Clinical Psychology Review 100 (2023) 102236

Contents lists available at ScienceDirect

Clinical Psychology Review

journal homepage: www.elsevier.com/locate/clinpsychrev

Review

Efficacy and safety of transcranial magnetic stimulation for treating major depressive disorder: An umbrella review and re-analysis of published meta-analyses of randomised controlled trials

S. Brini [a,*], N.I. Brudasca [a], A. Hodkinson [b], K. Kaluzinska [c], A. Wach [c], D. Storman [d], A. Prokop-Dorner [e], P. Jemioło [f], M.M. Bala [d]

[a] Division of Health Services Research and Management School of Health Sciences, City, University of London, London, United Kingdom
[b] National Institute for Health Research (NIHR), School for Primary Care Research, Manchester Academic Health Science Centre, University of Manchester, United Kingdom
[c] Jagiellonian University Medical College, Faculty of Medicine, Students' Scientific Group of Systematic Reviews, Kraków, Poland
[d] Chair of Epidemiology and Preventive Medicine, Department of Hygiene and Dietetics, Jagiellonian University Medical College, Kraków, Poland
[e] Chair of Epidemiology and Preventive Medicine, Department of Medical Sociology, Jagiellonian University Medical College, Kraków, Poland
[f] AGH University of Science and Technology, Kraków, Poland

ARTICLE INFO

Keywords:
Transcranial magnetic stimulation
Major depressive disorder
Systematic review
Meta-analysis
Prediction intervals
Heterogeneity

ABSTRACT

Objectives: We re-analysed data from published meta-analyses testing the effects of Transcranial Magnetic Stimulation (TMS) on Major Depressive Disorder (MDD) in adults. We applied up-to-date meta-analytic techniques for handling heterogeneity including the random-effects Hartung-Knapp-Sidik-Jonkman method and estimated 95% prediction intervals. Heterogeneity practices in published meta-analyses were assessed as a secondary aim.

Study design and setting: We performed systematic searches of systematic reviews with meta-analyses that included randomised controlled trials assessing the efficacy, tolerability, and side effects of TMS on MDD. We performed risk of bias assessment using A MeaSurement Tool to Assess Reviews (AMSTAR) 2 and re-analysed meta-analyses involving 10 or more primary studies.

Results: We included 29 systematic reviews and re-analysed 15 meta-analyses. Authors of all meta-analyses interpreted findings to suggest TMS is safe and effective for MDD. Our re-analysis showed that in 14 out of 15 meta-analyses, the 95% prediction intervals included the null and captured values in the opposite effect direction. We also detected presence of small-study effects in some meta-analyses and 24 out of 25 systematic reviews received an AMSTAR 2 rating classed as *critically low*.

Conclusion: Authors of all included meta-analyses interpreted findings to suggest TMS is safe and effective for MDD despite lack of comprehensive investigation of heterogeneity. Our re-analysis revealed the direction and magnitude of treatment effects vary widely across different settings. We also found high risk of bias in the majority of included systematic reviews and presence of small-study effects in some meta-analyses. Because of these reasons, we argue TMS for MDD may not be as effective and potentially less tolerated in some populations than current evidence suggests.

1. Introduction

The application of Transcranial Magnetic Stimulation (TMS) for the management of mood disorders has been investigated extensively. Several meta-analyses support its safety and effectiveness in treating Major Depressive Disorder (MDD) documenting improvements in MDD symptomatology (Berlim, Van den Eynde, & Daskalakis, 2013a; Berlim, Van den Eynde, & Daskalakis, 2013b; Kozel & George, 2002; Leggett et al., 2015; Valiengo et al., 2022). Some authors argue the evidence supporting the efficacy and safety of the TMS treatment for depression is "substantive and unequivocal" (Fitzgerald et al., 2021; Fitzgerald et al., 2021; Fitzgerald, George, & Pridmore, 2021). Although the use of TMS

* Corresponding author at: School of Health Sciences, City, University of London, 10 Northampton Square, London EC1V OHB, United Kingdom.
E-mail address: stefano.brini@city.ac.uk (S. Brini).

https://doi.org/10.1016/j.cpr.2022.102236
Received 13 October 2021; Received in revised form 15 October 2022; Accepted 24 November 2022
Available online 8 December 2022

for treating MDD was approved in 2008 by the United States Food and Drug Administration (McClintock et al., 2018), and it is now used in healthcare settings worldwide for this condition (Bolu et al., 2021; Carpenter et al., 2012; Dowling, Bonwick, Dharwadkar, & Ng, 2020; Janicak et al., 2013), there are also authors who have challenged the clinical utility of this intervention (Amad & Fovet, 2022; Malhi et al., 2021).

There is some evidence to suggest that poorly designed and underpowered trials may have resulted in excess rates of false positives in TMS literature for mood disorders, including MDD (Amad et al., 2019), possibly skewing the results of some meta-analyses toward favouring TMS over controls. Indeed, some meta-analyses have indicated that TMS for depression may not be beneficial in all populations (Couturier, 2005; Martin et al., 2003; Rodriguez-Martin et al., 2002). The 2020 Royal Australian and New Zealand College of Psychiatrists clinical practice guidelines for the management of mood disorders, among other authors, are concerned the estimates from meta-analyses in this field may not be representative of the true value (Amad et al., 2019; Malhi et al., 2021a; Malhi et al., 2021b). The application of prediction intervals (PI), as opposed to confidence intervals (CI) which represent the degree of precision around the average treatment effect, provides an additional index of heterogeneity (true differences among studies) that describe the direction and magnitude of TMS treatment effects across different settings. Because of this reason, PI should be routinely reported in meta-analysis addressing clinical questions such as in the TMS literature for depression (Borenstein, 2019; Deeks, Higgins, & Altman, 2019; IntHout, Ioannidis, Rovers, & Goeman, 2016).

Despite their relevance for clinical practice, PIs are seldom calculated in random-effects meta-analyses (Aytug, Rothstein, Zhou, & Kern, 2012; Borenstein, 2019; Carlson & Ji, 2011; IntHout et al., 2016; Moher, 1995; Paterson, Harms, Steel, & Credé, 2016; Riley, Higgins, & Deeks, 2011) across different scientific fields such as anaesthesiology (Umberham et al., 2017), exercise science (Impellizzeri, McCall, & van Smeden, 2021), and mixed medical settings (IntHout et al., 2016). Studies that have re-analysed existing meta-analysis whilst reporting the PIs have often found that the width of the PI tends to be wider than the width of the CI in the original results. For example, in a re-analysis of Cochrane meta-analyses involving mixed medical settings the inclusion of 95% PIs revealed that in 347 out of 479 meta-analyses the intervals overlapped the null and occasionally favoured the opposite effect direction (IntHout et al., 2016). This indicates that some individuals may experience no effect or even adverse outcomes. The findings from the Cochrane re-analyses (IntHout et al., 2016) are in line with results of other re-analyses of existing meta-analyses (Aytug et al., 2012; Carlson & Ji, 2011; Chiolero, Santschi, Burnand, Platt, & Paradis, 2012; Graham & Moran, 2012; Kelley & Kelley, 2009; Paterson et al., 2016).

Methodological simulation studies have also found that standard random-effects meta-analysis using the DerSimonian and Laird method of pooling effect sizes from primary studies is often unreliable when non-trivial heterogeneity is present (Langan et al., 2019). Authors performing random-effects meta-analyses often apply this method in estimating the weighted pool effect (DerSimonian & Laird, 1986; Follmann & Proschan, 1999; IntHout, Ioannidis, & Borm, 2014; Röver, Knapp, & Friede, 2015). This approach, particularly in meta-analysing with a small number of studies with varied sample sizes, or in the presence of non-trivial heterogeneity, can lead to more false-positive results (DerSimonian & Laird, 1986; Follmann & Proschan, 1999) than the Hartung-Knapp-Sidik-Jonkman (HKSJ) method, for example (IntHout et al., 2014; Röver et al., 2015). The HKSJ is more likely to generate wider CI around the weighted pool estimate increasing precision and lowering the risk of type I error (Follmann & Proschan, 1999; Langan et al., 2019; Sánchez-Meca & Marín-Martínez, 2008). Since most random-effects meta-analyses are likely to detect presence of non-trivial heterogeneity and to include fewer than 20 studies, the application of the HKSJ, rather than other methods, is strongly recommended (IntHout et al., 2014; Thorlund, Wetterslev, Awad, Thabane, & Gluud, 2011). The frequency in which the HKSJ method is applied in meta-analyses investigating the efficacy and safety of TMS for MDD, however, is currently not known.

To our knowledge, there are no studies exploring the assessment of heterogeneity, the application of PI, or whether the HKJS method is routinely used in meta-analyses assessing the efficacy and safety of TMS in treating MDD. We aimed to apply robust meta-analytic methods and re-analyse published meta-analyses in this field. Specifically, we aimed to re-analyse meta-analyses by applying the HKSJ method and by estimating 95% PI (Borenstein, 2019; Graham & Moran, 2012; Guddat, Grouven, Bender, & Skipka, 2012; IntHout et al., 2016). The primary aims of the re-analyses were to explore any potential changes in the interpretation of meta-analytic results between the original published findings with our re-evaluation of the same evidence. Heterogeneity practices and risk of bias were assessed as secondary aims.

2. Methods

2.1. Protocol and registration

We followed a protocol registered a priori in PROSPERO (CRD42020165516).

2.2. Eligibility criteria

We included studies published as systematic reviews with meta-analyses of randomised controlled trials (RCTs) that assessed the application of TMS in participants aged 18 years or older. At least 50% of participants had to have a current diagnosis of MDD according to any standardised diagnostic criterion [(e.g., Diagnostic and Statistical Manual of Mental Disorders (DSM-V) (American Psychiatric Association, 2013) and International Statistical Classification of Diseases and Related Health Problems (ICD-10)] (World Health Organization, 1992). We defined systematic reviews with meta-analysis as reviews that apply systematic and reproducible methods to identify, select, critically appraise relevant studies, and synthesize data from the studies that are included in the review (Higgins et al., 2019). No limit on the number of primary studies included in the meta-analysis was imposed as a cut-off point for the inclusion of the systematic reviews.

A concurrent secondary diagnosis of another neuropsychiatric disorder (e.g., dementia, epilepsy) in participants was not considered a limit. We included meta-analyses assessing the efficacy (e.g., remission, response, change in depression), safety, and/or tolerability of any type of TMS, irrespective of the protocol type, against any comparator (e.g., sham stimulation, active stimulation, pharmacotherapy). Meta-analyses that included any other clinical outcomes (e.g., anxiety, treatment-emergent adverse events) were also considered. Measures of the efficacy outcome had to be evaluated for each group through specific standardised scales [e.g., Hamilton Depression Rating Scale (HAMD) (Hamilton, 1960), Montgomery-Asberg Depression Rating Scale (MADRS) (Montgomery & Åsberg, 1979), Beck Depression Inventory (BDI)] (Beck, Steer, & Brown, 1996). Tolerability and safety had to be measured for each group by withdrawal from the trial and by the incidence of adverse effects (e.g., headaches, nausea, hospitalisation), respectively.

2.3. Information sources, search strategy and study selection

An electronic search was performed on the 14th of January 2020 and was updated on the 5th of September 2021, and again on the 3rd of April 2022 in the following databases: MEDLINE, Embase, Cochrane Library (CENTRAL), Web of Science, PsycINFO, Database of Abstracts of Reviews of Effects, National Health Service Economic Evaluation Database, and Health Technology Assessment. The search strategy was adapted for each database. No restrictions, such as language or date were applied. We used EndNote and Covidence to manage the selection process (Lefebvre et al., 2019). Study selection was undertaken by two

independent reviewers at each step of identification and full-text assessment using Covidence (DS, APD, SB, NIB). Titles and abstracts for all records were screened and irrelevant papers were excluded. For the potentially eligible studies, the full text was examined for compliance with the inclusion criteria. We also screened the reference lists of the records that met the inclusion criteria. At each step, disagreements were resolved by discussion and, if required, in consultation with a third independent reviewer (MMB, SB, DS).

