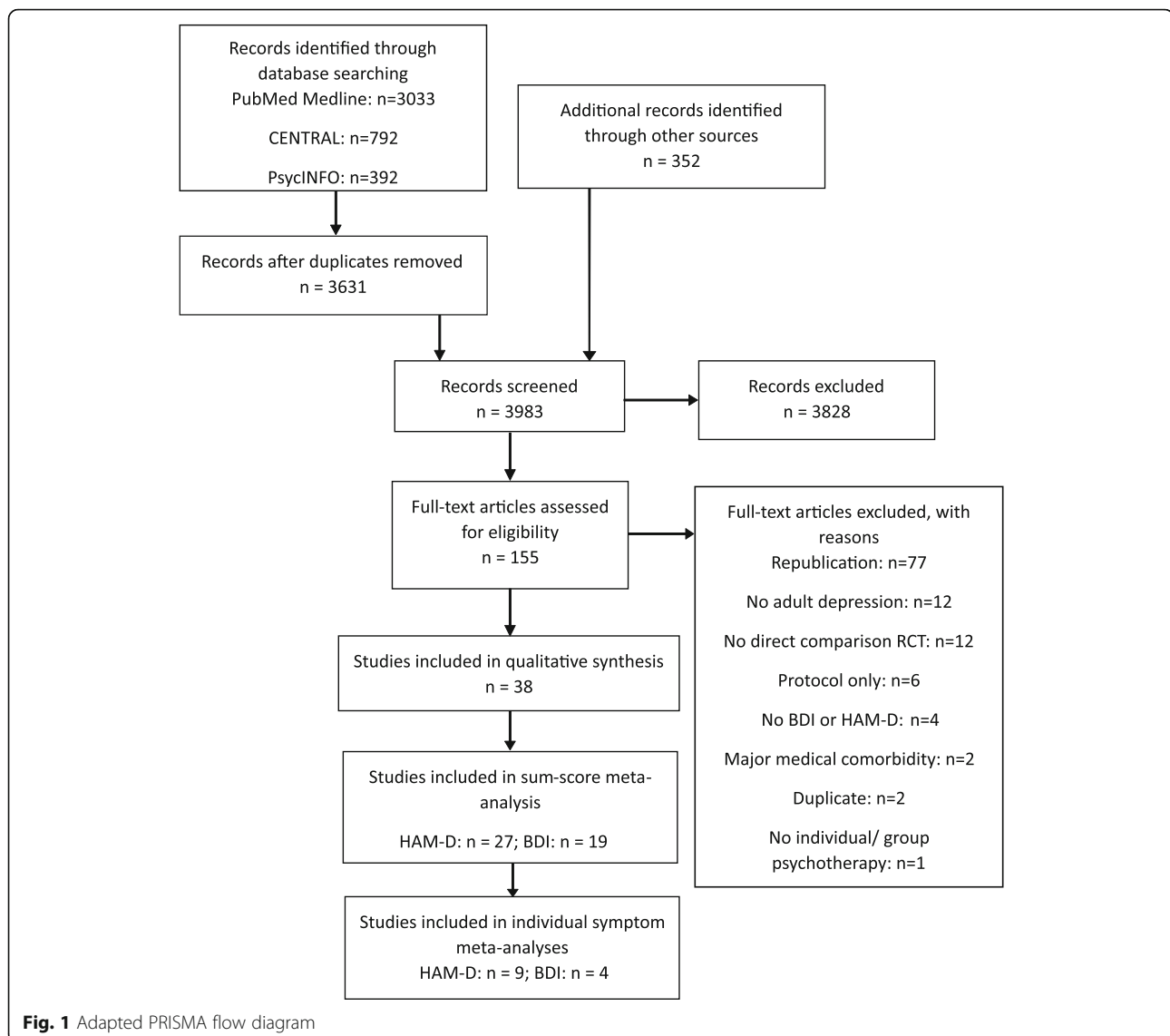
#### **Risk of bias**

Results from Cochrane Risk of Bias tool ratings by NK and JKB are presented in Fig. 2 for individual risk of bias categories. Most studies have high overall risk of bias or reason for some concerns. Considering that the overall risk of bias takes forward ratings of individual bias categories, however, it is unsurprising that individual bias categories are more diverse. In particular, low risk of bias arises from deviations from intended interventions while selection of reported results and randomisation process ambiguities were more concerning.

#### **Sum-score meta-analyses**

As previously done, we repeated sum-score meta-analyses for both HAM-D and BDI scales whenever outcome data were available in original studies. This allowed meta-analysis for 27 and 19 studies with 2433 and 1548 patients for HAM-D and BDI, respectively. Corresponding to completer only analyses on the individual symptom level, endpoint statistics for completers only were used in meta-analyses. We deviated from this procedure for 1 study, which only reported intention-to-treat statistics at endpoint [86]. Meta-analyses did not reveal significant differences in endpoint depressive symptom severity between psychotherapy and ADM for the HAM-D (SMD = 0.00, 95% CI -0.09–0.10; Cochrane's  $Q = 38.0$ , degrees of freedom [df] = 26,  $p = 0.060$ ;  $I^2 = 18.4\%$ ; Fig. 3) and the BDI (SMD = -0.05, 95% CI -0.24–0.15; Cochrane's  $Q = 61.1$ , df = 18,  $p < 0.001$ ;  $I^2 = 69.2\%$ ; Fig. 4).

We assessed publication bias for HAM-D and BDI sum-score meta-analyses by visual funnel plot inspection (Additional file 7: Fig. S1–S2) and Egger's test; neither



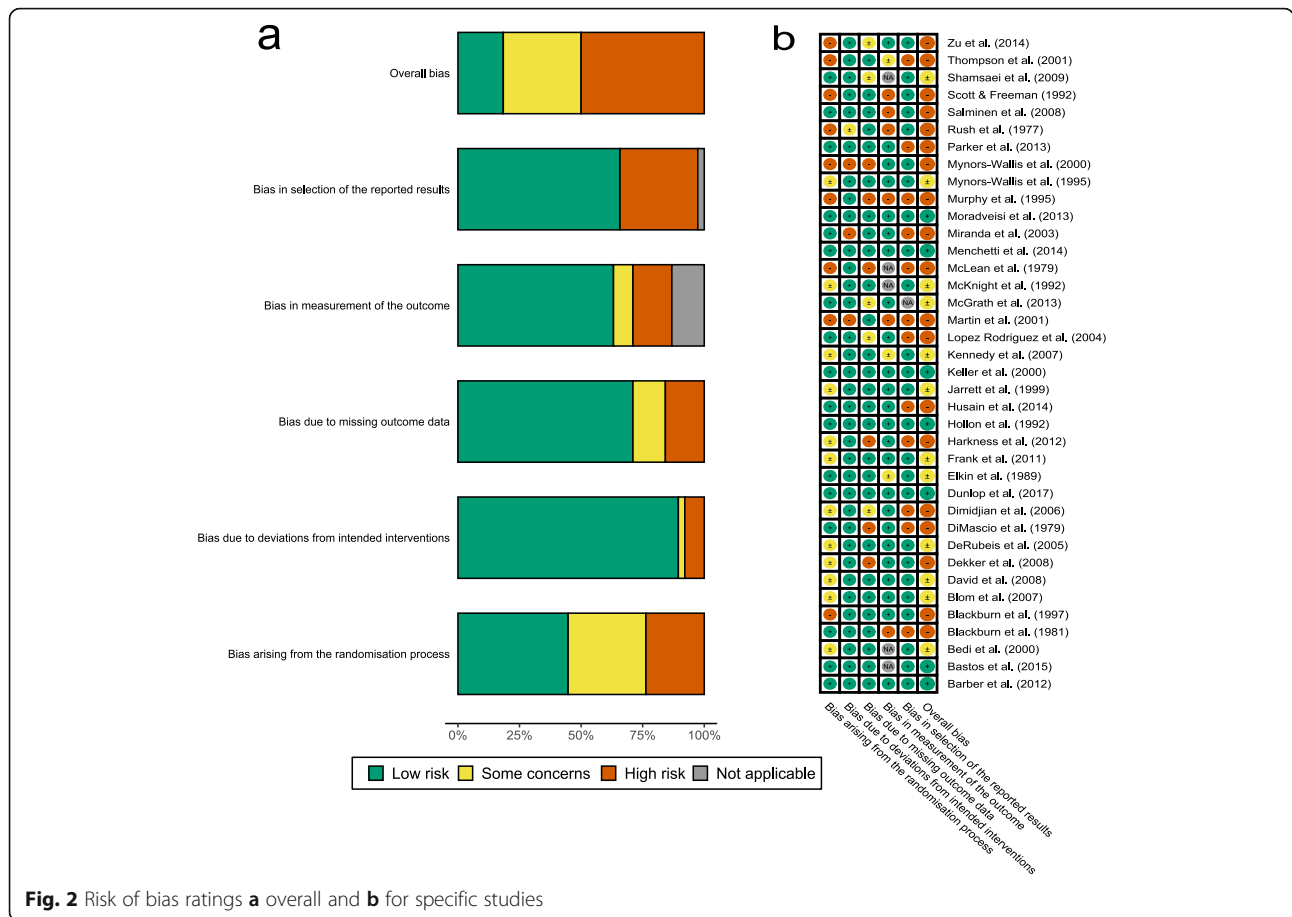
revealed indications of publication bias in favour of psychotherapy or ADM (HAM-D:  $z = -0.81$ ,  $p = 0.418$ ; BDI:  $z = -0.72$ ,  $p = 0.474$ ).

Beyond these pre-registered analyses, we further explored sum-score meta-analysis results by evaluating differential dropout between study arms. In particular, we performed meta-regression analyses including percentage of greater dropout in ADM compared to psychotherapy as moderator. These results revealed significant moderation of differential dropout for the HAM-D ( $p = 0.018$ ) but not for the BDI ( $p = 0.124$ ; Additional file 7: Table S7). Importantly, RCTs with greater dropout in their ADM arm(s) were those with more favourable effects for psychotherapy and vice versa. Additional file 7: Figs. S3-S4 visualise results for sum-score meta-analyses for HAM-D and BDI, respectively. We also explored differential dropout in a separate meta-analysis of 35

studies with information on baseline and endpoint sample sizes. Overall, 458 of 2163 (21.2%) and 557 of 2133 (26.1%) patients dropped out from psychotherapy and ADM arms, respectively. This difference was significant and indicated that psychotherapy was more acceptable to patients (OR = 0.74, 95% CI 0.59–0.92; Cochrane's  $Q = 57.7$ ,  $df = 34$ ,  $p = 0.007$ ;  $I^2 = 40.8\%$ ; see Additional file 7: Fig. S5 for the forest plot).

