Assessment of blinding in randomized controlled trials of antidepressants for depressive disorders 2000−2020: A systematic review and meta-analysis

Yi-Hsuan Lin,[a] Ethan Sahker,[a,b] Kiyomi Shinohara,[a] Noboru Horinouchi,[a] Masami Ito,[a] Madoka Lelliott,[a] Andrea Cipriani,[c,d] Anneka Tomlinson,[e] Christopher Baethge,[f] and Toshi A. Furukawa[a]*

[a]Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine and School of Public Health, Kyoto, Japan
[b]Population Health and Policy Research Unit, Graduate School of Medicine, Kyoto University, Kyoto, Japan
[c]Department of Psychiatry, University of Oxford, Oxford, UK
[d]Oxford Health NHS Foundation Trust, Warneford Hospital, Oxford, UK
[e]Manchester Pharmacy School, University of Manchester, Oxford Road, Manchester M13 9PT, UK
[f]Department of Psychiatry and Psychotherapy, University of Cologne Medical School, Cologne, Germany

eClinicalMedicine
2022;50: 101505
Published online xxx
https://doi.org/10.1016/j.eclinm.2022.101505

Summary

Background In double-blind randomized controlled trials (RCTs) of antidepressants, blinding can be broken due to the apparent side effects, and unsuccessful blinding can lead to overestimation of effect sizes. New generation antidepressants with less severe side effects may be less susceptible to broken blinding. However, successfulness of blinding in new generation antidepressant trials and its influence on trial effect size estimates remain unclear.

Methods Extending a previous systematic review assessing blinding successfulness in psychiatric trials (2000-2010), we searched PubMed/Medline for double-blinded antidepressant RCTs (2010-2020) for trials assessing blinding success. Our primary outcome was the degree of blinding successfulness, measured as kappa statistics between guesses and true allocations. We used random-effects meta-analysis to synthesize studies. We used meta-regression and Pearson's r to examine the relationship between blinding success and effect sizes. This study is registered with PROSPERO (CRD42021249973).

Findings Among 154 eligible studies, 11 (7·1%) contained information on blinding assessment between 2010 and 2020. Five studies were added from the previous review, and altogether nine of the 16 studies provided usable data. Agreement in individual studies ranged from κ=-0·14 to 0·38. The summary agreement between guesses and the truth was 0·21 (95% CI: 0·14 to 0·28) among patients and 0·17 (95% CI: 0·05 to 0·30) among assessors. Blinding success was not associated with effect size (patients: $r = 0·37$, $p = 0·32$; assessors: $r = 0·28$; $p = 0·72$). Meta-regression also failed to find a significant relationship between blinding success and depression effect sizes (β=0·06, $p = 0·09$).

Interpretation Less than 10% of the antidepressant RCTs reported blinding assessment. The results in new generation antidepressant trials indicated that patients and assessors were unlikely to be able to judge treatment allocation. There was little evidence that the extent of unblinding biased the effect size estimates of new generation antidepressants.

Funding None.

Keywords: Antidepressant; Trial; Blinding; Assessment; Successfulness

Introduction

Randomized controlled trials (RCTs) are widely acknowledged as the strongest study design to establish evidence of healthcare interventions. Blinding is often required to prevent bias during intervention administration and participant assessment.[1] Unblinded healthcare

*Corresponding author at: Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine/School of Public Health, Yoshida Konoe-cho, Sakyo-ku, Kyoto 606-8501, Japan.
E-mail address: furukawa@kuhp.kyoto-u.ac.jp (T.A. Furukawa).

Research in context

Evidence before this study

Considerable debate continues about the overstatement of antidepressant effects, partly due to unsuccessful blinding in so-called double-blind trials. Existing reviews have found that only 2-8% of psychiatric treatment trials assessed blinding successfulness, but these reviews combined varied treatments and populations. The most recent review focused specifically on antidepressants for depression and suggested that blinding has failed in antidepressant trials, but missed several relevant trials, and excluded trials with a double-dummy blinding strategy, which is an ideal design.

Added value of this study

The current systematic review provided an update of blinding assessment including newly published antidepressant studies from 2010 to 2020. To our knowledge, this is the first comprehensive meta-analysis of blinding successfulness in antidepressant trials. We identified 16 double-blind antidepressant trials examining successfulness of their blinding, nine of which provided usable data for the meta-analysis of successful blinding in the comparison between new generation antidepressants and placebo. The present findings showed that blinding was well maintained in these trials, and that there was no association between effect sizes and blinding successfulness.

Implications of all the available evidence

These findings allay some of the doubts behind antidepressant efficacy by demonstrating that broken blinding did not lead to risks of performance and assessment bias. We recommend increased assessment and reporting of blinding in RCTs to further improve the strength of the evidence. Future research should seek to extend the analysis to other psychotropic drugs.

workers providing additional treatments for a certain group of participants would confound the results (performance bias). Outcome assessments may be distorted if assessors know which intervention participants received, and unblinded participants may report differently than blinded participants (assessment bias).[2] Moreover, knowing treatment allocations may lead to unbalanced loss to follow-up between groups (attrition bias). Empirical data have repeatedly shown that lack of blinding is likely to exaggerate effect sizes of active treatment in clinical trials.[3,4]

Blinding is especially important in psychiatric studies since subjective clinical rating scales are usually used to assess outcomes.[5] Among the mental disorders, depression is one of the leading causes of global burden of disease.[6] Antidepressants have indeed been one of the most commonly prescribed drugs since the introduction of serotonin selective reuptake inhibitors (SSRI) in the 1990s, and the consumption of antidepressants has more than doubled in countries belonging to the Organization for Economic Co-operation and Development (OECD) between 2000 and 2017.[7] Though researchers have conducted many trials supporting the evidence of antidepressant efficacy, there is still argument about antidepressant effectiveness. Major detractors of antidepressants usually highlight the relatively small effect size of these drugs against placebo, claiming blinding is untrustworthy, and real-world effect sizes of antidepressants are expected to be even smaller.[8] In double-blind trials of antidepressants, blinding may still be broken due to apparent side effects of antidepressants.[9] Detailed descriptions of specific side effects in the informed consent can further facilitate the discrimination between arms.[10] Participants or evaluators identifying the group medication status can lead to overstating response on a new drug's or understating response on placebo.