2.4. Data extraction and coding

A customized fully pilot-tested data extraction sheet in Microsoft Excel was used to collect information from each included systematic review and meta-analysis. We extracted the following data: 1) summary measures with the 95% CI, 2) the statistical method for pooling (i.e. fixed/random), and the weighting approach if fixed-effects (i.e., Inverse Variance or Mantel-Haenszel) or method of estimating τ^2 if random-effects (i.e. DerSimonian-Laird, HKSJ, empirical Bayes, or other), 3) the methods for quantifying heterogeneity and between-study variance (i.e., Q-statistic and its *p*-value, I^2, τ^2), 4) the methods for calculating small study effects, together with the tool used, if any, 5) number of RCTs included in the meta-analysis, total sample size, participant demographic (age, gender), type of TMS protocol frequency, intensity, brain region, outcome including methods for diagnosis, results, and interpretation of results were also collected if reported in the included systematic reviews. Data collection about demographics was performed by two independent reviewers (KK, AW, PJ, DS) and for outcomes, extraction was performed by one reviewer (DS or PJ) and independently checked by a 2nd reviewer (NIB or DS). Disagreements were resolved by discussion and, if required, in consultation with a 3rd independent reviewer (DS). Three authors were contacted for supplementary materials (Hung et al., 2020; Leggett et al., 2015; Zhang et al., 2015) and further data was received from one author (Leggett et al., 2015).

2.5. Risk of bias

The methodological quality of the identified studies was assessed using the current version of A MeaSurement Tool to Assess Reviews tool (AMSTAR 2) (Shea et al., 2017). The tool consists of 16 items covering domains relevant to the quality of systematic reviewing, for which possible assessment responses include *Yes*, *Partial Yes*, or *No*. In our assessments, we followed available guidance documents (Shea et al., 2017), where items number 2 (protocol), 4 (literature search), 7 (justification for study exclusions), 9 (assessment of risk of bias in individual studies), 11 (methods of synthesis), 13 (incorporation of risk of bias assessment in the interpretation of the results), and 15 (assessment of publication bias) were identified as critical domains. Following the guidance, we classified overall confidence in the results of the review as high (maximum of one non-critical weakness), moderate (> one non-critical weakness), low (one critical flaw), or critically low (> one critical flaw). The assessment of methodological quality and risk of bias was carried out along with the extraction process by two reviewers independently (KK, AW, PJ, DS). Disagreements were resolved by discussion.

2.6. Data synthesis and analysis

All analyses were carried out in R version 4.0.5 (R Foundation for Statistical Computing) using the packages (*meta*, *metafor*). For the re-analysis with the 95% PI and to investigate small-study effects we followed published recommendations (Borenstein, 2019; Hedges & Vevea, 1998; Lau, Ioannidis, Terrin, Schmid, & Olkin, 2006). The re-analysis was only performed in meta-analyses with at least 10 RCTs to ensure accuracy when applying HKSJ random-effects and calculation of the PI (Borenstein, 2020; Partlett & Riley, 2017). Before re-analysis, we transformed the individual study effects extracted from each meta-analysis using the statistical software Comprehensive Meta-Analysis Version 3. This meant that meta-analysis reporting relative risk (RR), risk difference (RD), or odds ratio (OR) study effects, were transformed to the corresponding log scale (i.e., log RR, log OR, or log RD) with their associated standard error (SE) estimates. For studies reporting the individual study effects as standardised mean difference (SMD) or the raw mean difference (MD), the SE was calculated by using the standard formula given the 95% CI (Higgins, Li, & Deeks, 2019).

Once the data were transformed and ready with the required effect size measure and corresponding SE, the re-analysis of the meta-analysis was done firstly by repeating the exact pooling method used in the original published study. Then, we re-analysed the meta-analysis using HKSJ adjustment reporting the new effect estimate (95% CI), I^2 (95% CI), τ^2, and the PIs. For OR and RR pooled estimates the predicted probabilities were also reported using the margins package in R (Leeper, 2021). Small-study effects were examined visually through funnel plots and using Egger's statistical test. Our results of the re-analysis were presented in Table 4 to enable a full methodological critique of the findings of our re-analysis in light of the original findings.

3. Results

3.1. Study selection

The database search retrieved a total of 543 references. After 201 duplicates were removed, a further 291 articles were excluded after screening the title and abstract. A total of 45 studies were retained for full-text screening, of which 29 systematic reviews with meta-analysis met the eligibility criteria. The study selection process is presented in Fig. 1.

The included systematic reviews were published between 2001 and 2022. The majority ($k = 25$, 86%) of studies (Berlim et al., 2013a; Berlim et al., 2013b; Berlim, Van den Eynde, Tovar-Perdomo, & Daskalakis, 2014; Chen et al., 2013; Chen, Zhao, Liu, Fan, & Xie, 2017; Chu et al., 2021; Couturier, 2005; Gaynes et al., 2014; Gellersen & Kedzior, 2019; Hung et al., 2020; Kedzior, Azorina, & Reitz, 2014; Kozel & George, 2002; Lam, Chan, Wilkins-Ho, & Yatham, 2008; Liu, Zhang, Zhang, & Li, 2014; McNamara, Ray, Arthurs, & Boniface, 2001; Schutter, 2009; Schutter, 2010; Slotema, Blom, Hoek, & Sommer, 2010; Sonmez et al., 2019; Valiengo et al., 2022; Voigt, Leuchter, & Carpenter, 2021; Wang et al., 2022; Wei et al., 2017; Xie, Chen, & Wei, 2013; Zhang et al., 2015) did not specify the setting for the treatment and four studies (Kedzior, Reitz, Azorina, & Loo, 2015; Leggett et al., 2015; Martin et al., 2003; Rodriguez-Martin et al., 2002) which provided such information included studies in patients both hospitalised and in outpatient clinics. Patient demographic and clinical characteristics were not reported in all systematic reviews (Table 1). Ethnicity was not reported in any of the studies. The different types of mood disorders participants had been diagnosed with, included: primary MDD, depression in bipolar disorder, major depressive episode, dysthymia, treatment-resistant depression, and minor depression, which were diagnosed with Diagnostic and Statistical Manual of Mental Disorders or International Classification of Diseases, Chinese Classification of Mental Disorders, or Mini-Mental State Examination. Most primary studies eligible in the reviews included patients with *treatment-resistant depression* which generally refers to "*inadequate response to at least one antidepressant trial of adequate doses and duration among patients suffering from unipolar depressive disorders*" (Fava, 2003).

3.2. Methodological quality of included studies

Most included systematic reviews (28/29; 97%) received a rating classed as *critically low quality* due to more than one flaw in critical items (Berlim et al., 2013a; Berlim et al., 2013b; Berlim et al., 2014; Chen et al., 2013; Chen et al., 2017; Chu et al., 2021; Couturier, 2005; Gaynes et al., 2014; Gellersen & Kedzior, 2019; Hung et al., 2020; Kedzior et al., 2014; Kedzior et al., 2015; Kozel & George, 2002; Lam et al., 2008;

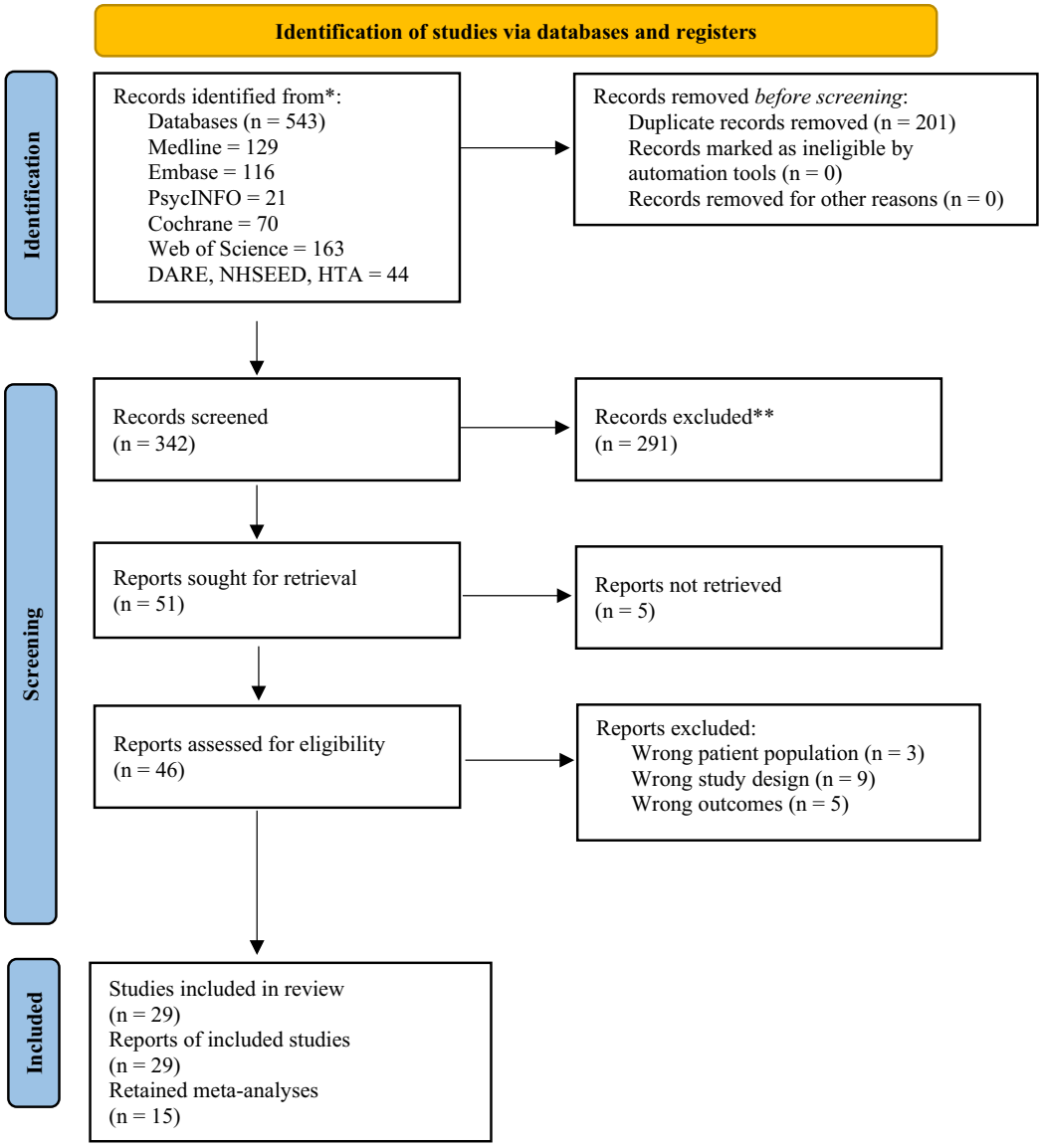

Fig. 1. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) PRISMA 2010 statement.
DARE: Database of Abstracts of Reviews of Effects.
NHSEED: National Health Service Economic Evaluation Database.
HTA: Health Technology Assessment.

Leggett et al., 2015; Liu et al., 2014; Martin et al., 2003; McNamara et al., 2001; Schutter, 2009; Schutter, 2010; Slotema et al., 2010; Sonmez et al., 2019; Valiengo et al., 2022; Voigt et al., 2021; Wang et al., 2022; Wei et al., 2017; Xie et al., 2013; Zhang et al., 2015), only one study received a rating classed as *low quality* (with one flaw in critical item) (Rodriguez-Martin et al., 2002). The majority (26/29; 90%) of included systematic reviews did not meet criteria for item 2 (did not refer to the existence of the protocol) and did not fully meet criteria for item 4 (comprehensive search strategy; 27/29;93%), did not use satisfactory tool for risk of bias assessment of primary studies (item 9–80%) and did not account for risk of bias in the interpretation of the results (item 13–88%), did not perform adequate investigations of publication bias (item 15–76%), and more thna half did not report information on excluded studies (item 7–56%). Most included systematic reviews used appropriate methods for statistical combination of the results (item 11–88%). Among non-critical items of the AMSTAR 2 tool, the most common concerns in the included systematic reviews were: missing information about sources of funding of primary studies (item 10–100%), insufficient description of included studies (item 8–88%), assessment of the potential impact of risk of bias on the results (item 12–84%), missing explanation of the selection of study designs (item 3–80%), did not report on duplicate study selection and data extraction (item 5–76% and item 6–64%). Most studies specified the research question and inclusion criteria using the Population, Intervention, Control, Outcome (PICO) format (item 1–96%), reported on conflict of interest (item 16–96%), and provided satisfactory explanation and discussion of heterogeneity (item 14–72%).

3.3. *Re-analysis of the meta-analyses*

3.3.1. *Efficacy*

Of the 12 meta-analyses that were eligible for re-analyses involving efficacy as an outcome, the median number of studies included was 15 (interquartile range (IQR): 7) and the median number of patients involved was 817 (interquartile range: 650) (Table 2). The RR was reported as the summary measure of response rates in three articles

Table 1
Demographic and clinical characteristics of participants within each systematic review.