#### Individual symptom meta-analyses

We conducted individual symptom meta-analyses for 9 and 4 studies providing individual symptom data for HAM-D and BDI, respectively. Importantly, the number of studies and, correspondingly, the number of patients included in meta-analyses varied across symptoms with ranges of 421–1166 (median = 1166) and 379–502 (median = 501) completer patients for HAM-D and BDI



**Fig. 2** Risk of bias ratings **a** overall and **b** for specific studies

items, respectively. The reasons for differing sample sizes across symptoms varied for the following reasons: We had data from fewer studies assessing the BDI, questionnaire versions differed, and the lack of variance in a particular symptom of a particular study necessitated the removal of such a study from a meta-analysis. Regarding differing questionnaire versions, for instance, some studies used a HAM-D version with more than 17 items. For the BDI, the items differ in versions I and II. Aggregated pooled meta-analytic effect sizes can be seen in Fig. 5. Additional file 8: Tables S8-S9 include forest plots with sample sizes and heterogeneity statistics for each symptom of HAM-D and BDI, respectively. Additional file 8: Fig. S6 shows effect size comparisons of HAM-D and BDI symptoms per symptom category (symptom categories as defined in Additional file 3: Table S2). These comparisons indicated similar effect sizes between symptom categories of guilt, loss of energy, and suicidality while there was a strong divergence between symptom categories of work and interests and loss of libido.

As meta-analytic results indicate, few symptoms showed nominally significant differences between psychotherapy and ADM (favouring ADM: HAM-D item 4 [insomnia: early], BDI item 2 [pessimism] and BDI item

13 [indecisiveness]; favouring psychotherapy: BDI item 19 [concentration difficulty]). Importantly, after family-wise error correction (using the Benjamini-Hochberg method) [95], none of these differences remained significant (all  $p > 0.05$ ).

### Sensitivity analyses

To determine the influence of choice of effect size metric, we repeated individual symptom meta-analyses using “simple” ORs and SMDs in comparison to our metric of choice, the proportional OR (pOR). These sensitivity analyses provided a technical replication of our results indicating that choice of pOR as effect size metric was not driving our results (see Additional file 8 for details).

### Exploratory comparison with Boschloo et al. [96]

Following pre-acceptance of this registered report, Boschloo et al. published a highly similar report to this study as RCTs comparing CBT (as one form of psychotherapy) versus ADM were meta-analysed on the individual symptom level [96]. The HAM-D was used as the sole outcome rating scale, but SMDs between treatments



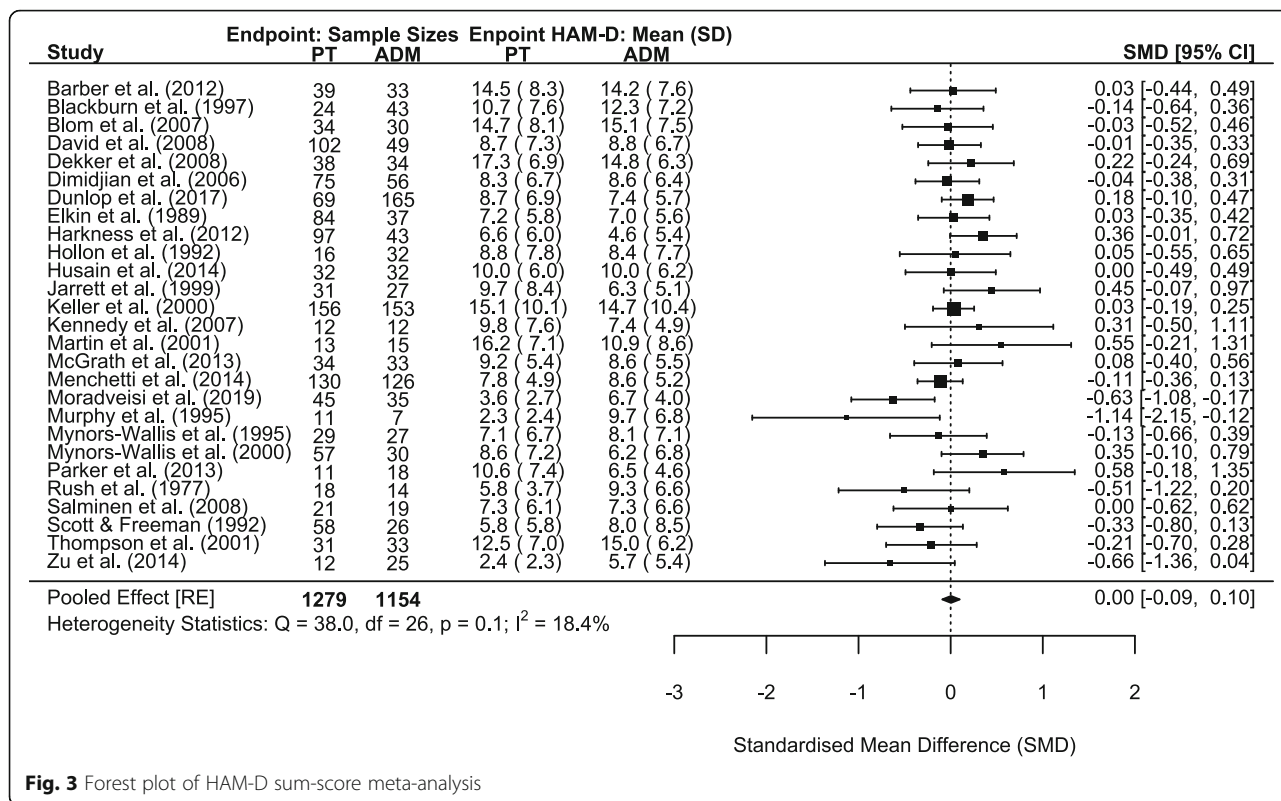


Fig. 3 Forest plot of HAM-D sum-score meta-analysis

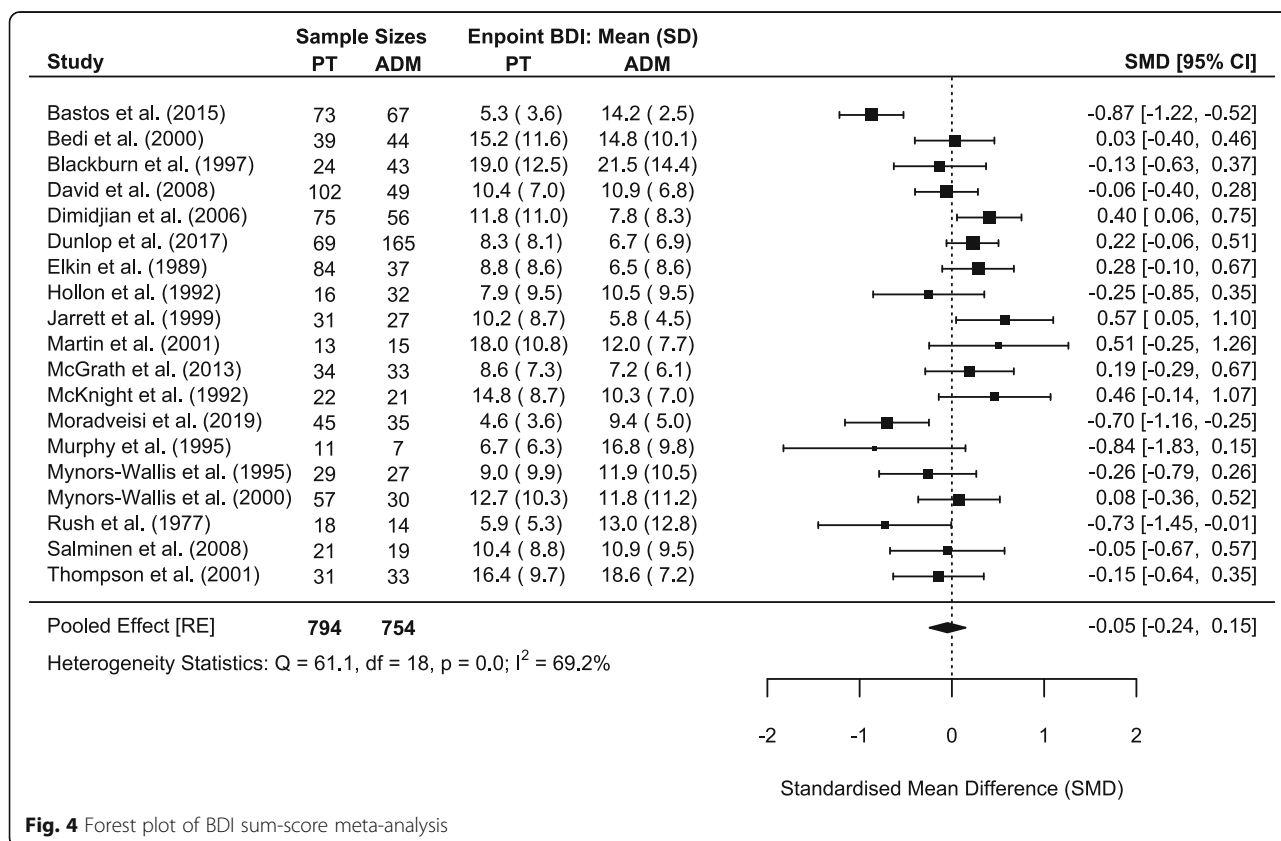
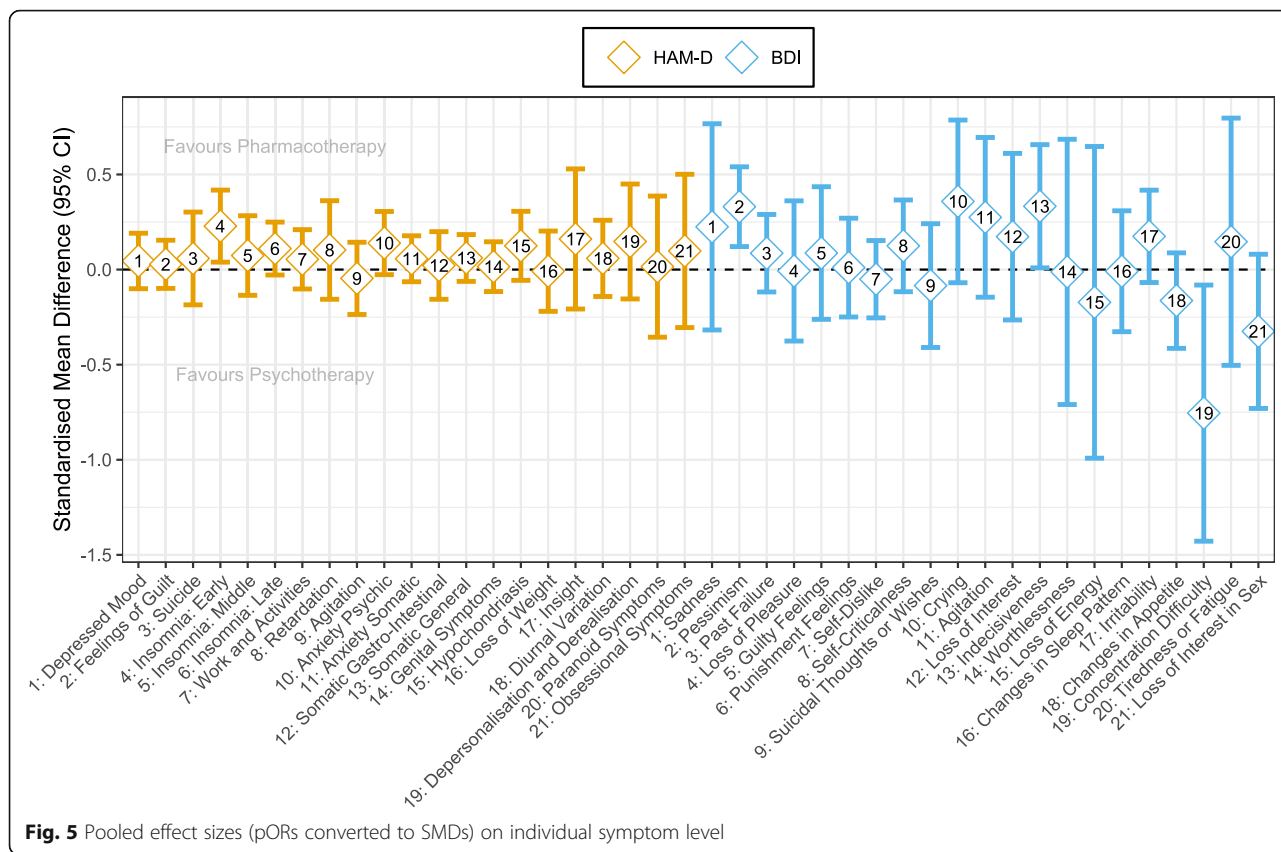