Study	Setting	Mean age (SD) in years	Total: % women	rTMS: % women	Controls: % women	Mood disorder type	Diagnostic tool	Number of studies with patients on medication / all studies	Number of studies with TRD participants
(Berlim et al., 2013a)	NR	48.5 (NR)	57.0	56.2	58.1	MDD, BD	DSM, ICD-10	6/7 – augmentation	7
(Berlim et al., 2013b)	NR	49.9 (NR)	66.0	65.6	66.4	MDD, BD	DSM, ICD-10	6/8 - augmentation	7
(Berlim et al., 2014)	NR	47.5 (NR)	56.6	58.6	54.4	MDD	DSM-IV or later, ICD-10	21/29 - augmentation	23
(Chen et al., 2013)	NR	49.2 (NR)	62.7	61.1	64.3	MDD, BD	NR	6/7 - augmentation	6
(Chen et al., 2017)	NR	48.5 (NR)	62.8	NR	NR	MDD, BD	NR	19/25 - augmentation	21
(Chu et al., 2021)	NR	44.2 (11.5)	60.8	59.6	64.8	MDD, BD	DSM-IV, ICD-10	8/10 - augmentation	8
(Couturier, 2005)	NR	NR	NR	NR	NR	MDE	DSM-IV	NR	2
(Gaynes et al., 2014)	NR	NR	NR	NR	NR	MDD, BD	NR	15/27 - augmentation, 6/27 - switch strategy, 5/27 mixed (either switch or augmentation), 1/27 - NR	27
(Gellersen & Kedzior, 2019)	NR	45.1 (NR)	52.9	NR	NR	MDD	DSM-IV	5/8 - concurrent antidepressants; 1/8 - only some patients on medication	8
(Hung et al., 2020)	NR	50.7 (NR)	49.6	48.9	50.2	MDD	DSM-IV, DSM-IV-TR, DSM-V, ICD-10; diagnosis of TRD based on previous definition by Fava et al. (Fava, 2003) or Maudsley staging method Fekadu, Donocik, and Cleare (2018)	2/3 - concurrent antidepressants	3
(Kedzior et al., 2014)	NR	43.8 (NR)	62.5	NR	NR	MDD, MDE, dysthymia	DSM-IV, ICD-10	11/14 - antidepressants (stable doses), 1/14- antidepressants were started on day 1 concurrently with rTMS, 2/14- antidepressant free (or only mood stabilizers)	12
(Kedzior et al., 2015)	4 studies - inpatients; 5 studies - outpatients; rest - NR	NR	62	NR	NR	MDD, MDE, dysthymia	DSM-IV, ICD-10	5/16- antidepressants were started concurrently with rTMS on day 1; 9/16 – continued medications (stable doses)	8
(Kozel & George, 2002)	NR	NR	NR	NR	NR	unspecified	NR	7/12 continued stable medications;2/12 started medications (sertraline),	NR
(Lam et al., 2008)	NR	NR	NR	NR	NR	MDD with an explicit definition of TRD that included at least one failed trial of an AD	DSM, ICD	17/24 - patients stayed on same medications (stable doses); 1/24 - all patients started on escitalopram 20 mg; 1/24 - all patients started on sertraline 50 mg	23
(Leggett et al., 2015)	15 studies outpatients, 5 studies inpatients, 7 studies both, 43 studies NR	45.4 (NR)	59.3	60.6	57.7	MDD, BD	NR	NR	70
(Liu et al., 2014)	NR	NR	NR	NR	NR	MDD	DSM-IV, ICD-10	6/7 – patients continued their treatment;1/7 – patients received Escitalopram (10-20 mg/d)	7

(*continued on next page*)

Table 1 (*continued*)

Study	Setting	Mean age (SD) in years	Total: % women	rTMS: % women	Controls: % women	Mood disorder type	Diagnostic tool	Number of studies with patients on medication / all studies	Number of studies with TRD participants
(Martin et al., 2003)	2 studies -in patients; 4 studies - outpatients; 2 studies - both in- and outpatients; 5 studies - NR	NR	60.5	NR	NR	MDD or BD (depressed phase), 1 study with minor depression	DSM-IV, ICD-10	10/14 – patients continued their medications; 1/14 - all patients started on sertraline 50 mg	7
(McNamara et al., 2001)	NR	NR	NR	NR	NR	MDE, 1 study - at least three episodes of depression that had been resistant to medication	DSM-III-R, DSM-IV	4/5 - augmentation (antidepressants)	1
(Rodriguez-Martin et al., 2002)	4 studies outpatients; 2 studies inpatients; 3 studies both; 5 studies NR	NR	61.74	NR	NR	MDD, BD (depressed phase)	DSM-IV, DSM-III-R	10/14 - patients continued their medication; 1/14 - same medicament for all patients for 6 weeks before and then during the study; 1/14 study - psychiatric medication discontinued, clonazepam started for all patients	8
(Schutter, 2009)	NR	49.1 (7.5)	NR	NR	NR	MDE without psychotic features	DSM-IV	NR	17
(Schutter, 2010)	NR	50.0 (NR)	NR	NR	NR	MDE	DSM-IV	NR	4
(Slotema et al., 2010)	NR	NR	NR	NR	NR	unspecified	DSM, ICD	17/40 – Continuation of antidepressants 5/40 antidepressants started simultaneously with rTMS 1/3 Psychotropic allowed (without washout period); 2/3 benzodiazepine only (with washout period)	NR
(Sonmez et al., 2019)	NR	NR	NR	NR	NR	MDE, unipolar TRD	DSM-IV, MINI (DSM-IV)		2
(Valiengo et al., 2022)	NR	NR	NR	NR	NR	MDD	DSM-III or later, ICD	NR	14
(Voigt et al., 2021)	NR	NR	59.0	NR	NR	MDD	NR	NR	4
(Wang et al., 2022)	NR	NR	NR	NR	NR	MDD	e.g. DSM-V, ICD-10	17/53 augmentation; 8/53 - mixed	21
(Wei et al., 2017)	NR	43.2 (NR)	56.5	56.9	56	unspecified	DSM-IV, CCMD-3, ICD-10	29/29 Antidepressants	3
(Xie et al., 2013)	NR	47.5 (NR)	66	65	67	MDD, BD	NR	3/9 - rTMS as add-on strategy; 2/9 - mixed strategy (some patients maintained their usual medication regimen)	6
(Zhang et al., 2015)	NR	NR	NR	NR	NR	MDD, BD	DSM-IV, ICD10, MINI	NR	10

NR: not reported; BD: depression in bipolar disorder; MDD: major depressive disorder; MDE: major depressive episode; PTSD: post-traumatic stress disorder; SD: standard deviation; TRD: treatment resistant depression; HKSJ: Hartung-Knapp-Sidik-Jonkman
* Ethnicity was not reported in any of the studies

(Gaynes et al., 2014; Leggett et al., 2015; Zhang et al., 2015) and of remission rates in one (Leggett et al., 2015); SMD as a summary measure of change in depression scores in four articles (Kedzior et al., 2014; Valiengo et al., 2022; Wei et al., 2017; Zhang et al., 2015); OR as a summary measure of response rates in two articles (Berlim et al., 2014; Valiengo et al., 2022) and of remission rates in one article (Berlim et al., 2014); and cumulative SMD of change in depression scores in one article (Kozel & George, 2002). For three systematic reviews (30%) the statistical method of pooling RCTs in their meta-analyses was not reported (Gaynes et al., 2014; Kozel & George, 2002; Leggett et al., 2015) (30%). Therefore, we made best efforts to try to determine the exact method used by running multiple analysis with different pooling methods until we obtained results close to the original results. Once this was achieved, a re-analysis of 12 meta-analyses was performed and results were compared to the original results of the original meta-analyses. In four studies (Kedzior et al., 2015; Kozel & George, 2002; Leggett et al., 2015; Wei et al., 2017), the re-analysis did not match the original published meta-analysis results and deviated by more than one decimal place based on the effect size (see Table 2; Fig. 2-4). Estimated heterogeneity with the I^2 measure in six (Kedzior et al., 2014; Kedzior et al., 2015; Leggett et al., 2015; Wei et al., 2017; Zhang et al., 2015) of the re-analyses deviated by >1% compared with what was reported in the journal publications.

The re-analysis using HKSJ random-effects did not alter the significance in any of the re-analysed pooled effects when compared to the original pooled effect estimates. The I^2 values also did not deviate by >10% in either direction. In all meta-analyses, the width of the PI was wider than the width of the CI and in 14 meta-analyses out of 15 the PI included the null value and, which is in contrast with the corresponding significant 95% CIs as calculated using HKSJ random-effects. In terms of

Table 2
Methods used for quantifying heterogeneity and between-study variance, use of prediction interval and HKSJ method in included reviews. Red colour represents the method was not performed, while the green colour represents the method was performed.

PI: prediction intervals
HKSJ: Hartung-Knapp-Sidik-Jonkman

the actual percentage increase in changes of response/remission, the estimates were found to range from a 62% to 78% increase. Egger's test was statistically significant in six meta-analyses (Berlim et al., 2014; Kedzior et al., 2014; Leggett et al., 2015; Zhang et al., 2015) indicating presence of small-study effects. Visual inspection of these six funnel plots also revealed slight asymmetry indicating small-study effects and corroborating results from Egger's test.

3.3.2. *Acceptability*

For acceptability, two reviews that involved a medium of 20 studies and 1293 participants (IRQ: 102) were re-analysed (Berlim et al., 2014; Wang et al., 2022) (Table 2). The re-analysis of this review complied with the original results, the PI was non-significant following the non-significant pooled effects from both the original results and re-analysis using HKSJ (Fig. 5). Small-study effects was not detected with Egger's test. The actual percentage increase in drop-out over the two pooled review estimates ranged from 47% to 68.2%.

3.3.3. *Side effects*

For patient side-effects only one review (Wei et al., 2017) involving 20 studies and 1353 participants was re-analysed. There was a greater risk of side-effects after re-analysis (RR = 2.14 vs. RR = 1.96) and small-study effects were detected with Egger's test ($p = 0.0196$).

4. Discussion

4.1. *Findings*

We re-analysed 12 meta-analyses testing the efficacy of TMS on MDD for three outcomes (i.e., change in depression scores, remission, and response rates) presented in nine articles (Berlim et al., 2014; Gaynes et al., 2014; Kedzior et al., 2014; Kedzior et al., 2015; Kozel & George, 2002; Leggett et al., 2015; Valiengo et al., 2022; Wei et al., 2017; Zhang et al., 2015). Two meta-analyses assessed acceptability rate presented in two articles (Berlim et al., 2014; Wang et al., 2022) and one meta-analysis investigated the incidence rate of side effects of TMS combined with antidepressants (Wei et al., 2017). Eleven out of the 12 original meta-analyses with efficacy outcomes favoured TMS over sham TMS. One meta-analysis did not observe improvements in response rates when an active TMS paradigm comparator (i.e., a combination of unilateral repetitive-TMS, sham repetitive-TMS, or both) was used as opposed to using an inactive comparator (Zhang et al., 2015). For the side effects outcome, a higher incidence rate was found for the TMS combined with antidepressants group as compared with sham (Wei et al., 2017) (see Fig. 5). Authors of most included meta-analyses interpreted findings in support of TMS being clinically beneficial in improving symptoms of MDD (Berlim et al., 2014; Gaynes et al., 2014; Kedzior et al., 2014; Kozel & George, 2002; Leggett et al., 2015; Valiengo et al., 2022; Wei et al., 2017) (see Figs. 2, 3, 4). While the PIs of most re-analysed meta-analyses showed a substantial portion of participants with MDD likely achieving clinically relevant improvements in their conditions (Turkoz et al., 2021), it remains unclear how much improvement can be expected in future patients since the magnitude of the effect varied substantially across the different re-analysed meta-analyses.

No meta-analysis had applied the HKSJ method in estimating the pooled treatment effect (Table 4), despite this being the recommended method for synthesising primary studies especially in random effects meta-analyses with a low number of studies and in the presence of non-trivial heterogeneity (IntHout et al., 2014). In our re-analyses using the HKSJ method, we found the width of the 95% CI around the pooled estimates was generally wider compared to the width of the 95% CI in the original meta-analyses (Table 4). This suggests the meta-analytic point estimates had lower precision than previous evidence might have suggested (Cumming, 2014). We also calculated the 95% PIs around each pooled estimate and found the width of these intervals exceeded the width of the CI we had calculated using the HKSJ method in our re-analysis (Fig. 2-5). In all meta-analyses, except for one assessing response rate (Gaynes et al., 2014), the PIs crossed over the line of no difference and included values favouring the control condition. Therefore, the PIs cannot exclude the possibility of TMS being associated with high rates of drop-out and side-effects in some

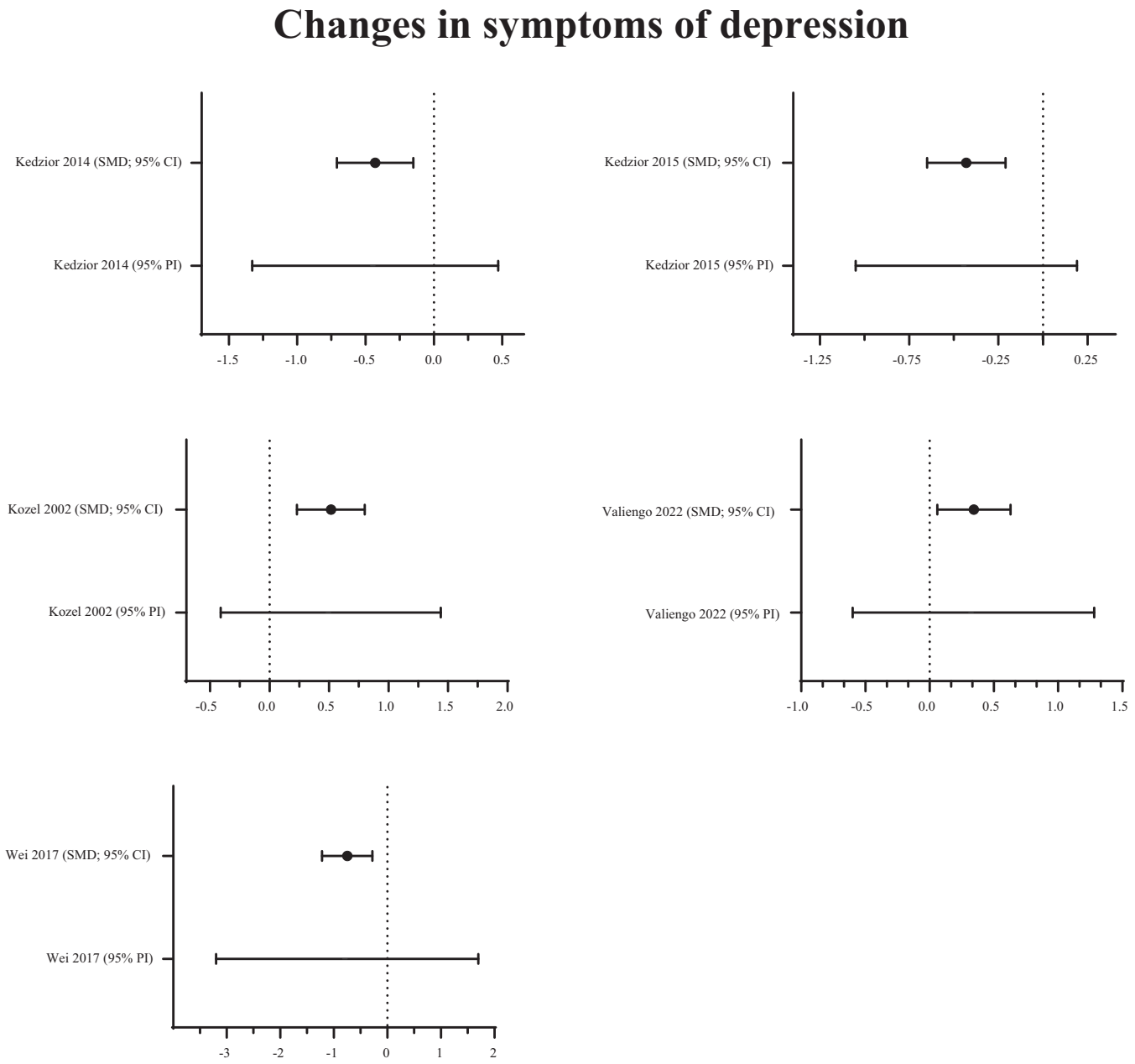

Fig. 2. Forest plots showing the weighted pooled average estimates in standardised mean difference (SMD) and their corresponding 95% confidence intervals (CI) for each study assessing changes in symptoms of depression. Each forest plot also includes the 95% prediction intervals (PI) beneath their corresponding weighted pooled average estimates. The 95% PIs show the distribution of treatment effects across different settings. The stacked vertical line represents the null hypothesis.

populations (IntHout et al., 2016).

Our results are in line with several previous similar studies. One published meta-analysis showed pharmacists' interventions were on average useful in lowering systolic/diastolic blood pressure (Santschi, Chiolero, Burnand, Colosimo, & Paradis, 2011). In a later report, the inclusion of PIs in the same meta-analysis showed the same interventions likely had no effect on diastolic blood pressure in some populations (Chiolero et al., 2012). In another study, authors re-analysed Cochrane meta-analyses exploring whether applying 95% PI would produce different results compared to the original CI of the meta-analyses (IntHout et al., 2016). Out of 479 meta-analyses with non-trivial heterogeneity, in 347 meta-analyses the 95% CI excluded the null indicating a beneficial effect for the average participant. However, in the same meta-analyses, the 95% PI included the null value indicating that some individuals may experience no effect or an adverse outcome relative to controls despite the original CI indicating average benefit (IntHout et al., 2016). In another study, authors re-analysed data from existing meta-analyses of RCTs testing the effects of Nordic hamstring exercise on hamstring injury (Impellizzeri et al., 2021). Similar to our study, authors re-analysed existing meta-analyses in this field by applying the HKSJ method and 95% PI (Impellizzeri et al., 2021). In line with our findings, the authors of the original meta-analyses had focused on average point estimates without accounting for between-study heterogeneity and concluded Nordic hamstring exercise can reduce risk of hamstring injury by 50% (Attar et al., 2017; Raya-Gonzalez, Castillo, & Clemente, 2021; Van Dyk, Behan, & Whiteley, 2019). However, the re-analysis of the same data did not support existing recommendations (Impellizzeri et al., 2021).

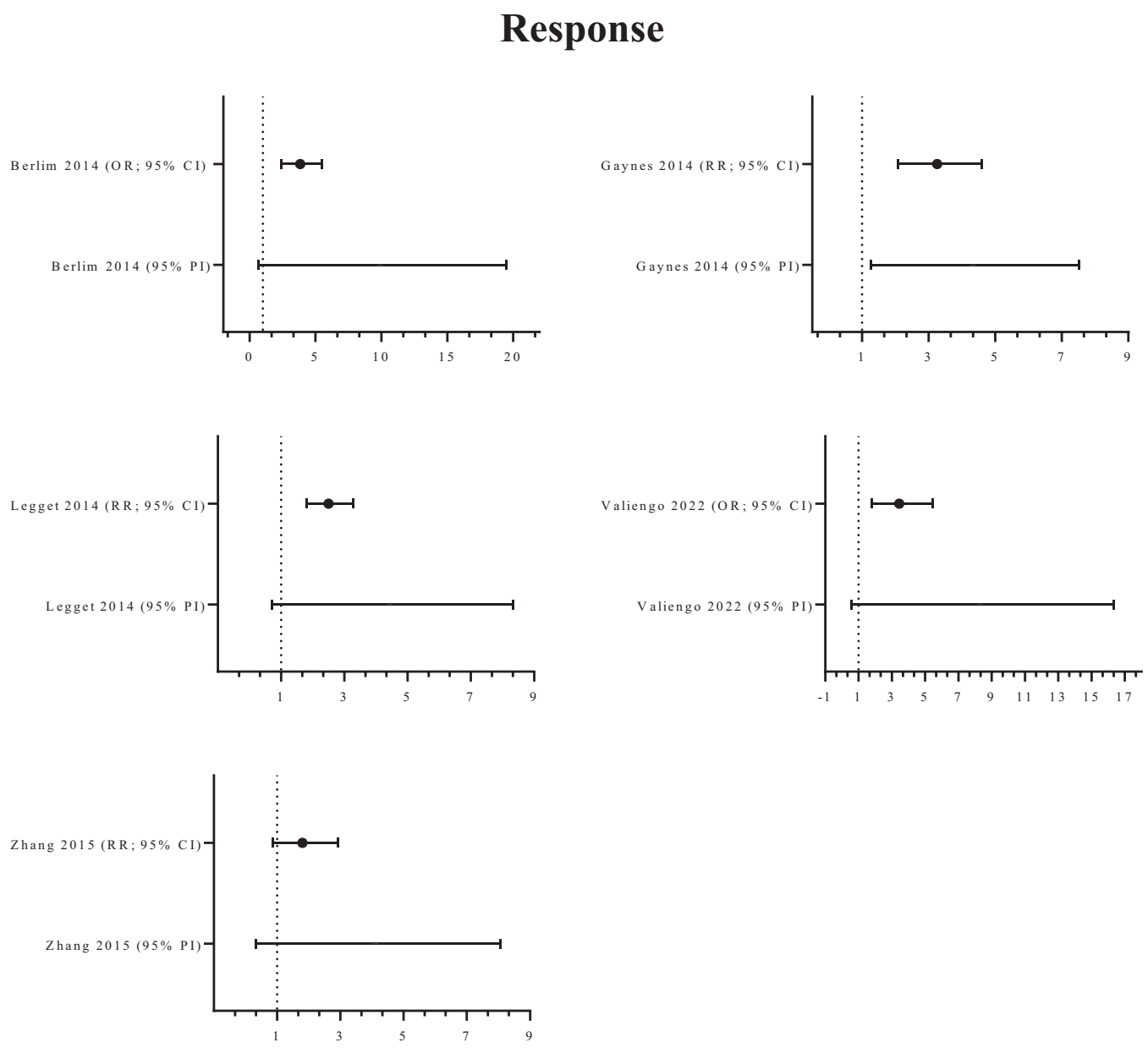

Fig. 3. Forest plots showing the weighted pooled average estimates in odds ratio (OR) relative risk (RR) and their corresponding 95% confidence intervals (CI) for each study assessing response. Each forest plot also includes the 95% prediction intervals (PI) beneath their corresponding weighted pooled average estimates. The 95% PIs show the distribution of treatment effects across different settings. The stacked vertical line represents the null hypothesis.

4.1.1. Heterogeneity

We provide a summary table showing the reasons or lack thereof the authors provided for exploring heterogeneity (Supplementary Table 1). Most authors did not provide sufficient information why heterogeneity had not been explored while others mentioned that presence of heterogeneity did not go beyond what is expected by chance and provided *p* values for Q alongside the I^2 value. These indices, however, are insufficient to comprehensively quantify the amount of heterogeneity, especially in meta-analyses investigating clinical questions that aim to determine in what populations a treatment may be effective, ineffective, or harmful (Borenstein, 2019). Our results, and that of others (Umberham et al., 2017), suggest the lack of comprehensive investigation of heterogeneity is prevalent among published random-effects meta-analyses in the field of TMS for treating MDD.