Fig. 4 Forest plot of BDI sum-score meta-analysis



were calculated as differences in symptom scores from baseline to study endpoint.

As we deemed it important for the field, we performed an exploratory comparison of our results with theirs. Details of this exploratory comparison are provided in Additional file 8. In sum, symptom-specific effect sizes from our meta-analyses were not associated with effect sizes from Boschloo et al. and only with small-to-moderate correlation when we restricted our meta-analysis to RCTs investigating CBT only.

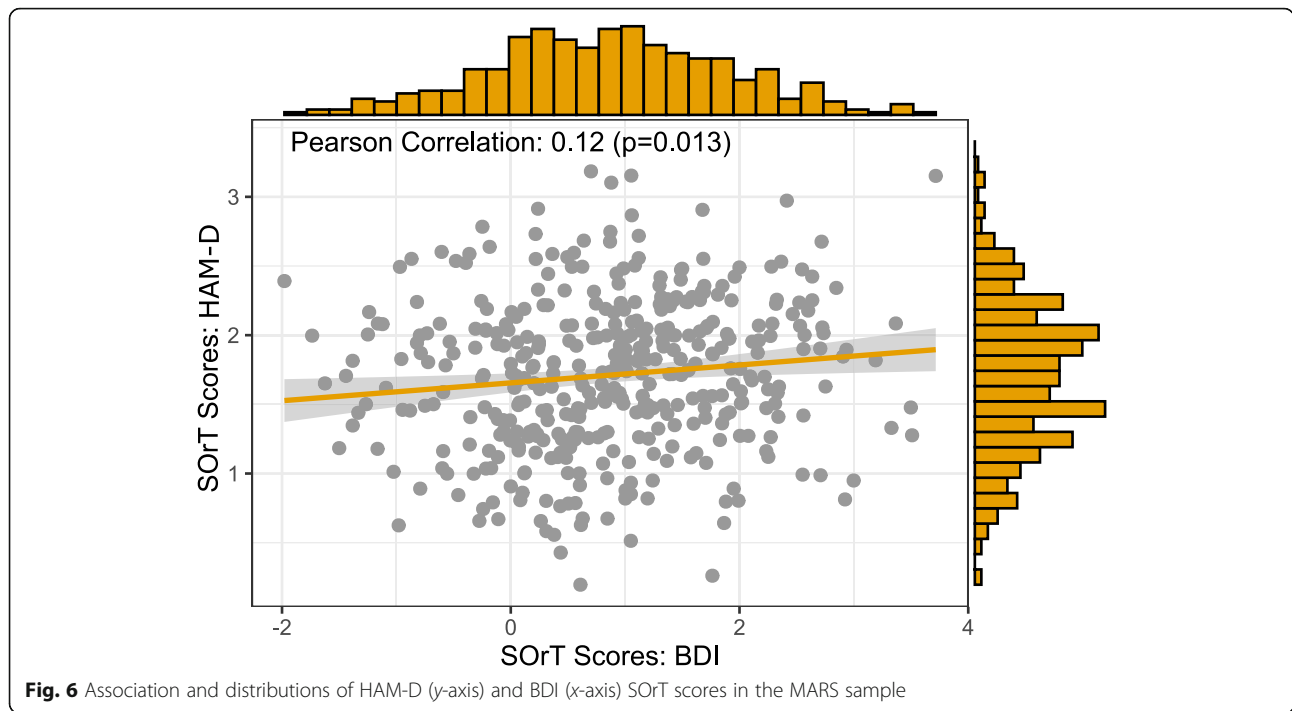
**Step 2: Development and validation of the Symptom-Oriented Therapy (SO<sub>r</sub>T) metric**

**Validation in MARS study**

We validated the SO<sub>r</sub>T metric using (i) a descriptive summary, (ii) assessment of clinical and demographic associations with our metric, and (iii) cross-validation between BDI- and HAM-D-based SO<sub>r</sub>T scores.

First, HAM-D and BDI SO<sub>r</sub>T metrics were approximately normally distributed in the sample (HAM-D: min = 0.20, median = 1.72, mean = 1.71, max = 3.18, SD = 0.55; BDI: min = -1.98, median = 0.87, mean = 0.86, max = 3.72, SD = 1.04; Fig. 6). Contrary to our expectations, however, SO<sub>r</sub>T metric distributions were not centred around 0 and, in case of HAM-D SO<sub>r</sub>T scores, did not even include 0. A likely reason for the

distributions not being centred around 0 is that most symptom-specific effect sizes were positive (i.e. favouring ADM) with 15/17 positive effect sizes for the HAM-D and 12/21 for the BDI. Second, we assessed clinical and sociodemographic correlates of respective SO<sub>r</sub>T scores. Importantly, we found high and small-to-moderate correlations between SO<sub>r</sub>T scores to their respective scale (HAM-D: Pearson’s  $r = 0.81, p < 0.001$ ; BDI: Pearson’s  $r = 0.35, p < 0.001$ ), which likely arises as an artefact of most effect sizes being positive. This also counters our initial goal of creating a treatment allocation metric (indicating psychotherapy versus ADM) as opposed to a mere measure of symptom severity. To highlight the distinction between a treatment allocation metric versus symptom severity associations, we report linear regression analyses, unadjusted and adjusted for baseline HAM-D and BDI sum-scores, of SO<sub>r</sub>T metric outcome on different sociodemographic and clinical predictor variables (Table 1). These analyses failed to reveal any consistent associations with our metric. Third, the small correlation of HAM-D and BDI SO<sub>r</sub>T scores (Pearson’s  $r = 0.12$ ; cf. Fig. 6) was below the level considered meaningful for two treatment allocation metrics with identical goals (i.e. indicating favourable treatment by psychotherapy or ADM). In sum, our SO<sub>r</sub>T metric did not seem to



be a valid measure with any clinical or sociodemographic correlates.

**Validation in PRedICT study**

The SOrT metric was further evaluated in data from the PRedICT trial [44, 57]. Again, SOrT scores were computed based on patients’ baseline symptom scores on

HAM-D and BDI. Importantly, however, SOrT scores were based on effect sizes of individual symptom meta-analyses excluding PRedICT data to ascertain independence of our validation data (see comparison of effect sizes in Additional file 9: Table S11). As with the MARS sample, distributions and correlations of HAM-D and BDI SOrT scores are visualised in Additional file 9: Fig.

**Table 1** Linear regression analyses of SOrT scores on sociodemographic and clinical predictor variables in the MARS sample

Predictor variable	HAM-D SOrT metric outcome				BDI SOrT metric outcome			
	Unadjusted		Adjusted <sup>†</sup>		Unadjusted		Adjusted <sup>†</sup>	
	Beta (SE)	p	Beta (SE)	p	Beta (SE)	p	Beta (SE)	p
<b>Sociodemographic</b>								
Female sex	0.125 (0.054)	0.021	0.044 (0.032)	0.167	-0.084 (0.103)	0.414	-0.173 (0.097)	0.075
Age	> 0.001 (0.002)	0.983	-0.001 (0.001)	0.631	0.007 (0.004)	0.074	0.008 (0.003)	0.027
Married	0.025 (0.055)	0.648	-0.045 (0.033)	0.176	-0.103 (0.105)	0.328	-0.035 (0.099)	0.725
Education: German A-levels	-0.136 (0.055)	0.014	-0.025 (0.034)	0.447	0.035 (0.106)	0.742	0.109 (0.100)	0.276
<b>Clinical</b>								
Baseline HAM-D sum-score	0.075 (0.003)	< 0.001	-	-	0.021 (0.009)	0.017	-0.002 (0.009)	0.815
Baseline BDI sum-score	0.019 (0.002)	< 0.001	0.004 (0.002)	0.022	0.036 (0.005)	< 0.001	-	-
Discharge HAM-D sum-score	0.011 (0.005)	0.026	0.002 (0.003)	0.590	0.014 (0.010)	0.152	0.006 (0.009)	0.560
Discharge BDI sum-score	0.006 (0.003)	0.087	0.002 (0.002)	0.443	0.011 (0.007)	0.087	-0.003 (0.007)	0.620
History of depression	-0.011 (0.069)	0.869	0.049 (0.039)	0.214	0.082 (0.130)	0.526	-0.069 (0.123)	0.576
Weeks of hospitalisation	0.010 (0.004)	0.009	< 0.001 (0.002)	0.829	0.010 (0.007)	0.173	> -0.001 (0.007)	0.891
Anxiety (HAMA)	0.030 (0.003)	< 0.001	-0.003 (0.003)	0.325	0.010 (0.006)	0.112	-0.001 (0.006)	0.817

Separate linear regression analyses were performed for each predictor variable. Model intercepts are not reported to simplify result interpretation. Binary predictor variables are termed in a way that makes dummy coding obvious (e.g. female sex: 0 = male, 1 = female). *Abbreviation:* HAMA Hamilton Anxiety Rating Scale.