The majority of included meta-analyses reported I^2 and applied terms including low, medium, or high heterogeneity to describe the magnitude of I^2 from 0 to 100% (Berlim et al., 2014; Gaynes et al., 2014; Kedzior et al., 2014; Kedzior et al., 2015; Kozel & George, 2002; Leggett et al., 2015; Wei et al., 2017; Zhang et al., 2015). However, these terms can be misleading and offer no value to relevant clinical questions (Borenstein, 2019; Borenstein, 2020; Borenstein, Higgins, Hedges, & Rothstein, 2017; IntHout et al., 2016). This is because I^2 is an index of proportion that describes the percentage of heterogeneity that is due to true differences between studies rather than sampling error (Borenstein, 2019; Borenstein et al., 2017). Because of this reason, while "low" heterogeneity could indeed indicate presence of non-trivial heterogeneity, the true effects may still vary widely across different populations. By contrast, in the presence of "high" heterogeneity, the treatment effects may still be consistent across different populations (Borenstein, 2019; Borenstein, 2020). As such, the I^2 does not provide useful clinical information and should necessarily be reported alongside the PIs, which can describe how wide or narrow the distribution of treatment effects

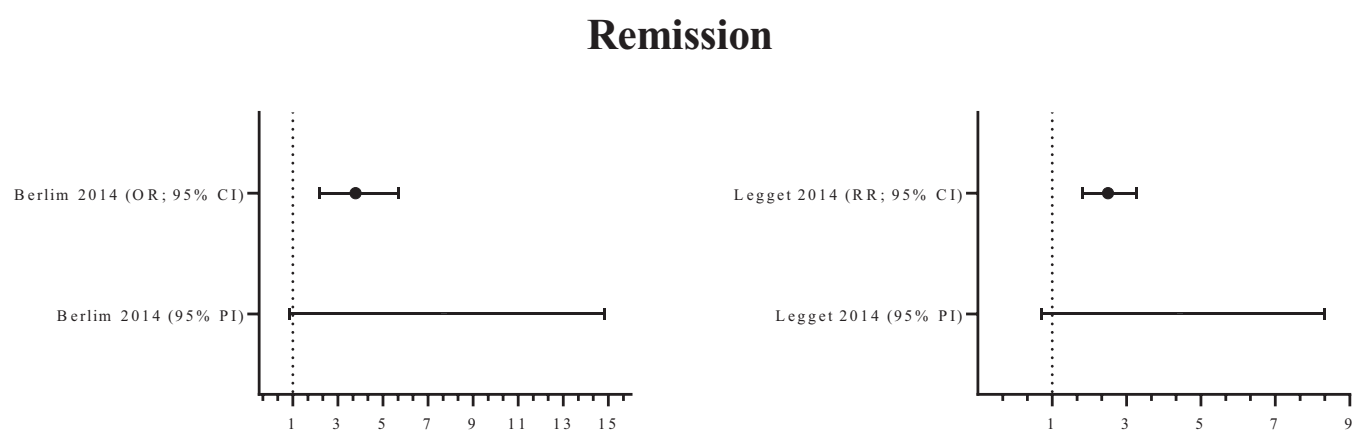

Fig. 4. Forest plots showing the weighted pooled average estimates in odds ratio (OR) or relative risk (RR) and their corresponding 95% confidence intervals (CI) for each study assessing remission from depression. Each forest plot also includes the 95% prediction intervals (PI) beneath their corresponding weighted pooled average estimates. The 95% PIs show the distribution of treatment effects across different settings. The stacked vertical line represents the null hypothesis.

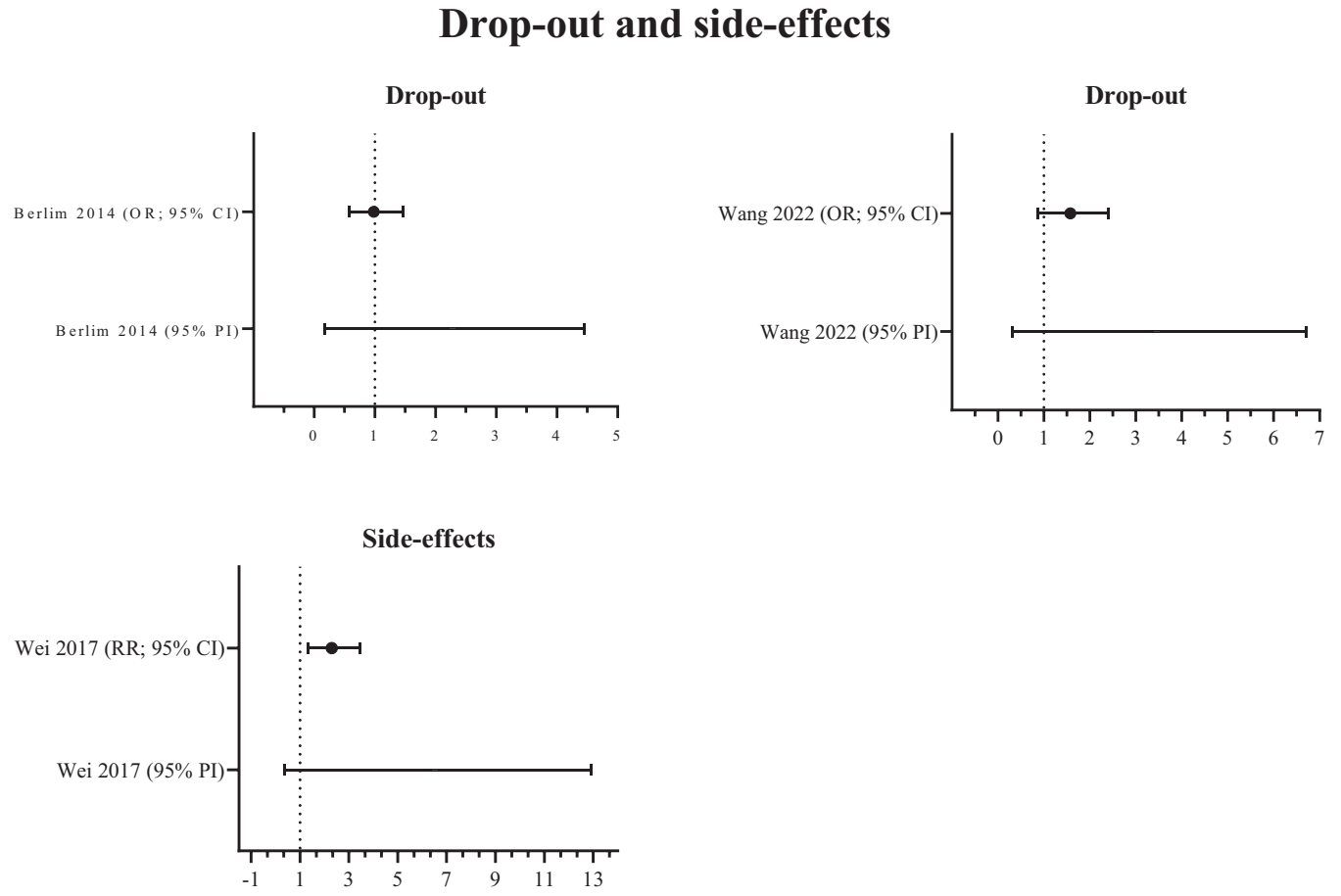

Fig. 5. Forest plots showing the weighted pooled average estimates in odds ratio (OR) or relative risk (RR) and their corresponding 95% confidence intervals (CI) for each study assessing rate of drop-out and side-effects. Each forest plot also includes the 95% prediction intervals (PI) beneath their corresponding weighted pooled average estimates. The 95% PIs show the distribution of treatment effects across different settings. The stacked vertical line represents the null hypothesis.

may be across different settings (Borenstein, 2019; Borenstein et al., 2017; IntHout et al., 2016).

4.2. Risk of bias and small-study effects

Risk of bias assessments showed several systematic reviews were of critically low quality due to more than one flaw in critical items according to the AMSTAR 2 tool (Shea et al., 2017). Specifically, these reviews lacked information about pre-registration of study protocol (which is necessary for reducing biases such as selective reporting), literature searches, and of satisfactory application of risk of bias tools as well as lack of interpreting results in light of risk of bias. Some

systematic reviews did not perform study screening, data extraction, and other relevant methodological processes in pairs of reviewers, which is important to reduce error in extracting data from primary studies (Li, Higgins, & Deeks, 2019). Many of the included reviews poorly reported the details of their selection criteria and characteristics of included studies. While results from our assessment indicated most systematic reviews had adequately discussed and reported heterogeneity, the AMSTAR 2 guidelines do not specifically state what indices of heterogeneity should be reported (Shea et al., 2017), despite clear reporting and discussion of heterogeneity such as including PI are key to the interpretation of meta-analytic results (Deeks et al., 2019). That the AMSTAR 2 revealed most included systematic reviews carried low risk of bias in addressing heterogeneity should be interpreted with caution. Overall, out of the 29 systematic reviews identified, none could be classed as high quality according to the AMSTAR 2 (Shea et al., 2017).

Small-study effects refers to the phenomenon whereby smaller studies sometimes show greater effect sizes than larger studies (Schwarzer, Carpenter, & Rücker, 2015; Sterne, Gavaghan, & Egger, 2000). Meta-analysts can apply Egger's test or visually inspect funnel plots to explore small-study effects within a meta-analysis (Rücker, Carpenter, & Schwarzer, 2011). A significant Egger's test or an asymmetrical funnel plot may provide evidence for presence of small-study effects (Rücker et al., 2011; Sterne & Harbord, 2004). These tests, however, are often erroneously used to test for publication bias (Ioannidis, 2008; Lau et al., 2006; Sterne et al., 2011; Sterne & Harbord, 2004). This practice is erroneous because Egger's test and funnel plots cannot directly test for publication bias, rather they can test for small-study effects (Borenstein, 2019). Publication bias is one of many possible sources of small-study effects; others may include but are not limited to selective reporting or presence of non-trivial heterogeneity (Egger, Smith, Schneider, & Minder, 1997; Ioannidis, 2008; Sterne & Harbord, 2004; Sterne et al., 2011). We found most meta-analyses did not specifically investigate small-study effects but applied Egger's test or visually inspected funnel plots to explore publication bias.

Results from the AMSTAR 2 revealed multiple systematic reviews did not adequately address small-study effects (Table 3). We analysed small-study effects using Egger's test in meta-analyses that included 10 or more RCTs (Lau et al., 2006). Five meta-analyses met this criterion and Egger's test revealed that within these meta-analyses the RCTs with smaller samples showed larger effect sizes than RCTs with larger samples indicating that some sources of bias such as publication bias or selective reporting may have biased the treatment effects (Lau et al., 2006). That selective reporting in this field may be one source of small-study effects is not surprising since our quality assessment revealed a lack of information regarding pre-registration of study protocol for several of our included systematic reviews. Moreover, small sample sizes tend to inflate the effect size and increase the risk of type I error (Ioannidis, 2005; Pereira, Horwitz, & Ioannidis, 2012). Sample sizes tend to be small in neuroscience (Button et al., 2013), psychology and cognitive neuroscience (Szucs & Ioannidis, 2017; Szucs & Ioannidis, 2021), and the possible presence of excess of significant findings in the field of TMS research in neuropsychiatric disorders (Amad et al., 2019) may be one source of small-study effects.

4.3. Limitations

We acknowledge there are several limitations to our study. We could not further explore study-level factors that might have contributed to the unexplained heterogeneity. This is because our included systematic reviews did not provide sufficient data to perform subgroup analyses or meta-regression. We also did not narratively discuss potential sub-group differences (e.g., differences between men and women, severity in MDD, etc.) that might have explained differences in the magnitude and direction in the meta-analytic point-estimates. The lack of sufficient demographic and clinical information about participants in the included systematic reviews would have rendered a narrative discussion of

Table 3
Results of methodological quality using the A MeaSurement Tool to Assess Reviews tool (AMSTAR 2) of included systematic reviews.

Study	1	2	3	4	5	6	7	8	9 a (RCT)	9b	10	11	12	13	14	15	16	No. of [illegible]	No. of [illegible]	Overall
(Berlim et al., 2013a)										*								5	6	CL
(Berlim et al., 2013b)										*								5	5	CL
(Berlim et al., 2014)										*								6	5	CL
(Chen et al., 2013)										*								5	5	CL
(Chen et al., 2017)										*								6	5	CL
(Chu et al., 2021)										*								6	5	CL
(Couturier, 2005)										*								5	6	CL
(Gaynes et al., 2014)										*								6	7	CL
(Gellersen & Kedzior, 2019)										*								5	5	CL
(Hung et al., 2020)										*								6	5	CL
(Kedzior et al., 2014)										*								5	7	CL
(Kedzior et al., 2015)										*								4	5	CL
(Kozel & George, 2002)										*								5	5	CL
(Lam et al., 2008)										*								3	5	CL
(Leggett et al., 2015)										*								5	4	CL
(Liu et al., 2014)										*								5	6	CL
(Martin et al., 2003)										*								4	5	CL
(McNamara et al., 2001)										*								5	7	CL
(Rodriguez-Martin et al., 2002)										*								1	2	L
(Schutter, 2009)										*								6	6	CL
(Schutter, 2010)										*								6	6	CL
(Slotema et al., 2010)										*								6	6	CL
(Sonmez et al., 2019)										*								6	6	CL
(Valiengo et al., 2022)										*								5	3	CL
(Voigt et al., 2021)										*								4	4	CL
(Wang et al., 2022)										*								3	2	CL
(Wei et al., 2017)										*								5	5	CL
(Xie et al., 2013)										*								6	4	CL
(Zhang et al., 2015)										*								3	4	CL

[1] (No or Partial Yes response for Q2, Q4, Q7, Q9, Q11, Q13, Q15); [2] No or Partial Yes for Q1, Q3, Q5, Q6, Q8, Q10, Q12, Q14, Q16
* only RCTs
Items
1. Did the research questions and inclusion criteria for the review include the components of PICO?
2. Did the report of the review contain an explicit statement that the review methods were established prior to conduct of the review and did the report justify any significant deviations from the protocol?*
3. Did the review authors explain their selection of the study designs for inclusion in the review?
4. Did the review authors use a comprehensive literature search strategy? *
5. Did the review authors perform study selection in duplicate?
6. Did the review authors perform data extraction in duplicate?
7. Did the review authors provide a list of excluded studies and justify the exclusions? *
8. Did the review authors describe the included studies in adequate detail?
9a. Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?* RCT
9b. Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review? nonRCT*
10. Did the review authors report on the sources of funding for the studies included in the review?
11a. If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results? *RCT
11b. If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?*
12. If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?
13. Did the review authors account for RoB in individual studies when interpreting/ discussing the results of the review? *
14. Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?
15. If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review? *
16. Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?
* Critical domain
Yes
H: High
M: Moderate
L: Low
CL: Critically low

Table 4
Comparison of the original published results with re-analysis of each meta-analysis including the prediction intervals for efficacy, acceptability, and side effects outcomes.