<sup>†</sup>Models were adjusted for respective (HAM-D or BDI) baseline symptom covariate

S11. Contrary to our findings in MARS, we found a small, negative correlation between HAM-D and BDI SOrT scores (Pearson's  $r = -0.13$ ,  $p = 0.054$ ). Distributions were comparable for the HAM-D-based SOrT metric (min = 0.13, median = 1.15, mean = 1.15, max = 2.19, SD = 0.34) and more negative and variable for the BDI-based SOrT metric; this is consistent with a smaller number of trials included in BDI meta-analyses (min = -5.27, median = -0.98, mean = -1.04, max = 2.09, SD = 1.39).

To compare optimal versus non-optimal SOrT-based treatment allocation, we initially pre-registered a sample split based on SOrT score valence (i.e. positive scores indicating optimal allocation to ADM and vice versa). As SOrT score distributions highlight, however, this approach was only possible for the BDI (23% of patients with positive SOrT scores) but not for the HAM-D (0% of patients with positive SOrT scores). Consequently, we decided to report an additional, exploratory classification of "optimal" versus "non-optimal" treatment allocation based on a SOrT score median split (i.e. patients above the median optimally treated by ADM and vice versa). Pre-registered and exploratory comparisons are reported in Table 2. None of these analyses revealed any potential benefit of allocating patients to treatment based on their SOrT scores. We performed additional linear regression analyses of these comparisons, unadjusted and adjusted for the HAM-D at baseline, to delineate potential dependency of the SOrT metric on symptom severity (Additional file 9: Tables S12-S14). These results align with pre-registered  $t$  tests.

### Exploratory analyses

We conducted two further sets of exploratory analyses to compare our results with those by Boschloo et al. [96] and altering computation of the SOrT metric to  $SOrT = \frac{\sum m_i s_i}{\sum s_i}$ , so that there were no artifactual correlations with symptom severity (see Additional file 9 for details). Results suggested that symptom-based metrics

from Boschloo et al. [96] and with altered SOrT score computation did not offer reliable advantages for allocation to psychotherapy versus ADM.

### Discussion

This registered report outlines a detailed investigation of the comparative effectiveness of psychotherapy and ADM for the treatment of individual depressive symptoms and whether symptom-specific effectiveness information can serve precision allocation. We did not find ADM or psychotherapy to be more effective than the respective other treatment at study endpoints in depressive symptom sum-scores on HAM-D and BDI scales. Similarly, there was no clear advantage of either treatment for individual depressive symptoms. Using individual symptom meta-analysis results, we evaluated the *Symptom-Oriented Therapy* (SOrT) metric, which combines meta-analytic effect size estimates with patients' symptom profiles prior to treatment. Validation analyses in MARS and PRedICT studies did not indicate that the SOrT metric constituted a valid measure, nor that it should be used as a precision metric to indicate favourable allocation to psychotherapy versus ADM.

Our findings of comparable effectiveness of psychotherapy and ADM at the end of acute treatment closely align with older [40–42] and more recent [97] meta-analytic work. Cuijpers and colleagues only recently conducted an extensive network meta-analysis quantifying and ranking multiple different treatments for adult depression, including psychotherapy, ADM, and their combination [97, 98]: Across multiple sets of sensitivity analyses (e.g. high versus low risk of bias, optimised psychotherapy/ADM, excluding placebo-controlled studies) and looking at groups of studies with moderate, severe, and chronic depression, psychotherapy and ADM consistently showed comparable effectiveness. Yet, psychotherapy was more acceptable (as defined in terms of lower dropout rates) compared to ADM. It is reassuring that our findings, albeit with smaller number of included

**Table 2** Evaluation of SOrT-based treatment allocation

Method	Pre-registered	12-week HAM-D, mean (SD)		$p$ value <sup>†</sup>
		Optimal	Non-Optimal	
HAM-D-based SOrT allocation				
Valence	Yes	–	–	–
Median split	No	8.50 (6.41)	7.17 (5.86)	0.091
Extreme groups	Yes	8.86 (6.65)	7.08 (5.85)	0.074
BDI-based SOrT allocation				
Valence	Yes	7.62 (6.55)	8.03 (5.97)	0.617
Median split	No	7.63 (6.14)	8.09 (6.24)	0.563
Extreme groups	Yes	7.01 (6.16)	7.66 (6.15)	0.510

<sup>†</sup> $p$  values are based on independent samples  $t$  tests

studies, fully replicate this report. Our exploratory meta-regression of the HAM-D further demonstrated that effectiveness estimates of specific studies were moderated by differential dropout in study arms. Accordingly, if only ADM or psychotherapy is available to patients (rather than the combination, which is most effective [97]), we agree with clinical recommendations by Cuijpers and colleagues in that psychotherapy would be preferable to ADM based on its greater acceptability. This indication may be particularly relevant for patients who are likely not to complete a full course of treatment and prone to dropout (e.g. younger patients [99]).

Symptom-specific meta-analyses did not show significant differential treatment effects of psychotherapy or ADM for specific depressive symptoms following multiple comparison corrections. Although we identified nominally significant treatment differences for some symptoms, these do not align with prior research findings [36, 38, 96]. Hence, it remains unclear whether true symptom-specific treatment differences exist or whether reports from our study and previous literature reflect false positive findings (see Additional file 10 for further discussion).

The absence of clear meta-analytic symptom differences may also explain why our proposed SOrT metric did not seem to offer any benefit for treatment allocation. Assuming meta-analytic effect sizes were representative of noise rather than signal, this would mean SOrT scores were merely superimposing (meta-analytic) noise onto patients' baseline symptom profiles. Our failure to find predictive utility of the precision sum-score from Boschloo and colleagues [96] and of a SOrT metric based on their reported effect sizes adds to this discouraging conclusion that symptom scores may not be of value in predicting psychotherapy versus ADM response.

It is important to emphasise, however, that the SOrT metric and other symptom-based allocation metrics (cf. [96, 100]) are likely specific to the comparison being made, so are restricted to the comparison of psychotherapy versus ADM in this report. Khazanov and colleagues, for instance, recently showed that patients with greater distress and anhedonia prior to treatment responded more favourably to a combination of cognitive therapy and ADM rather than to ADM alone. This suggests individual symptom information may be valuable for indicating combination versus monotherapy in depression [100]. Similarly, there are reports that low-grade inflammation shows specificity for somatic symptoms of depression (e.g. sleep problems, low energy, or increased appetite) [101–104], so ongoing studies evaluating immunotherapy as a treatment for depression may benefit from symptom-based treatment allocation rules. We therefore encourage researchers to apply the SOrT metric to other treatment comparisons to evaluate its

value as a precision medicine tool. It could be helpful, therefore, to adjust SOrT scores for symptom severity as we highlighted in our exploratory analyses (cf. Additional file 9). Moreover, future research could evaluate a combination of symptom-weighted approaches (such as the SOrT metric) with significance thresholds. Specifically, symptom-based metrics might benefit from only including meta-analytic weights (and corresponding symptoms) that pass a specific meta-analytic significance threshold for the comparison in question, thus optimising the signal-to-noise ratio. It is interesting to note that this approach would be similar to how polygenic risk scores are created (e.g. using PRSice software [105]), which weight effect alleles by effect sizes, if a specific significance threshold has been reached for this allele. In sum, we hope symptom-based precision medicine tools receive further attention in future research despite the limited value for decision-making on allocation to psychotherapy versus ADM in depression.

#### Limitations

The present investigation has five major limitations. First, we were not able to include all studies identified from the literature in meta-analyses. This low coverage was due to varied reasons such as general data unavailability [59, 62, 63, 86, 89, 90], insufficient reporting of sample sizes and summary statistics for inclusion in sum-score meta-analysis [68, 75, 85, 94], and/or failure of authors to respond to inquiries (cf. Additional file 11: Table S19). Based on this, we encourage more thorough reporting of summary statistics in original trials (ideally including completer and intention-to-treat samples) and also advocate publication of summary statistics in meta-analyses rather than effect sizes alone; this would enhance reproducibility. Despite limited data availability, it is reassuring that conclusions from sum-score meta-analyses match the most recent meta-analysis with the largest power [97]. Sample sizes for individual symptom meta-analyses were also relatively large and exceeded those of similar work by Boschloo et al. [96], which can be attributed to our broader focus on RCTs comparing ADM with psychotherapy (rather than to CBT only).

Second, we cannot conclude anything regarding symptom-specific differences between intervention comparisons other than psychotherapy versus ADM (as discussed before), and it is unclear if conclusions outside this study's inclusion criteria would be meaningful. Our inclusion criteria, for instance, focussed on acute treatment of adult depression, so it is possible that symptom-specific differences following treatment with ADM versus psychotherapy exist in children, adolescents, or older patients. Similarly, symptom-specific differences may arise at follow-up in the form of residual symptoms (also

not addressed in this report), which was recently shown in a RCT re-analysis [38].

Third, our analyses focussed on study completer data, though an intention-to-treat approach would be clinically more interesting. The data format used for present analyses did not allow for the identification of individual participants, which facilitated data sharing from original RCTs. Due to non-identifiability of individuals, however, we were not able to use imputation techniques for intention-to-treat analyses on the symptom level. Consequently, we could not verify whether our findings would have been different if all randomised participants had been included, rendering this an open question for future research.

Fourth, inferences from symptom-specific analyses are limited by the low reliability of individual symptoms. Depression measures, such as the HAM-D and BDI used in the present report, are not designed for reliable symptom-specific assessments [32, 33]. Thus, our failure to find evidence for symptom-specific treatment effects and/or differences between present findings and Boschloo et al. [96] may be a consequence of low reliability of symptom assessments. Future research should ideally combine evidence from multiple indicators per assessed symptom to increase reliability.

Lastly, validation of the SOrT metric was only performed in data from the PRoDICT study. Because the PRoDICT study enrolled only treatment-naïve patients [44, 57], it may not have provided the optimal sample for validating the SOrT metric, which was derived from more mixed samples. Future research should thus evaluate the utility of symptom-based metrics in other samples, which would require individual patient data. Ideally, cross-validation procedures could be used, which would average over unique characteristics of multiple validation samples.