					Results of original published meta-analysis			Re-analysis using original method			Re-analysis with HKSJ and 95% CI and prediction interval (PI), I^2, τ^2, and assessment of Small-Study effects		
Meta-analysis	Outcome	No. of studies	Total no. of patients	Effect measure	Effect estimate	CI	I^2 (τ^2)	Meta-analytic method used	Effect estimate 95% CI	I^2 (τ^2)	Effect estimate, with (95% CI) [PI] and % change based on outcome	I^2 (95% CI) (τ^{2*})	Assessment of small-study effects
Efficacy													
(Berlim et al., 2014)	Response	29	1371	OR	3.31	2.35, 4.64	2.97% (NR)	DL-REs	3.31 (2.35, 4.64)	3% (0.027)	3.62 (2.43, 5.48) [0.68, 19.51] 78.4% (70.8, 84.6) %	3 (0, 43) % (0.63)	$P < 0.0001$
(Berlim et al., 2014)	Remission	15	975	OR	3.30	2.04, 5.33	0% (NR)	DL-REs	3.30 (2.04, 5.33)	0% (0)	3.50 (2.16, 5.68) [0.83, 14.81] 77.8% (68.4, 85.0) %	0 (0, 54) % (0.39)	$P = 0.395$
(Gaynes et al., 2014)	Response	14	653	RR	3.38	2.24, 5.10	0% (NR)	NR (but assumed they had used DL-REs)	3.109 (1.99, 4.86)	0% (0)	3.09 (2.07, 4.60) [1.27, 7.52] 75.6% (67.4, 82.1) %	0 (0, 55) % (0.12)	$P = 0.092$
(Kedzior et al., 2014)	Change in depression	14	659	SMD	−0.42	−0.66, −0.18	50% (NR)	IV-RE	−0.43 (−0.67, −0.19)	53.5% (0.11)	−0.43 (−0.71, −0.15) [−1.33, 0.47]	54 (16, 75) % (0.15)	$P = 0.042$
(Kedzior et al., 2015)	Change in depression	16	496	SMD	−0.48	−0.70, −0.25	27% (NR)	IV-RE	−0.43 (−0.65, −0.22)	25.3% (0.045)	−0.43 (−0.65, −0.21) [−1.05, 0.19]	25 (0, 59) % (0.07)	$P = 0.443$
(Kozel & George, 2002)	Change in depression	12	230	cumulative effect size of Hedges' d	0.53	0.24, 0.82	NR (NR)	NR but assumed MH-FEs	0.52 (0.25, 0.78)	80.3% (0.1627)	0.52 (0.23, 0.80) [−0.41, 1.44]	80 (66, 88) % (0.15)	NA
(Leggett et al., 2015)	Response	31	1377	RR	2.35	1.70, 3.25	36.1% (NR)	NR but assumed DL-REs	2.25 (1.67, 3.03)	26.5% (0.154)	2.43 (1.81, 3.27) [0.71, 8.33] 70.8% (64.4, 76.6) %	27 (0, 53) % (0.34)	$P < 0.0001$
(Leggett et al., 2015)	Remission	18	1010	RR	2.24	1.53, 3.27	1.1% (NR)	NR but assumed DL-REs	2.23 (1.54, 3.23)	0% (0)	2.73 (1.73, 4.31) [0.66, 11.32] 73.2% (63.3, 81.2) %	0 (0, 50) % (0.40)	$P = 0.0298$
(Valiengo et al., 2022)	Change in depression	12	528	SMD	0.36	0.13, 0.60	28% (0.04)	IV-RE	0.36 (0.11, 0.60)	39% (0.06)	0.34 (0.06, 0.63) [−0.60, 1.28]	39 (0, 69) % (0.16)	$P = 0.8559$
(Valiengo et al., 2022)	Response	14	563	OR	3.26	2.11, 5.04	0% (0.00)	MH-RE	3.24 (2.10, 5.01)	0% (0.00)	3.12 (1.79, 5.44) [0.60, 16.31] 75.7% (64.1, 84.5) %	0 (0, 55) % (0.51)	$P = 0.6528$
(Wei et al., 2017)	Change in depression	29	1659	SMD	−0.84	−1.19, −0.48	90% (0.81)	IV-REs	−0.75 (−1.07, −0.42)	89.4% (0.675)	−0.75 (−1.22, −0.28) [−3.20, 1.70]	89 (85, 92) % (1.37)	$P = 0.239$
(Zhang et al., 2015)	Response	10	623	RR	1.5	0.91, 2.47	57% (0.27)	MH DL-REs	1.46 (0.91, 2.34)	52.2% (0.223)	1.60 (0.88, 2.92) [0.32, 8.06] 61.5% (46.8, 74.5) %	52 (1.5, 77) % (0.42)	$P = 0.0163$
Acceptability													
(Berlim et al., 2014)	Drop-out	22	1191	OR	0.97	0.61, 1.52	0% (NR)	DL-REs	0.97 (0.61, 1.54)	0% (0)	0.91 (0.57, 1.46) [0.18, 4.46] 47.6% (36.3, 59.3) %	0 (0, 46) % (0.53)	$P = 0.393$
(Wang et al., 2022)	Drop-out	18	2021	OR	1.44	0.83, 2.52	0% (0.00)	MH-RE	1.44 (0.83, 2.52)	0% (0.00)	1.45 (0.87, 2.41) [0.31, 6.72] 59.2% (46.5, 59.3) %	0% (0.47)	$P = 0.8284$
Side effect													
(Wei et al., 2017)	Side effect	20	1353	RR	1.96	1.47, 2.61	38% (NR)	MH-FE	1.98 (1.61, 2.44)	78% (0.92)	2.14 (1.33, 3.44) [0.35, 12.93] 68.2% (57.1, 77.5) %	78 (67, 86) % (0.68)	P = 0.0196

MH: Mantel-Haenszel pooling (Method of choice unless otherwise stated); IV: Inverse Variance Pooling; DL-REs: DerSimonian-Laird- Random-Effects; IV-FEs: Inverse-Variance-Fixed-Effects; HKSJ: Hartung-Knapp-Sidik-Jonkman

Values in bold are statistically significant at the alpha level of 0.05; OR: odd ratio; RR: relative risk; SMD: standardised mean difference; NR: not reported; CI: confidence intervals; PI: prediction intervals

* Hartung- Knapp-Sidik-Jonkman method for random-effects meta-analysis was used with tau estimated using 'Sidik-Jonkman' approach in the R function 'metagen'

potential subgroup differences limited. In addition, we only re-analysed meta-analyses including at least 10 primary randomised controlled trials because this is the least number of studies needed to generate reliable PI estimates (Borenstein, 2019; Borenstein, 2020). Because of this reason, we drew a smaller number of meta-analyses from all the available evidence in TMS literature for mood disorders. This might have potentially introduced bias in our conclusions.

4.4. Implications and suggestions for future research

The clinical management of MDD is complex with several first-line treatments currently available including psychotherapy and pharmacotherapy with the application of TMS for MDD typically considered for cases of MDD-treatment resistant (Kiebs, Hurlemann, & Mutz, 2019; Lefaucheur et al., 2020; Malhi, Bell, Bassett, et al., 2021b; Milev et al., 2016). The selection and implementation of the most appropriate treatment for the management of mood disorders and indeed MDD should align with the clinical features the patient is manifesting. Tailoring the treatment of MDD with the individual's clinical profile is complicated by the multifaceted pathophysiological nature of mood disorders and the individual's demographic characteristics. For example, psychotherapy may be more appropriate for patients whose mood disorder likely arose from a traumatic life event where habitual negative thinking exacerbates depressive symptoms (Malhi et al., 2021).

Selecting the correct treatment for MDD therefore, requires a thorough assessment of the biopsychosocial mechanisms that may be responsible for the clinical manifestations of the mood disorder (Malhi, Bell, Murray, et al., 2021a; Malhi, Bell, Outhred, et al., 2021). The available meta-analyses, however, do not provide sufficient information that would allow clinicians to discern at which stage to position TMS in the clinical management of MDD (Malhi, Bell, Outhred, et al., 2021). This is because primary studies in this field have not provided sufficient demographic and clinical information (Malhi, Bell, Mannie, et al., 2021) that would allow meta-analysts to critically explore presence of heterogeneity when pooling treatment effects from primary studies. This, in turn, can result in authors of meta-analyses relying on the average pooled treatment effects, which by default, precludes the interpretation of meta-analytic results to all populations, therefore limiting the external validity and clinical utility of the findings (Borenstein, 2019).

Creating homogenous groups of patients based on demographic and clinical factors including age, sex, disease severity, brain chemistry and electrophysiology would allow clinicians to biotype patients and personalise TMS interventions to achieve greater efficacy and reduce risk of harm (Modak & Fitzgerald, 2021). For example, an individual's neural signature may be used in selecting patients with MDD who are more likely to benefit from TMS (Modak & Fitzgerald, 2021). Different neural connectivity patterns in the frontostriatal and limbic brain systems may predict which patients with depression may derive greater benefit from receiving TMS. In one study, a larger proportion of individuals (82.5%) with depression classed as biotype 1, as opposed to biotype 2, 3, and 4, experienced >25% improvement on the Hamilton Depression Rating Scale after receiving TMS on the dorsomedial prefrontal cortex (Drysdale et al., 2017). The available evidence from our included meta-analyses and primary RCTs however, do not currently provide such levels of demographic or clinical detail necessary to biotype patients and personalise TMS for treating MDD. In the absence of evidence necessary to tailor the delivery of TMS to patients with greater probability of achieving clinically relevant outcomes, clinicians may wish to adopt alternative treatments such as electroconvulsive therapy (Espinoza & Kellner, 2022), ketamine or esketamine (McIntyre et al., 2021), or antidepressants (Suchting et al., 2021) for treating MDD.

4.5. Overall interpretation and conclusion

The PIs describe the magnitude of the effect across different settings and provide information on what effect may be expected in future patients (Borenstein, 2019; Borenstein, 2020; Guddat et al., 2012; IntHout et al., 2016). Clinicians are interested in knowing whether a treatment effect is consistently beneficial across different populations or whether it is null or harmful in some individuals. The weighted pooled average of meta-analysis and its CIs cannot address this relevant clinical question, while PIs can provide information about the extent to which the treatment may in some cases produce no benefit or harm in some patients. We found that TMS is likely ineffective in treating MDD in some populations and may even be associated with the opposite intended effect such as elevated symptoms of MDD. Moreover, the critically low methodological quality in the included systematic reviews questions whether the observed treatment effect estimates in the original meta-analyses are close to the true value. The view that substantial heterogeneity such as differences in study methodology, underpowered trials, and excess significance in this field of research are serious limitations that question the extent to which TMS should be applied to the average patient to treat MDD is shared by some authors (Malhi, Bell, Outhred, et al., 2021). We also share these concerns and argue that future research should address the observed limitations while aiming to identify which populations with MDD are more likely to benefit from TMS and which are less likely or are at risk of harmful effects.

Declaration of Competing Interest

The authors declare no conflict of interest

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cpr.2022.102236.