## Conclusion

In conclusion, we report the largest symptom-specific meta-analysis of direct comparisons of psychotherapy and ADM for depression. We did not find robust indications for symptom-specific effectiveness differences between treatments and this also extended to the sum-score level. We introduced the Symptom-Oriented Therapy (SOrT) metric as a precision treatment allocation tool, but failed to demonstrate its usefulness for improved allocation to psychotherapy or ADM. Though future symptom-specific work looking at other treatment comparisons could be valuable, our findings suggest that symptom information does not inform on whether any individual patient should receive psychotherapy or ADM.

## Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12916-020-01623-9>.

**Additional file 1: Table S1.** Search Strategy for Identification of RCTs of Psychotherapy versus Pharmacotherapy for Depression.

**Additional file 2.** Standardised author contacting and quality evaluation procedure.

**Additional file 3 Tables S2-S4.** Content overlap assessment of Beck Depression Inventory (BDI) and Hamilton Rating Scale for Depression (HAM-D). Table S2- Content Overlap of HAM-D and BDI-II Items Sorted by Item Numbering. Table S3- Content Overlap of HAM-D and BDI-II Items Sorted by Equivalent Items. Table S4- Content Overlap of BDI-I and BDI-II Items Sorted by Item Numbering.

**Additional file 4: Table S5.** Development and considerations of a Symptom-Oriented Therapy (SOrT) metric. Table S5- Hypothetical SOrT Metric Computation for Three Patients.

**Additional file 5.** Timeline and timeline adherence of the registered report.

**Additional file 6: Table S6.** Study overview of RCTs included in qualitative synthesis.

**Additional file 7: Table S7, Figs. S1-S5.** Sum-score meta-analysis results. Table S7- Meta-regression of HAM-D and BDI sum-score meta-analyses on differential dropout. Fig. S1- Funnel plot of HAM-D sum-score meta-analysis. Fig. S2- Funnel plot of BDI sum-score meta-analysis. Fig. S3- Meta-Regression of HAM-D sum-score meta-analysis on differential dropout. Fig. S4- Meta-Regression of BDI sum-score meta-analysis on differential dropout. Fig. S5- Funnel plot of HAM-D sum-score meta-analysis. Fig. S6- Forest plot of dropout meta-analysis.

**Additional file 8: Tables S8-S10, Figs. S6-S10.** Individual symptom meta-analysis results and related sensitivity and exploratory analyses. Table S8- Individual symptom forest plots for the HAM-D. Table S9- Individual symptom forest plots for the BDI. Table S10- Correlations between meta-analytic effect size metrics. Fig. S6- Effect size comparison of HAM-D and BDI per symptom type. Fig. S7- Individual symptom effect sizes per effect size metric used in meta-analyses. Fig. S8- Associations between meta-analytic effect size metrics. Fig. S9- Effect size association to Boschloo et al. Fig. S10- Effect size association to Boschloo et al. for RCTs with CBT only.

**Additional file 9: Tables S11-S18, Fig. S11.** Validation analyses of SOrT metric in MARS and PRoDICT samples and related exploratory analyses. Table S11- Effect sizes of individual symptom meta-analyses with versus without Dunlop et al. Table S12- Linear regression analysis of 12-week HAM-D sum-scores on SOrT-based BDI treatment allocation match following valence split. Table S13- Linear regression analysis of 12-week HAM-D sum-scores on SOrT-based treatment allocation match following median split. Table S14- Linear regression analysis of 12-week HAM-D sum-scores on SOrT-based treatment allocation match following 2/3 extreme group split. Table S15- Treatment allocation match prediction using Boschloo et al.-based SOrT and sum-scores. Table S16- Regression-based treatment allocation match prediction using Boschloo et al.-based SOrT and sum-scores. Table S17- Association of symptom severity with updated SOrT scores. Table S18- Treatment allocation match prediction using updated SOrT metric. Fig. S11- Association and distributions of HAM-D and BDI SOrT scores in PRoDICT.

**Additional file 10.** Discussion on nominally significant treatment differences of psychotherapy and ADM for specific depressive symptoms.

**Additional file 11: Tables S19-S20.** Overview of available data and materials from original studies and presented work. Table S19 Availability of Original Study Data. Table S20. Online File Overview as Available on <https://osf.io/qzjc9/>.

## Abbreviations

ADM: Antidepressant medication; BDI: Beck Depression Inventory; CBT: Cognitive behavioural therapy; CENTRAL: Cochrane Central Register of Controlled Trials; CI: Confidence interval; HAMA: Hamilton Anxiety Rating Scale; HAM-D: Hamilton Rating Scale for Depression; IPD: Individual patient

data; MARS: Munich Antidepressant Response Signature; MDD: Major Depressive Disorder; PAI: Personalised Advantage Index; PReDICT: Emory Predictors of Remission in Depression to Individual and Combined Treatments; (p)OR: (Proportional) Odds ratio; RCT: Randomised controlled trial; SD: Standard deviation; SE: Standard error; SMD: Standardised mean difference; SoRT: Symptom-Oriented Therapy

### Acknowledgements

We are grateful to all original authors, research groups, funders, and participants of original RCTs comparing psychotherapy and ADM in depression that constituted the foundation for this investigation. We specifically thank the National Institute of Mental Health (NIMH) as sponsor of the Barber et al. (MH 061410 [JP Barber]), Jarrett et al. (MH-45043 & MH-38238 [RB Jarrett]), and McGrath et al. (R01 MH073719 [HS Mayberg], T32 GM08695 [CL McGrath], K23 MH086690 [BW Dunlop], & K23 MH077869 [PE Holtzheimer]) studies, as well as studies funded by the Australian National Health and Medical Research Council (#1037196 and #GNT1176689 [G Parker]).

We also thank Josefine Moultrie for her support in developing and piloting the search strategy and Britta Dumser for helping with the comparison of BDI and HAM-D symptoms and support on the Cochrane Risk of Bias tool. Lastly, we thank Marc Volkert for sharing his thoughts on the choice of effect size metric for meta-analysis and formulation of the SoRT metric.

### Authors' contributions

NK had the initial idea for the project and its conceptualisation. NK drafted the manuscript with supervision and support from JKB. NK and JKB both performed the literature search, data extraction, and risk of bias ratings. NK was responsible for all statistical analyses and figures. NK and MR were responsible for the search strategy and timeline. JF was responsible for IPD and data management sections. HSM, WEC, BWD, CBN, MK, DNK, BAA, NH, RBJ, JRV, MM, GP, JPB, AGB, JD, and JP were responsible for providing and sharing expertise on their RCT data and related analyses. NK and JKB assume responsibility for the accuracy and integrity of this work. All authors critically reviewed the manuscript and have given approval to the final version of the manuscript.

### Funding

This project is funded by the Max Planck Institute of Psychiatry. The MARS project was supported by grants of the German Federal Ministry of Education and Research (BMBF), project no. 01ES0811 (Molecular Diagnostics) and 01EE1401D (German Research Network for Mental Disorders). PReDICT study was supported by the following National Institutes of Health grants: P50 MH077083; R01 MH080880; UL1 RR025008; M01 RR0039; and the Fuqua family foundations. Forest Labs and Elli Lilly Inc. donated the study medications, escitalopram and duloxetine, respectively, and were otherwise uninvolved in the study design, data collection, data analysis, or interpretation of findings.

### Availability of data and materials

For maximum transparency and reproducibility, we attempt to provide as much data and materials as possible. In Additional file 11: Table S19, we outline to what capacity we can share data from original RCTs identified in the systematic review. In Additional file 11: Table S20, we provide a list of additional data and analysis scripts as available on the open science framework (OSF) platform under <https://osf.io/qzjc9/>. We are not able to publicly share original data from the MARS study. Original data from PReDICT, however, is available as required by NIH data sharing policy and can be requested by contacting original authors.

### Ethics approval and consent to participate

For the systematic review part of this study, ethics approval was not applicable as it is necessary inclusion criteria for the to-be-evaluated RCTs. Validation of the SoRT metric was conducted using the Munich Antidepressant Response Signature (MARS) study and Emory Predictors of Remission in Depression to Individual and Combined Treatments (PReDICT) study. MARS has received approval from the Ethics Committee of the Ludwig Maximilians University in Munich, Germany, and PReDICT from the Emory Institutional Review Board and the Grady Hospital Research Oversight Committee. Both studies were conducted in concordance with the

Declaration of Helsinki [43, 44]. Participants' written informed consent was ascertained in MARS and PReDICT studies [43, 44].

### Consent for publication

Consent for publication was not generally applicable for the systematic review as the data from studies included in meta-analyses were in an individual *symptom* format (see explanation in data extraction and acquisition section), which is fully anonymous. However, we still checked whether authors of RCTs described that written informed consent was obtained, also because we allowed authors to send us individual *patient* data in case this made sharing of their data more convenient.

### Competing interests

NK, MR, JF, BAA, AGB, MM, GP, AGB, JD, JP, and JKB do not have any competing financial or other interests relating to the content of this study. HSM holds intellectual property in the field of deep brain stimulation for depression and is a consultant to Abbott Labs who has licenced the IP. WEC is a board member of Hugarheill ehf, an Icelandic company dedicated to the prevention of depression, and he receives book royalties from John Wiley & Sons. His research is supported by the NIH, the Mary and John Brock Foundation, and the Fuqua family foundations. He is a consultant to the George West Mental Health Foundation and is a member of the Scientific Advisory Boards of the ADAA and AIM for Mental Health. BWD has received research support from Acadia, Assurex Health, Axsome, Intra-Cellular Therapies, Janssen, National Institute of Mental Health, and Takeda. He has served as a consultant to Assurex Health and Aptinix. CBN has received funding from the National Institutes of Health and the Stanley Medical Research Institute. In the last 3 years, he has served as a consultant to Xhale, Takeda, Taisho Pharmaceutical Inc., Signant Health, Sunovion Pharmaceuticals Inc., Janssen Research & Development LLC, Magstim, Inc., Navitor Pharmaceuticals, Inc., TC MSO, Inc., Intra-Cellular Therapies, Inc., EMA Wellness, Gerson Lehrman Group (GLG), and Acadia Pharmaceuticals, and served on the Board of Directors for the Gratitude America, Anxiety Disorders Association of America (ADAA), and Xhale Smart, Inc. CBN is a stockholder in Xhale, Celgene, Seattle Genetics, Abbvie, OPKO Health, Inc., Antares, BI Gen Holdings, Inc., Corcept Therapeutics Pharmaceuticals Company, TC MSO, Inc., Trends in Pharma Development, LLC, and EMA Wellness, and serves on the Scientific Advisory Boards of the American Foundation for Suicide Prevention (AFSP), Brain and Behavior Research Foundation (BBRF), Xhale, ADAA, Skyland Trail, Signant Health, Laureate Institute for Brain Research (LIBR), Inc. CBN reports income sources or equity of \$10,000 or more from American Psychiatric Publishing, Xhale, Signant Health, CME Outfitters, Intra-Cellular Therapies, Inc., Magstim, and EMA Wellness, and has patents on the method and devices for transdermal delivery of lithium (US 6,375,990B1), the method of assessing antidepressant drug therapy via transport inhibition of monoamine neurotransmitters by ex vivo assay (US 7,148,027B2), and Compositions, Compositions, Methods of Synthesis, and Methods of Treatment (CRF Receptor Binding Ligand) (US 8,551, 996 B2). MK is supported by the NIH and The John J. McDonnell and Margaret T.O. O'Brien Foundation. DNK receives grant funding from the National Institute of Mental Health (NIMH). NH is the chair of the board of trustees of Manchester Global Foundation (MGF) which was founded in 2015 as a Charitable Incorporated Organisation (CIO) registered in England and Wales. NH is the past Trustee of Pakistan Institute of Living & Learning (PILL), Abaseen Foundation and Lancashire Mind. NH has attended educational events sponsored by pharmaceutical industry. RBJ's medical center collects the payments from the cognitive therapy she provides to patients. RBJ is a paid consultant to the National Institute of Mental Health and is a paid reviewer for UpToDate. She owns stock equity in Amgen, Johnson and Johnson, and Procter and Gamble. JRV is a paid reviewer for UpToDate. JPB received medication and placebo from Pfizer and NIMH funding for the data he contributed to the present study. MEK reports the following potential conflicts of interest: speakers bureau honoraria and other continuing medical education activity: AstraZeneca (Switzerland), Eli Lilly (Switzerland), Lundbeck (Switzerland), Vifor (Switzerland), and Zeller (Switzerland), as well as advisory panel payment from Lundbeck Switzerland.

### Author details

<sup>1</sup>Max Planck Institute of Psychiatry, Kraepelinstraße 2-10, 80804 Munich, Germany. <sup>2</sup>International Max Planck Research School for Translational Psychiatry (IMPRS-TP), Munich, Germany. <sup>3</sup>Department of Neurology and

Neurosurgery, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>4</sup>Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA, USA. <sup>5</sup>Department of Psychology, Emory University, Atlanta, GA, USA. <sup>6</sup>Institute for Early Life Adversity Research, University of Texas Dell Medical School in Austin, Austin, TX, USA. <sup>7</sup>Department of Psychiatry and Human Behavior, Brown University School of Medicine, Providence, RI, USA. <sup>8</sup>Department of Psychology, Stony Brook University, Stony Brook, NY, USA. <sup>9</sup>Department of Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA 94304, USA. <sup>10</sup>Division of Psychology and Mental Health, The University of Manchester, Manchester, UK. <sup>11</sup>Department of Psychiatry, The University of Texas Southwestern Medical Center, Dallas, TX, USA. <sup>12</sup>Department of Psychology, Truman State University, Kirksville, MO, USA. <sup>13</sup>Department of Biomedical and Neuromotor Sciences, University of Bologna, Bologna, Italy. <sup>14</sup>School of Psychiatry, University of New South Wales, Sydney, NSW, Australia. <sup>15</sup>Gordon F. Derner School of Psychology, Adelphi University, Garden City, New York, USA. <sup>16</sup>Contemporary Institute of Psychoanalysis and Transdisciplinarity of Porto Alegre, Porto Alegre, Brazil. <sup>17</sup>Department of Research, Arkin Mental Health Care, Amsterdam, Netherlands.

Received: 12 April 2018 Accepted: 7 May 2020

Published online: 05 June 2020

## References

- Otte C, Gold SM, Penninx BWJH, Pariante CM, Etkin A, Fava M, et al. Major depressive disorder. *Nat Rev Dis Prim.* 2016;2:16065. <https://doi.org/10.1038/nrdp.2016.65>.
- McClintock SM, Husain MM, Wisniewski SR, Nierenberg AA, Stewart JW, Trivedi MH, et al. Residual symptoms in depressed outpatients who respond by 50% but do not remit to antidepressant medication. *J Clin Psychopharmacol.* 2011;31:180–6. <https://doi.org/10.1097/JCP.0b013e31820ebd2c>.
- Treadway MT, Waskom ML, Dillon DG, Holmes AJ, Park MTM, Chakravarty MM, et al. Illness progression, recent stress and morphometry of hippocampal subfields and medial prefrontal cortex in major depression. *Biol Psychiatry.* 2015;77:285–94. <https://doi.org/10.1016/j.biopsych.2014.06.018>.
- Chang CH, Chen MC, Hong Qiu M, Lu J. Ventromedial prefrontal cortex regulates depressive-like behavior and rapid eye movement sleep in the rat. *Neuropharmacology.* 2014;86:125–32. <https://doi.org/10.1016/j.neuropharm.2014.07.005>.
- Lekman M, Laje G, Charney D, Rush AJ, Wilson AF, Sorant AJM, et al. The FKBP5-gene in depression and treatment response—an association study in the sequenced treatment alternatives to relieve depression (STAR\*D) cohort. *Biol Psychiatry.* 2008;63:1103–10. <https://doi.org/10.1016/j.biopsych.2007.10.026>.
- Bromet E, Andrade LH, Hwang I, Sampson NA, Alonso J, de Girolamo G, et al. Cross-national epidemiology of DSM-IV major depressive episode. *BMC Med.* 2011;9:90. <https://doi.org/10.1186/1741-7015-9-90>.
- Kessler RC, Bromet EJ. The epidemiology of depression across cultures. *Annu Rev Public Health.* 2013;34:119–38. <https://doi.org/10.1146/annurev-publhealth-031912-114409>.
- Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet.* 2015;386:743–800. [https://doi.org/10.1016/S0140-6736\(15\)60692-4](https://doi.org/10.1016/S0140-6736(15)60692-4).
- Köhler O, Benros ME, Nordentoft M, Farkouh ME, Iyengar RL, Mors O, et al. Effect of anti-inflammatory treatment on depression, depressive symptoms, and adverse effects: a systematic review and meta-analysis of randomized clinical trials. *JAMA Psychiatry.* 2014;71:1381–91. <https://doi.org/10.1001/jamapsychiatry.2014.1611>.
- Kappelmann N, Lewis G, Dantzer R, Jones PB, Khandaker GM. Antidepressant activity of anti-cytokine treatment: a systematic review and meta-analysis of clinical trials of chronic inflammatory conditions. *Mol Psychiatry.* 2018;23:335–43. <https://doi.org/10.1038/mp.2016.167>.
- Singh JB, Fedgchin M, Daly E, Xi L, Melman C, De Bruecker G, et al. Intravenous esketamine in adult treatment-resistant depression: a double-blind, double-randomization. Placebo-Controlled Study *Biol Psychiatry.* 2016;80:424–31. <https://doi.org/10.1016/j.biopsych.2015.10.018>.
- Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet.* 2018. [https://doi.org/10.1016/S0140-6736\(17\)32802-7](https://doi.org/10.1016/S0140-6736(17)32802-7).
- Barth J, Munder T, Gerger H, Nüesch E, Trelle S, Znoj H, et al. Comparative efficacy of seven psychotherapeutic interventions for patients with depression: a network meta-analysis. *PLoS Med.* 2013;10:e1001454. <https://doi.org/10.1371/journal.pmed.1001454>.
- Cuijpers P, Cristea IA, Karyotaki E, Reijnders M, Huibers MJH. 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. <https://doi.org/10.1002/wps.20346>.
- Adli M, Wiethoff K, Baghai TC, Fisher R, Seemüller F, Laakmann G, et al. How effective is algorithm-guided treatment for depressed inpatients? Results from the randomized controlled multicenter German algorithm project 3 trial. *Int J Neuropsychopharmacol.* 2017;20:721–30. <https://doi.org/10.1093/ijnp/pyx043>.
- Gaynes B, Warden D, Trivedi M, Wisniewski S, Fava M, Rush AJ. What did STAR\*D teach us? Results from a large-scale, practical, clinical trial for patients with depression. *Psychiatr Serv.* 2009;60:1439–45. <https://doi.org/10.1176/appi.ps.60.11.1439>.
- Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR\*D report. *Am J Psychiatry.* 2006;163:1905–17. <https://doi.org/10.1176/appi.ajp.163.11.1905>.
- Paul GL. Strategy of outcome research in psychotherapy. *J Consult Psychol.* 1967;31:109–18. <https://doi.org/10.1037/h0024436>.
- Iniesta R, Stahl D, McGuffin P. Machine learning, statistical learning and the future of biological research in psychiatry. *Psychol Med.* 2016;46:2455–65. <https://doi.org/10.1017/S0033291716001367>.
- Kessler RC, van Loo HM, Wardenar KJ, Bossarte RM, Brenner LA, Cai T, et al. Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. *Mol Psychiatry.* 2016;21:1366–71. <https://doi.org/10.1038/mp.2015.198>.
- Cuijpers P. Personalized treatment for functional outcome in depression. *Medicographia.* 2014;36:476–81.
- Dunlop BW, Rajendra JK, Craighead WE, Kelley ME, McGrath CL, Choi KS, et al. Functional connectivity of the subcallosal cingulate cortex and differential outcomes to treatment with cognitive-behavioral therapy or antidepressant medication for major depressive disorder. *Am J Psychiatry.* 2017;174:533–45. <https://doi.org/10.1176/appi.ajp.2016.16050518>.
- Niles AN, Loerinc AG, Krull JL, Roy-Byrne P, Sullivan G, Sherbourne CD, et al. Advancing personalized medicine: application of a novel statistical method to identify treatment moderators in the coordinated anxiety learning and management study. *Behav Ther.* 2017;48:490–500. <https://doi.org/10.1016/j.beth.2017.02.001>.
- Kraemer HC. Discovering, comparing, and combining moderators of treatment on outcome after randomized clinical trials: a parametric approach. *Stat Med.* 2013;32:1964–73. <https://doi.org/10.1002/sim.5734>.
- 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:e83875. <https://doi.org/10.1371/journal.pone.0083875>.
- Webb CA, Trivedi MH, Cohen ZD, Dillon DG, Fournier JC, Goer F, et al. Personalized prediction of antidepressant v. placebo response: evidence from the EMBARC study. *Psychol Med.* 2018;1–10. doi:<https://doi.org/10.1017/S0033291718001708>.
- Huibers MJH, Cohen ZD, Lemmens LHJM, Arntz A, Peeters FPML, Cuijpers P, et al. Predicting optimal outcomes in cognitive therapy or interpersonal psychotherapy for depressed individuals using the personalized advantage index approach. *PLoS One.* 2015;10:e0140771. <https://doi.org/10.1371/journal.pone.0140771>.
- Cohen ZD, Kim TT, Van HL, Dekker JJM, Driessen E. A demonstration of a multi-method variable selection approach for treatment selection: recommending cognitive-behavioral versus psychodynamic therapy for mild to moderate adult depression. *Psychother Res.* 2020;30:137–50. <https://doi.org/10.1080/10503307.2018.1563312>.
- Vittengl JR, Anna Clark L, Thase ME, Jarrett RB. Initial steps to inform selection of continuation cognitive therapy or fluoxetine for higher risk responders to cognitive therapy for recurrent major depressive disorder. *Psychiatry Res.* September 2016;2017(253):174–81.
- Weitz E, Kleiboer A, van Straten A, Hollon SD, Cuijpers P. Individual patient data meta-analysis of combined treatments versus psychotherapy (with or without