References

Amad, A., & Fovet, T. (2022). rTMS for depression: The difficult transition from research to clinical practice. *Australian and New Zealand Journal of Psychiatry, 56*(1), 14–15.
Amad, A., Jardri, R., Rousseau, C., Larochelle, Y., Ioannidis, J. P. A., & Naudet, F. (2019). Excess significance bias in repetitive transcranial magnetic stimulation literature for neuropsychiatric disorders. *Psychotherapy and Psychosomatics, 88*(6), 363–370.
American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.
Attar, A., Wesam Saleh, A., Soomro, N., Sinclair, P. J., Pappas, E., & Sanders, R. H. (2017). Effect of injury prevention programs that include the Nordic hamstring exercise on hamstring injury rates in soccer players: A systematic review and meta-analysis. *Sports Medicine, 47*(5), 907–916.
Aytug, Z. G., Rothstein, H. R., Zhou, W., & Kern, M. C. (2012). Revealed or concealed? Transparency of procedures, decisions, and judgment calls in meta-analyses. *Organizational Research Methods, 15*(1), 103–133.
Beck, A. T., Steer, R. A., & Brown, G. (1996). Beck depression inventory–II. *Psychological Assessment*.
Berlim, M. T., Van den Eynde, F., & Daskalakis, Z. J. (2013a). A systematic review and meta-analysis on the efficacy and acceptability of bilateral repetitive transcranial magnetic stimulation (rTMS) for treating major depression. *Psychological Medicine, 43*(11), 2245.
Berlim, M. T., Van den Eynde, F., & Daskalakis, Z. J. (2013b). Clinically meaningful efficacy and acceptability of low-frequency repetitive transcranial magnetic stimulation (rTMS) for treating primary major depression: A meta-analysis of randomized, double-blind and sham-controlled trials. *Neuropsychopharmacology, 38*(4), 543–551.
Berlim, M. T., Van den Eynde, F., Tovar-Perdomo, S., & Daskalakis, Z. J. (2014). Response, remission and drop-out rates following high-frequency repetitive transcranial magnetic stimulation (rTMS) for treating major depression: A systematic review and meta-analysis of randomized, double-blind and sham-controlled trials. *Psychological Medicine, 44*(2), 225.
Bolu, A., Gündoğmuş, İ., Aydın, M. S., Fadıloğlu, D., Erken, Y., & Uzun, Ö. (2021). Ten years' data of transcranial magnetic stimulation (TMS): A naturalistic, observational study outcome in clinical practice. *Psychiatry Research, 301*, Article 113986.
Borenstein, M. (2019). *Common mistakes in meta-analysis and how to avoid them*. Biostat, Incorporated.
Borenstein, M. (2020). Research note: In a meta-analysis, the I 2 index does not tell us how much the effect size varies across studies. *Journal of Physiotherapy, 66*(2), 135–139.

Borenstein, M., Higgins, J. P. T., Hedges, L. V., & Rothstein, H. R. (2017). Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Research Synthesis Methods, 8*(1), 5–18.
Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365–376.
Carlson, K. D., & Ji, F. X. (2011). Citing and building on meta-analytic findings: A review and recommendations. *Organizational Research Methods, 14*(4), 696–717.
Carpenter, L. L., Janicak, P. G., Aaronson, S. T., Boyadjis, T., Brock, D. G., Cook, I. A., … Demitrack, M. A. (2012). Transcranial magnetic stimulation (TMS) for major depression: A multisite, naturalistic, observational study of acute treatment outcomes in clinical practice. *Depression and Anxiety, 29*(7), 587–596.
Chen, J., Zhao, L., Liu, Y., Fan, S., & Xie, P. (2017). Comparative efficacy and acceptability of electroconvulsive therapy versus repetitive transcranial magnetic stimulation for major depression: A systematic review and multiple-treatments meta-analysis. *Behavioural Brain Research, 320*, 30–36.
Chen, J., Zhou, C., Wu, B., Wang, Y., Li, Q., Wei, Y., … Xie, P. (2013). Left versus right repetitive transcranial magnetic stimulation in treating major depression: A meta-analysis of randomised controlled trials. *Psychiatry Research, 210*(3), 1260–1264.
Chiolero, A., Santschi, V., Burnand, B., Platt, R. W., & Paradis, G. (2012). Meta-analyses: With confidence or prediction intervals? *European Journal of Epidemiology, 27*(10), 823–825.
Chu, H., Cheng, C., Liang, C., Chang, W., Juan, C., Huang, Y., … Chen, M. (2021). Efficacy and tolerability of theta-burst stimulation for major depression: A systematic review and meta-analysis. *Progress in Neuro-Psychopharmacology and Biological Psychiatry, 106*, Article 110168.
Couturier, J. L. (2005). Efficacy of rapid-rate repetitive transcranial magnetic stimulation in the treatment of depression: A systematic review and meta-analysis. *Journal of Psychiatry and Neuroscience, 30*(2), 83.
Cumming, G. (2014). The new statistics: Why and how. *Psychological Science, 25*(1), 7–29.
Deeks, J. J., Higgins, J. P., Altman, D. G., & Cochrane Statistical Methods Group. (2019). Analysing data and undertaking meta-analyses. *Cochrane Handbook for Systematic Reviews of Interventions*, 241–284.
DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials, 7*(3), 177–188.
Dowling, N. L., Bonwick, R., Dharwadkar, N. P., & Ng, C. H. (2020). Repetitive transcranial magnetic stimulation for major depression: A naturalistic observational study in an Australian private hospital. *Psychiatry Research, 291*, Article 113275.
Drysdale, A. T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., … Etkin, A. (2017). Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature Medicine, 23*(1), 28–38.
Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *Bmj, 315*(7109), 629–634.
Espinoza, R. T., & Kellner, C. H. (2022). Electroconvulsive therapy. *New England Journal of Medicine, 386*(7), 667–672.
Fava, M. (2003). Diagnosis and definition of treatment-resistant depression. *Biological Psychiatry, 53*(8), 649–659.
Fekadu, A., Donocik, J. G., & Cleare, A. J. (2018). Standardisation framework for the Maudsley staging method for treatment resistance in depression. *BMC Psychiatry, 18*(1), 1–13.
Fitzgerald, P. B., George, M. S., & Pridmore, S. (2021). The evidence is in: Repetitive transcranial magnetic stimulation is an effective, safe and well-tolerated treatment for patients with major depressive disorder. *Australian and New Zealand Journal of Psychiatry*. https://doi.org/10.1177/00048674211043047, 486742110430–48674211043047.
Fitzgerald, P. B., Gill, S., Breakspear, M., Kulkarni, J., Chen, L., Pridmore, S., Purushothaman, S., Galletly, C., Clarke, P., & Ng, F. (2021). Revisiting the effectiveness of repetitive transcranial magnetic stimulation treatment in depression, again. *The Australian and New Zealand Journal of Psychiatry, 00048674211068788*.
Fitzgerald, P. B., Gill, S., Hussain, S., Sarma, S., Chamoli, S., Weiss, A., Garside, D., Purushothaman, S., Fasnacht, M., Simpson, B., Csizmadia, T., Dean, C., & Loo, C. (2021). The place of non-invasive brain stimulation in the RANZCP clinical practice guidelines for mood disorders. *Australian and New Zealand Journal of Psychiatry, 55*(4), 349–354.
Follmann, D. A., & Proschan, M. A. (1999). Valid inference in random effects meta-analysis. *Biometrics, 55*(3), 732–737.
Gaynes, B. N., Lloyd, S. W., Lux, L., Gartlehner, G., Hansen, R. A., Brode, S., … Lohr, K. N. (2014). Repetitive transcranial magnetic stimulation for treatment-resistant depression: A systematic review and meta-analysis. *The Journal of Clinical Psychiatry, 75*(5), 477–489.
Gellersen, H. M., & Kedzior, K. K. (2019). Antidepressant outcomes of high-frequency repetitive transcranial magnetic stimulation (rTMS) with F8-coil and deep transcranial magnetic stimulation (DTMS) with H1-coil in major depression: A systematic review and meta-analysis. *BMC Psychiatry, 19*.
Graham, P. L., & Moran, J. L. (2012). Robust meta-analytic conclusions mandate the provision of prediction intervals in meta-analysis summaries. *Journal of Clinical Epidemiology, 65*(5), 503–510.
Guddat, C., Grouven, U., Bender, R., & Skipka, G. (2012). A note on the graphical presentation of prediction intervals in random-effects meta-analyses. *Systematic Reviews, 1*(1), 1–5.
Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry, 23*(1), 56.
Hedges, L. V., & Vevea, J. L. (1998). Fixed-and random-effects models in meta-analysis. *Psychological Methods, 3*(4), 486.
Higgins, J. P., Li, T., & Deeks, J. J. (2019). Choosing effect measures and computing estimates of effect. In *Cochrane handbook for systematic reviews of interventions* (pp. 143–176).
Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane handbook for systematic reviews of interventions*. John Wiley & Sons.
Hung, Y. Y., Yang, L. H., Stubbs, B., Li, D. J., Tseng, P. T., Yeh, T. C., … Chu, C. S. (2020). Efficacy and tolerability of deep transcranial magnetic stimulation for treatment-resistant depression: A systematic review and meta-analysis. *Progress in Neuro-Psychopharmacology & Biological Psychiatry, 99*, Article 109850.
Impellizzeri, F. M., McCall, A., & van Smeden, M. (2021). Why methods matter in a meta-analysis: A reappraisal showed inconclusive injury preventive effect of Nordic hamstring exercise. *Journal of Clinical Epidemiology, 140*, 111–124.
IntHout, J., Ioannidis, J. P., & Borm, G. F. (2014). The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Medical Research Methodology, 14*(1), 1–12.
IntHout, J., Ioannidis, J. P. A., Rovers, M. M., & Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open, 6*(7), Article e010247. https://doi.org/10.1136/bmjopen-2015-010247.
Ioannidis, J. P. (2005). Why most published research findings are false. *PLoS Medicine, 2*(8), Article e124.
Ioannidis, J. P. A. (2008). Interpretation of tests of heterogeneity and bias in meta-analysis. *Journal of Evaluation in Clinical Practice, 14*(5), 951–957.
Janicak, P. G., Dunner, D. L., Aaronson, S. T., Carpenter, L. L., Boyadjis, T. A., Brock, D. G., … Bonneh-Barkay, D. (2013). Transcranial magnetic stimulation (TMS) for major depression: A multisite, naturalistic, observational study of quality of life outcome measures in clinical practice. *CNS Spectrums, 18*(6), 322–332.
Kedzior, K. K., Azorina, V., & Reitz, S. K. (2014). More female patients and fewer stimuli per session are associated with the short-term antidepressant properties of repetitive transcranial magnetic stimulation (rTMS): A meta-analysis of 54 sham-controlled studies published between 1997–2013. *Neuropsychiatric Disease and Treatment, 10*, 727.
Kedzior, K. K., Reitz, S. K., Azorina, V., & Loo, C. (2015). Durability of the antidepressant effect of the high-frequency repetitive transcranial magnetic stimulation (rTMS) in the absence of maintenance treatment in major depression: A systematic review and meta-analysis of 16 double-blind, randomized, sham-controlled trials. *Depression and Anxiety, 32*(3), 193–203.
Kelley, G. A., & Kelley, K. S. (2009). Impact of progressive resistance training on lipids and lipoproteins in adults: A meta-analysis of randomized controlled trials. *Preventive Medicine, 48*(1), 9–19.
Kiebs, M., Hurlemann, R., & Mutz, J. (2019). Repetitive transcranial magnetic stimulation in non-treatment-resistant depression. *The British Journal of Psychiatry, 215*(2), 445–446.
Kozel, F. A., & George, M. S. (2002). Meta-analysis of left prefrontal repetitive transcranial magnetic stimulation (rTMS) to treat depression. *Journal of Psychiatric Practice, 8*(5), 270–275.
Lam, R. W., Chan, P., Wilkins-Ho, M., & Yatham, L. N. (2008). Repetitive transcranial magnetic stimulation for treatment-resistant depression: A systematic review and metaanalysis. *The Canadian Journal of Psychiatry, 53*(9), 621–631.
Langan, D., Higgins, J. P., Jackson, D., Bowden, J., Veroniki, A. A., Kontopantelis, E., … Simmonds, M. (2019). A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Research Synthesis Methods, 10*(1), 83–98.
Lau, J., Ioannidis, J. P. A., Terrin, N., Schmid, C. H., & Olkin, I. (2006). The case of the misleading funnel plot. *Bmj, 333*(7568), 597.
Leeper, T. J. (2021). *Margins: Marginal effects for model objects*.
Lefaucheur, J., Aleman, A., Baeken, C., Benninger, D. H., Brunelin, J., Di Lazzaro, V., … Hummel, F. C. (2020). Evidence-based guidelines on the therapeutic use of repetitive transcranial magnetic stimulation (rTMS): An update (2014–2018). *Clinical Neurophysiology, 131*(2), 474–528.
Lefebvre, C., Glanville, J., Briscoe, S., Littlewood, A., Marshall, C., Metzendorf, M., Noel-Storr, A., Rader, T., Shokraneh, F., & Thomas, J. (2019). Searching for and selecting studies. *Cochrane Handbook for Systematic Reviews of Interventions*, 67–107.
Leggett, L. E., Soril, L. J., Coward, S., Lorenzetti, D. L., MacKean, G., & Clement, F. M. (2015). Repetitive transcranial magnetic stimulation for treatment-resistant depression in adult and youth populations: A systematic literature review and meta-analysis. *The Primary Care Companion for CNS Disorders, 17*(6).
Li, T., Higgins, J. P., & Deeks, J. J. (2019). Collecting data. In *Cochrane Handbook for Systematic Reviews of Interventions* (pp. 109–141).
Liu, B., Zhang, Y., Zhang, L., & Li, L. (2014). Repetitive transcranial magnetic stimulation as an augmentative strategy for treatment-resistant depression, a meta-analysis of randomized, double-blind and sham-controlled study. *BMC Psychiatry, 14*(1), 1–9.
Malhi, G. S., Bell, E., Bassett, D., Boyce, P., Bryant, R., Hazell, P., Hopwood, M., Lyndon, B., Mulder, R., & Porter, R. (2021b). The 2020 Royal Australian and New Zealand College of Psychiatrists clinical practice guidelines for mood disorders. *The Australian and New Zealand Journal of Psychiatry, 55*(1), 7–117.
Malhi, G. S., Bell, E., Mannie, Z., Bassett, D., Boyce, P., Hopwood, M., … Lyndon, B. (2021). Profiling rTMS: A critical response. *The Australian and New Zealand Journal of Psychiatry, 55*(4), 355–365.
Malhi, G. S., Bell, E., Murray, G., Boyce, P., Lyndon, B., Bassett, D., Mulder, R., Hopwood, M., Porter, R., & Singh, A. B. (2021a). The positioning of rTMS. *Australian and New Zealand Journal of Psychiatry, 55*(2), 125–128.
Malhi, G. S., Bell, E., Outhred, T., Singh, A. B., Hopwood, M., Mulder, R., … Mannie, Z. (2021). Is rTMS ready for primetime? *Canadian Journal of Psychiatry, 66*(10), 873–877.
Martin, J. L. R., Barbanoj, M. J., Schlaepfer, T. E., Thompson, E., Pérez, V., & Kulisevsky, J. (2003). Repetitive transcranial magnetic stimulation for the treatment