- pill placebo), pharmacotherapy or pill placebo for adult depression: a protocol. *BMJ Open*. 2017;7:e013478. <https://doi.org/10.1136/bmjopen-2016-013478>.
31. Fried EI, Nesse RM. The impact of individual depressive symptoms on impairment of psychosocial functioning. *PLoS One*. 2014;9.
  32. Fried EI, Nesse RM. Depression sum-scores don't add up: why analyzing specific depression symptoms is essential. *BMC Med*. 2015;13:72. <https://doi.org/10.1186/s12916-015-0325-4>.
  33. Fried EI. The 52 symptoms of major depression: lack of content overlap among seven common depression scales. *J Affect Disord*. 2017;208:191–7. <https://doi.org/10.1016/j.jad.2016.10.019>.
  34. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5®). Washington, DC: American Psychiatric Pub; 2013.
  35. Fried EI, Nesse RM. Depression is not a consistent syndrome: an investigation of unique symptom patterns in the STAR\*D study. *J Affect Disord*. 2015;172:96–102.
  36. Fournier JC, DeRubeis RJ, Hollon SD, Gallop R, Shelton RC, Amsterdam JD. Differential change in specific depressive symptoms during antidepressant medication or cognitive therapy. *Behav Res Ther*. 2013;51:392–8. <https://doi.org/10.1016/j.brat.2013.03.010>.
  37. Fried EI, van Borkulo CD, Epskamp S, Schoevers RA, Tuerlinckx F, Borsboom D. Measuring depression over time... Or not? Lack of unidimensionality and longitudinal measurement invariance in four common rating scales of depression. *Psychol Assess*. 2016;28:1354–67.
  38. Dunlop BW, Polychroniou PE, Rakofsky JJ, Nemeroff CB, Craighead WE, Mayberg HS. Suicidal ideation and other persisting symptoms after CBT or antidepressant medication treatment for major depressive disorder. *Psychol Med*. 2018;1–10. doi:<https://doi.org/10.1017/S0033291718002568>.
  39. Hieronymus F, Emilsson JF, Nilsson S, Eriksson E. Consistent superiority of selective serotonin reuptake inhibitors over placebo in reducing depressed mood in patients with major depression. *Mol Psychiatry*. 2015;21:523–30. <https://doi.org/10.1038/mp.2015.53>.
  40. Thoma NC, McKay D, Gerber AJ, Mirod BL, Edwards AR, Kocsis JH. A quality-based review of randomized controlled trials of cognitive-behavioral therapy for depression: an assessment and meta-regression. *Am J Psychiatry*. 2012;169:22–30. <https://doi.org/10.1176/appi.ajp.2011.11030433>.
  41. Huhn M, Tardy M, Maria Spinelli L, Kissling W, Förstl H, Pitschel-Walz G, et al. Efficacy of pharmacotherapy and psychotherapy for adult psychiatric disorders a systematic overview of meta-analyses. *JAMA Psychiatry*. 2014;71:706–15. <https://doi.org/10.1001/jamapsychiatry.2014.112>.
  42. Cuijpers P, Sijbrandij M, Koole SL, Andersson G, Beekman AT, Reynolds CF. The efficacy of psychotherapy and pharmacotherapy in treating depressive and anxiety disorders: a meta-analysis of direct comparisons. *World Psychiatry*. 2013;12:137–48. <https://doi.org/10.1002/wps.20038>.
  43. Hennings JM, Owashi T, Binder EB, Horstmann S, Menke A, Kloiber S, et al. Clinical characteristics and treatment outcome in a representative sample of depressed inpatients – findings from the Munich Antidepressant Response Signature (MARS) project. *J Psychiatr Res*. 2009;43:215–29. <https://doi.org/10.1016/j.jpsychires.2008.05.002>.
  44. Dunlop BW, Binder EB, Cubells JF, Goodman MM, Kelley ME, Kinkead B, et al. Predictors of remission in depression to individual and combined treatments (PReDICT): study protocol for a randomized controlled trial. *Trials*. 2012;13:1–18. <https://doi.org/10.1186/1745-6215-13-106>.
  45. Booth A, Clarke M, Dooley G, Ghersi D, Moher D, Petticrew M, et al. The nuts and bolts of PROSPERO: an international prospective register of systematic reviews. *Syst Rev*. 2012;1:1–8. <https://doi.org/10.1186/2046-4053-1-2>.
  46. Cuijpers P, van Straten A, Warmerdam L, Andersson G. Psychological treatment of depression: a meta-analytic database of randomized studies. *BMC Psychiatry*. 2008;8:36. <https://doi.org/10.1186/1471-244X-8-36>.
  47. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011;343:d5928. <https://doi.org/10.1136/bmj.d5928>.
  48. Munder T, Barth J. Cochrane's risk of bias tool in the context of psychotherapy outcome research. *Psychother Res*. 2017;1–9. doi:<https://doi.org/10.1080/10503307.2017.1411628>.
  49. Higgins JPT, Savović J, Page MJ, Sterne JAC. A revised tool to assess risk of bias in randomized trials (RoB 2.0). 2016. <https://sites.google.com/site/riskofbiastool/welcome/rob-2-0-tool>. Accessed 28 Feb 2018.
  50. Jelicic Kadic A, Vucic K, Dosenovic S, Sapunar D, Puljak L. Extracting data from figures with software was faster, with higher interrater reliability than manual extraction. *J Clin Epidemiol*. 2016;74:119–23. <https://doi.org/10.1016/j.jclinepi.2016.01.002>.
  51. Brant R. Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics*. 1990;46:1171–8. <https://doi.org/10.2307/2532457>.
  52. Agresti A, Kateri M. Categorical data analysis. In: Lovric M, editor. *International Encyclopedia of Statistical Science*. Berlin: Springer Berlin Heidelberg; 2011. p. 206–8. <https://doi.org/10.1007/978-3-642-04898-2>.
  53. Chinn S. A simple method for converting an odds ratio to effect size for use in meta-analysis. *Stat Med*. 2000;19:3127–31.
  54. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw*. 2010;36:1–48.
  55. R Core Team. R: a language and environment for statistical computing. 2017. <https://www.r-project.org/>. Accessed 19 Sept 2017.
  56. Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315:629–34.
  57. Dunlop BW, Kelley ME, Aponte-Rivera V, Mletzko-Crowe T, Kinkead B, Ritchie JC, et al. Effects of patient preferences on outcomes in the predictors of remission in depression to individual and combined treatments (PReDICT) study. *Am J Psychiatry*. 2017;174:546–56. <https://doi.org/10.1176/appi.ajp.2016.16050517>.
  58. Barber JP, Barrett MS, Gallop R, Rynn MA, Rickels K. Short-term dynamic psychotherapy versus pharmacotherapy for major depressive disorder. *J Clin Psychiatry*. 2012;73:66–73. <https://doi.org/10.4088/JCP.11m06831>.
  59. DiMascio A, Weissman MM, Prusoff BA, Neu C, Zwilling M, Klerman GL. Differential symptom reduction by drugs and psychotherapy in acute depression. *Arch Gen Psychiatry*. 1979;36:1450–6. <https://doi.org/10.1001/archpsyc.1979.01780130068008>.
  60. Dimidjian S, Hollon SD, Dobson KS, Schmaling KB, Kohlenberg RJ, Addis ME, et al. Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *J Consult Clin Psychol*. 2006;74:658–70.
  61. Elkin I, Shea T, Watkins JT, Imber SD, Sotsky SM, Collins JF, et al. National Institute of Mental Health treatment of depression collaborative research program. *Arch Gen Psychiatry*. 1989;46:971–82. <https://doi.org/10.1001/archpsyc.1989.01810110013002>.
  62. Frank E, Cassano GB, Rucci P, Thompson WK, Kraemer HC, Fagioliini A, et al. Predictors and moderators of time to remission of major depression with interpersonal psychotherapy and SSRI pharmacotherapy. *Psychol Med*. 2011;41:151–62. <https://doi.org/10.1017/S0033291710000553>.
  63. Harkness KL, Bagby RM, Kennedy SH. Childhood maltreatment and differential treatment response and recurrence in adult major depressive disorder. *J Consult Clin Psychol*. 2012;80:342–53. <https://doi.org/10.1037/a0027665>.
  64. Husain N, Chaudhry N, Fatima B, Husain M, Amin R, Chaudhry IB, et al. Antidepressant and group psychosocial treatment for depression: a rater blind exploratory RCT from a low income country. *Behav Cogn Psychother*. 2014;42:693–705. <https://doi.org/10.1017/S1352465813000441>.
  65. Jarrett RB, Schaffer M, McIntire D, Witt-Browder A, Kraft D, Risser RC. Treatment of atypical depression with cognitive therapy or phenelzine. *Arch Gen Psychiatry*. 1999;56:431. <https://doi.org/10.1001/archpsyc.56.5.431>.
  66. Keller MB, McCullough JP, Klein DN, Arnow B, Dunner DL, Gelenberg AJ, et al. A comparison of nefazodone, the cognitive behavioral-analysis system of psychotherapy, and their combination for the treatment of chronic depression. *N Engl J Med*. 2000;342:1462–70. <https://doi.org/10.1056/NEJM200005183422001>.
  67. Kennedy SH, Konarski JZ, Segal ZV, Lau MA, Bieling PJ, McIntyre RS, et al. Differences in brain glucose metabolism between responders to CBT and venlafaxine in a 16-week randomized controlled trial. *Am J Psychiatry*. 2007;164:778–88. <https://doi.org/10.1176/ajp.2007.164.5.778>.
  68. López Rodríguez J, López Butrón MA, Vargas Terrez BE, Villamil SV. Estudio doble ciego con antidepressivo, psicoterapia breve y placebo en pacientes con depresión leve a moderada. *Salud Ment*. 2004;27:53–61.
  69. Bastos AG, Guimaraes LSP, Trentini CM. The efficacy of long-term psychodynamic psychotherapy, fluoxetine and their combination in the outpatient treatment of depression. *Psychother Res*. 2015;25:612–24. <https://doi.org/10.1080/10503307.2014.935519>.
  70. Martin SD, Martin E, Rai SS, Richardson MA, Royall R. Brain blood flow changes in depressed patients treated with interpersonal psychotherapy or venlafaxine hydrochloride. *Arch Gen Psychiatry*. 2001;58:641. <https://doi.org/10.1001/archpsyc.58.7.641>.

71. McGrath CL, Kelley ME, Holtzheimer PE, Dunlop BW, Craighead WE, Franco AR, et al. Toward a neuroimaging treatment selection biomarker for major depressive disorder. *JAMA Psychiatry*. 2013;70:821–9. <https://doi.org/10.1001/jamapsychiatry.2013.143>.
72. McKnight DL, Nelson-Gray RO, Barnhill J. Dexamethasone suppression test and response to cognitive therapy and antidepressant medication. *Behav Ther*. 1992;23:99–111. [https://doi.org/10.1016/S0005-7894\(05\)80311-9](https://doi.org/10.1016/S0005-7894(05)80311-9).
73. McLean PD, Hakstian AR. Clinical depression: comparative efficacy of outpatient treatments. *J Consult Clin Psychol*. 1979;47:818–36. <https://doi.org/10.1037/0022-006X.47.5.818>.
74. Menchetti M, Rucci P, Bortolotti B, Bombi A, Scocco P, Kraemer HC, et al. Moderators of remission with interpersonal counselling or drug treatment in primary care patients with depression: randomised controlled trial. *Br J Psychiatry*. 2014;204:144–50. <https://doi.org/10.1192/bjp.bp.112.122663>.
75. Miranda J, Chung JY, Green BL, Krupnick J, Siddique J, Revicki DA, et al. Treating depression in predominantly low-income young minority women. *JAMA*. 2003;290:57–65. <https://doi.org/10.1001/jama.290.1.57>.
76. Moradveisi L, Huijbers MJH, Renner F, Arasteh M, Arntz A. Behavioural activation v. antidepressant medication for treating depression in Iran: randomised trial. *Br J Psychiatry*. 2013;202:204–11. <https://doi.org/10.1192/bjp.bp.112.113696>.
77. Murphy GE, Carney RM, Knesevich MA, Wetzel RD, Whitworth P. Cognitive behavior therapy, relaxation training, and tricyclic antidepressant medication in the treatment of depression. *Psychol Rep*. 1995;77:403–20. <https://doi.org/10.2466/pr0.1995.77.2.403>.
78. Mynors-Wallis LM, Gath DH, Lloyd-Thomas AR, Tomlinson D. Randomised controlled trial comparing problem solving treatment with amitriptyline and placebo for major depression in primary care. *BMJ*. 1995;310:441–5. <https://doi.org/10.1136/bmj.310.6977.441>.
79. Mynors-Wallis LM, Gath DH, Day A, Baker F. Randomised controlled trial of problem solving treatment, antidepressant medication, and combined treatment for major depression in primary care. *BMJ*. 2000;320:26–30. <https://doi.org/10.1136/bmj.320.7226.26>.
80. Bedi N, Chilvers C, Churchill R, Dewey M, Duggan C, Fielding K, et al. Assessing effectiveness of treatment of depression in primary care. Partially randomised preference trial. *Br J Psychiatry*. 2000;177 d:312–8. [internal-pdf://189.2.208.174/Bedi-2000-Assessing-effectiveness-of-treatment.pdf](https://doi.org/10.1192/bjp.rcpsych.org/content/bjprcpsych/177/4/312.full.pdf) LB - 01/8/30%65Cnhttp://bjpr.rcpsych.org/content/bjprcpsych/177/4/312.full.pdf.
81. Parker G, Blanch B, Paterson A, Hadzi-Pavlovic D, Sheppard E, Manicavasagar V, et al. The superiority of antidepressant medication to cognitive behavior therapy in melancholic depressed patients: a 12-week single-blind randomized study. *Acta Psychiatr Scand*. 2013;128:271–81. <https://doi.org/10.1111/acps.12049>.
82. Rush AJ, Beck AT, Kovacs M, Hollon S. Comparative efficacy of cognitive therapy and pharmacotherapy in the treatment of depressed outpatients. *Cognit Ther Res*. 1977;1:17–37. <https://doi.org/10.1007/BF01173502>.
83. Salminen JK, Karlsson H, Hietala J, Kajander J, Aalto S, Markkula J, et al. Short-term psychodynamic psychotherapy and fluoxetine in major depressive disorder: a randomized comparative study. *Psychother Psychosom*. 2008;77:351–7. <https://doi.org/10.1159/000151388>.
84. Scott AJ, Freeman CP. Edinburgh primary care depression study: treatment outcome, patient satisfaction, and cost after 16 weeks. *BMJ*. 1992;304:883–7. <https://doi.org/10.1136/bmj.304.6831.883>.
85. Shamsaei F, Rahimi A, Zarabian M, Sedehi M. Efficacy of pharmacotherapy and cognitive therapy, alone and in combination in major depressive disorder. *Hong Kong J Psychiatry*. 2008;18:76–80.
86. Thompson LW, Coon DW, Gallagher-Thompson D, Sommer BR, Koin D. Comparison of desipramine and cognitive/behavioral therapy in the treatment of elderly outpatients with mild-to-moderate depression. *Am J Geriatr Psychiatry*. 2001;9:225–40. <https://doi.org/10.1097/00019442-200108000-00006>.
87. Zu S, Xiang Y-T, Liu J, Zhang L, Wang G, Ma X, et al. A comparison of cognitive-behavioral therapy, antidepressants, their combination and standard treatment for Chinese patients with moderate–severe major depressive disorders. *J Affect Disord*. 2014;152–154:262–7. <https://doi.org/10.1016/j.jad.2013.09.022>.
88. Hollon SD, DeRubeis RJ, Evans MD, Wiemer MJ, Garvey MJ, Grove WM, et al. Cognitive therapy and pharmacotherapy for depression. *Arch Gen Psychiatry*. 1992;49:774. <https://doi.org/10.1001/archpsyc.1992.01820100018004>.
89. Blackburn IM, Bishop S, Glen AI, Whalley LJ, Christie JE. The efficacy of cognitive therapy in depression: a treatment trial using cognitive therapy and pharmacotherapy, each alone and in combination. *Br J Psychiatry*. 1981;139:181–9. <http://www.ncbi.nlm.nih.gov/pubmed/3140>.
90. Blackburn I-M, Moore RG. Controlled acute and follow-up trial of cognitive therapy and pharmacotherapy in out-patients with recurrent depression. *Br J Psychiatry*. 1997;171:328–34. <https://doi.org/10.1192/bjp.171.4.328>.
91. Blom MBJ, Jonker K, Dusseldorp E, Spinhoven P, Hoencamp E, Haffmans J, et al. Combination treatment for acute depression is superior only when psychotherapy is added to medication. *Psychother Psychosom*. 2007;76:289–97. <https://doi.org/10.1159/000104705>.
92. David D, Szentagotai A, Lupu V, Cosman D. Rational emotive behavior therapy, cognitive therapy, and medication in the treatment of major depressive disorder: a randomized clinical trial, posttreatment outcomes, and six-month follow-up. *J Clin Psychol*. 2008;64:728–46. <https://doi.org/10.1002/jclp.20487>.
93. Dekker JJM, Koelen JA, Van HL, Schoevers RA, Peen J, Hendriksen M, et al. Speed of action: the relative efficacy of short psychodynamic supportive psychotherapy and pharmacotherapy in the first 8 weeks of a treatment algorithm for depression. *J Affect Disord*. 2008;109:183–8. <https://doi.org/10.1016/j.jad.2007.10.015>.
94. DeRubeis RJ, Hollon SD, Amsterdam JD, Shelton RC, Young PR, Salomon RM, et al. Cognitive therapy vs medications in the treatment of moderate to severe depression. *Arch Gen Psychiatry*. 2005;62:409. <https://doi.org/10.1001/archpsyc.62.4.409>.
95. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B (Statistical Methodol)*. 1995;57:289–300.
96. Boschloo L, Bekhuis E, Erica S, Reijnders M, Derubeis RJ, Dimidjian S, et al. The symptom-specific efficacy of antidepressant medication vs. cognitive behavioral therapy in the treatment of depression : results from an individual patient data meta-analysis. *World Psychiatry*. 2019;18:183–91.
97. 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;19:92–107. <https://doi.org/10.1002/wps.20701>.
98. Craighead WE, Dunlop BW. Combination psychotherapy and antidepressant medication treatment for depression: for whom, when, and how. *Annu Rev Psychol*. 2014;65:267–300. <https://doi.org/10.1146/annurev.psych.121.208.131653>.
99. Zijlstra-Mano S, Keefe JR, Chui H, Rubin A, Barrett MS, Barber JP. Reducing dropout in treatment for depression. *J Clin Psychiatry*. 2016;77:e1584–90. <https://doi.org/10.4088/JCP.15m10081>.
100. Khazanov GK, Xu C, Dunn BD, Cohen ZD, DeRubeis RJ, Hollon SD. Distress and anhedonia as predictors of depression treatment outcome: a secondary analysis of a randomized clinical trial. *Behav Res Ther*. 2020;125:103507. <https://doi.org/10.1016/j.brat.2019.103507>.
101. Fried EI, von Stockert S, Haslbeck JMB, Lamers F, Schoevers RA, Penninx BWJH. Using network analysis to examine links between individual depressive symptoms, inflammatory markers, and covariates. *Psychol Med*. 2019;1–9. doi:<https://doi.org/10.1017/S0033291719002770>.
102. Jokela M, Virtanen M, Batty G, Kivimäki M. Inflammation and specific symptoms of depression. *JAMA Psychiatry*. 2016;73:87–8. <https://doi.org/10.1001/jamapsychiatry.2015.1977>.
103. Chu AL, Stochl J, Lewis G, Zammit S, Jones PB, Khandaker GM. Longitudinal association between inflammatory markers and specific symptoms of depression in a prospective birth cohort. *Brain Behav Immun*. 2019;76:74–81. <https://doi.org/10.1016/j.bbi.2018.11.007>.
104. Lamers F, Milaneschi Y, de Jonge P, Giltay EJ, Penninx BWJH. Metabolic and inflammatory markers: associations with individual depressive symptoms. *Psychol Med*. 2018;48:1102–10. <https://doi.org/10.1017/S0033291717002483>.
105. Choi SW, O'Reilly PF. PRSice-2: polygenic risk score software for biobank-scale data. *Gigascience*. 2019;8:1–6. <https://doi.org/10.1093/gigascience/giz082>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.