of depression: Systematic review and meta-analysis. *The British Journal of Psychiatry, 182*(6), 480–491.
McClintock, S. M., Reti, I. M., Carpenter, L. L., McDonald, W. M., Dubin, M., Taylor, S. F., … National Network of Depression Centers rTMS, Task Group, & American Psychiatric Association Council on Research Task Force on Novel Biomarkers, and Treatments. (2018). Consensus recommendations for the clinical application of repetitive transcranial magnetic stimulation (rTMS) in the treatment of depression. *The Journal of Clinical Psychiatry, 79*(1).
McIntyre, R. S., Rosenblat, J. D., Nemeroff, C. B., Sanacora, G., Murrough, J. W., Berk, M., … Ho, R. (2021). Synthesizing the evidence for ketamine and esketamine in treatment-resistant depression: An international expert opinion on the available evidence and implementation. *American Journal of Psychiatry, 178*(5), 383–399.
McNamara, B., Ray, J. L., Arthurs, O. J., & Boniface, S. (2001). Transcranial magnetic stimulation for depression and other psychiatric disorders. *Psychological Medicine, 31*(7), 1141.
Milev, R. V., Giacobbe, P., Kennedy, S. H., Blumberger, D. M., Daskalakis, Z. J., Downar, J., … Lam, R. W. (2016). Canadian Network for Mood and Anxiety Treatments (CANMAT) 2016 clinical guidelines for the management of adults with major depressive disorder: Section 4. neurostimulation treatments. *Canadian Journal of Psychiatry, 61*(9), 561–575.
Modak, A., & Fitzgerald, P. B. (2021). Personalising transcranial magnetic stimulation for depression using neuroimaging: A systematic review. *The World Journal of Biological Psychiatry, 22*(9), 647–669.
Moher, D. (1995). Meta-analysis of randomized controlled trials. A concern for standards. *JAMA : The Journal of the American Medical Association, 274*(24), 1962–1964.
Montgomery, S. A., & Åsberg, M. (1979). A new depression scale designed to be sensitive to change. *The British Journal of Psychiatry, 134*(4), 382–389.
Partlett, C., & Riley, R. D. (2017). Random effects meta-analysis: Coverage performance of 95% confidence and prediction intervals following REML estimation. *Statistics in Medicine, 36*(2), 301–317.
Paterson, T. A., Harms, P. D., Steel, P., & Credé, M. (2016). An assessment of the magnitude of effect sizes: Evidence from 30 years of meta-analysis in management. *Journal of Leadership and Organizational Studies, 23*(1), 66–81.
Pereira, T. V., Horwitz, R. I., & Ioannidis, J. P. A. (2012). Empirical evaluation of very large treatment effects of medical interventions. *Jama, 308*(16), 1676–1684.
Raya-Gonzalez, J., Castillo, D., & Clemente, F. M. (2021). Injury prevention of hamstring injuries through exercise interventions: an umbrella review. *The Journal of Sports Medicine and Physical Fitness, 61*(9), 1242–1251.
Riley, R. D., Higgins, J. P. T., & Deeks, J. J. (2011). Interpretation of random effects meta-analyses. *Bmj, 342*, Article d549.
Rodriguez-Martin, J. L., Barbanoj, J. M., Schlaepfer, T. E., Clos, S. S., Perez, V., Kulisevsky, J., & Gironell, A. (2002). Transcranial magnetic stimulation for treating depression. *Cochrane Database of Systematic Reviews, 2*.
Röver, C., Knapp, G., & Friede, T. (2015). Hartung-Knapp-Sidik-Jonkman approach and its modification for random-effects meta-analysis with few studies. *BMC Medical Research Methodology, 15*(1), 1–7.
Rücker, G., Carpenter, J. R., & Schwarzer, G. (2011). Detecting and adjusting for small-study effects in meta-analysis. *Biometrical Journal, 53*(2), 351–368.
Sánchez-Meca, J., & Marín-Martínez, F. (2008). Confidence intervals for the overall effect size in random-effects meta-analysis. *Psychological Methods, 13*(1), 31.
Santschi, V., Chiolero, A., Burnand, B., Colosimo, A. L., & Paradis, G. (2011). Impact of pharmacist care in the management of cardiovascular disease risk factors: A systematic review and meta-analysis of randomized trials. *Archives of Internal Medicine, 171*(16), 1441–1453.
Schutter, D. (2009). Antidepressant efficacy of high-frequency transcranial magnetic stimulation over the left dorsolateral prefrontal cortex in double-blind sham-controlled designs: A meta-analysis. *Psychological Medicine, 39*(1), 65–75.
Schutter, D. (2010). Quantitative review of the efficacy of slow-frequency magnetic brain stimulation in major depressive disorder. *Psychological Medicine, 40*(11), 1789–1795.
Schwarzer, G., Carpenter, J. R., & Rücker, G. (2015). Small-study effects in meta-analysis. In *Meta-analysis with R* (pp. 107–141). Springer.
Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., … Kristjansson, E. (2017). AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *Bmj, 358*.
Slotema, C. W., Blom, J. D., Hoek, H. W., & Sommer, I. E. (2010). Should we expand the toolbox of psychiatric treatment methods to include repetitive transcranial magnetic stimulation (rTMS)? A meta-analysis of the efficacy of rTMS in psychiatric disorders. *The Journal of Clinical Psychiatry, 71*(7), 873–884.
Sonmez, A. I., Camsari, D. D., Nandakumar, A. L., Voort, J. L. V., Kung, S., Lewis, C. P., & Croarkin, P. E. (2019). Accelerated TMS for depression: A systematic review and meta-analysis. *Psychiatry Research, 273*, 770–781.
Sterne, J. A., & Harbord, R. M. (2004). Funnel plots in meta-analysis. *The Stata Journal, 4*(2), 127–141.
Sterne, J. A., Sutton, A. J., Ioannidis, J. P., Terrin, N., Jones, D. R., Lau, J., … Schmid, C. H. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *Bmj, 343*.
Sterne, J. A. C., Gavaghan, D., & Egger, M. (2000). Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology, 53*(11), 1119–1129.
Suchting, R., Tirumalaraju, V., Gareeb, R., Bockmann, T., de Dios, C., Aickareth, J., … Selvaraj, S. (2021). Revisiting monoamine oxidase inhibitors for the treatment of depressive disorders: A systematic review and network meta-analysis. *Journal of Affective Disorders, 282*, 1153–1160.
Szucs, D., & Ioannidis, J. P. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLoS Biology, 15*(3), Article e2000797.
Szucs, D., & Ioannidis, J. P. (2021). Correction: Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLoS Biology, 19*(3), Article e3001151.
Thorlund, K., Wetterslev, J.ø., Awad, T., Thabane, L., & Gluud, C. (2011). Comparison of statistical inferences from the DerSimonian–Laird and alternative random-effects model meta-analyses – An empirical assessment of 920 Cochrane primary outcome meta-analyses. *Research Synthesis Methods, 2*(4), 238–253.
Turkoz, I., Alphs, L., Singh, J., Jamieson, C., Daly, E., Shawi, M., … Rush, A. J. (2021). Clinically meaningful changes on depressive symptom measures and patient-reported outcomes in patients with treatment-resistant depression. *Acta Psychiatrica Scandinavica, 143*(3), 253–263.
Umberham, B., Hedin, R., Detweiler, B., Kollmorgen, L., Hicks, C., & Vassar, M. (2017). Heterogeneity of studies in anesthesiology systematic reviews: A meta-epidemiological review and proposal for evidence mapping. *BJA: British Journal of Anaesthesia, 119*(5), 874–884.
Valiengo, L., Maia, A., Cotovio, G., Gordon, P. C., Brunoni, A. R., Forlenza, O. V., & Oliveira-Maia, A. J. (2022). Repetitive transcranial magnetic stimulation for major depressive disorder in older adults: Systematic review and meta-analysis. *The Journals of Gerontology: Series A, 77*(4), 851–860.
Van Dyk, N., Behan, F. P., & Whiteley, R. (2019). Including the Nordic hamstring exercise in injury prevention programmes halves the rate of hamstring injuries: A systematic review and meta-analysis of 8459 athletes. *British Journal of Sports Medicine, 53*(21), 1362–1370.
Voigt, J. D., Leuchter, A. F., & Carpenter, L. L. (2021). Theta burst stimulation for the acute treatment of major depressive disorder: A systematic review and meta-analysis. *Translational Psychiatry, 11*(1), 1–12.
Wang, W. L., Wang, S. Y., Hung, H. Y., Chen, M. H., Juan, C. H., & Li, C. T. (2022). Safety of transcranial magnetic stimulation in unipolar depression: A systematic review and meta-analysis of randomized-controlled trials. *Journal of Affective Disorders, 301*, 400–425.
Wei, Y., Zhu, J., Pan, S., Su, H., Li, H., & Wang, J. (2017). Meta-analysis of the efficacy and safety of repetitive transcranial magnetic stimulation (rTMS) in the treatment of depression. *Shanghai Archives of Psychiatry, 29*(6), 328.
World Health Organization. (1992). *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*. World Health Organization.
Xie, J., Chen, J., & Wei, Q. (2013). Repetitive transcranial magnetic stimulation versus electroconvulsive therapy for major depression: A meta-analysis of stimulus parameter effects. *Neurological Research, 35*(10), 1084–1091.
Zhang, Y. Q., Zhu, D., Zhou, X. Y., Liu, Y. Y., Qin, B., Ren, G. P., & Xie, P. (2015). Bilateral repetitive transcranial magnetic stimulation for treatment-resistant depression: A systematic review and meta-analysis of randomized controlled trials. *Brazilian Journal of Medical and Biological Research, 48*(3), 198–206.