To allay suspicion and criticism against unblinding in clinical trials, some studies assess whether healthcare providers, participants, and outcome assessors remained unaware of allocated treatments throughout the trial by asking them to guess the group allocation. Successful blinding can not only enhance the validity and credibility of RCTs, but also help readers appraise the quality of results.[11] Across medicine, less than 10% of studies report blindness assessment.[12] In psychiatry, the proportion appears much smaller at 2·5%.[13]

Since 1990, there has been an upsurge of so-called new generation antidepressants, which appear to have less severe side effects overall.[14,15] Therefore, these new generation antidepressants may be less susceptible to breaking of the blinding when being examined in RCTs. A previous systematic review has assessed blinding successfulness on studies published prior to March 2010 for schizophrenia and affective disorders.[13] However, existing summary evaluations of antidepressant trials do not account for chance in their analyses.

Recently, Scott et al.[16] published a systematic review of blinding successfulness in antidepressant trials. Their search excluded RCTs with a double-dummy blinding strategy, which is an ideal design used to compare drugs with very different appearances. For example, two drugs with two different forms are administered to one group, one being active and one being placebo. Further, their index of blinding successfulness may be clinically uninterpretable because they did not distinguish trials with two or more types of antidepressants, and they calculated Bang's BI,[44,45] a new blinding index that could take "don't know" response into account, for each of a study and then combined them using Hedges' g for each study; however, they did not pool the results across all the included studies. Thus, high-quality evidence for blinding successfulness in new generation antidepressants remains lacking.

Given the clinical importance of blinding effectiveness considering side effects of antidepressants, we updated the previous systematic review of Baethge[13] by incorporating newly published, pertinent evidence of blinding successfulness in antidepressant trials after 2010. Therefore, the present study investigated (i) the proportion of antidepressant RCTs with blinding assessment 2010-2020, (ii) degree of blinding successfulness in trials of new generation antidepressants, and (iii) relationship between blinding successfulness and trial effect sizes.

Methods

Search strategy and selection criteria

Our systematic review and meta-analysis followed PRISMA guidelines, and the study protocol is registered with PROSPERO (CRD42021249973). We included double-blinded RCTs comparing any antidepressants with placebo or active antidepressants for the treatment of depressive disorders (major depression, unipolar depression, dysthymia, minor depression, postpartum depression, or bereavement), as diagnosed by standard international criteria (Feighner criteria, Research Diagnostic Criteria, DSM-III, DSM-III-R, DSM-IV, DSM-IV-TR, DSM-5, ICD-9, and ICD-10). Antidepressant classification is shown in Appendix 1 p.2. We excluded studies recruiting participants with comorbid psychotic disorders, substance use disorders, or cognitive impairments, because patients with above comorbidities may be a different population from depressive patients or might not be able to make judgments on the treatment allocations due to recognition difficulties. To ensure the comparability of blinding effects, only studies with a placebo comparison were analyzed for blinding successfulness.

We retained antidepressant articles on depressive disorders in the previous systematic review,[37-41] and searched PubMed for newly published studies from March 01, 2010 to December 31, 2020 in English. Search terms included synonyms of double-blinded method, randomization, depressive disorders, and antidepressants (full search terms provided in Appendix 1 p.3). Y.H.L. conducted the literature search and three pairs from among six investigators (Y.H.L., E.S., K.S., M.I., N.H., and M.L.) independently screened studies and reviewed the full-text manuscripts, including supplementary materials, to determine inclusion. Reference lists of relevant antidepressant trials were also searched. Excluded duplicate reports were identified throughout the review process by study name, trial number, methodology, and specific patient characteristics.

Publications fulfilling the eligibility criteria were searched for information regarding blinding assessment. Blinding assessment was defined as any statement or data on guesses as to which treatment the participants received (experimental treatment vs. control) by patients, raters, or clinicians. Five pairs from among six investigators (Y.H.L. paired with E.S., K.S., M.I., N.H., and M.L.) independently extracted data of study characteristics, methods of blinding assessment, and summary results from studies with blinding assessments (details of data extracted were shown in Appendix 1 p.4). After extraction, researchers worked in pairs to find consensus on the extracted data. Authors were contacted for unreported information as needed.

Data analysis

Our primary outcome was the degree of blinding successfulness measured with Cohen's kappa (κ). We chose kappa statistics because it is commonly employed by trialists when assessing blinding success. Furthermore, kappa accounts for correct guesses by chance, and the statistical method of meta-analysis for kappa is well developed.[17] A κ value of 0 indicated successful blinding, and a κ value of 1 reveals that all the patients/assessors could correctly identify patients' treatment so that the blinding was totally broken. A positive value implied failure of blinding, whereby the majority of personnel correctly guessed the treatment allocation above random guessing.[18] A negative value from 0 to -0·20 indicated patients were unable to tell the treatment allocations, while a more extreme negative one implied blinding failure in the other direction. We adapted the established guidelines and defined a κ value of -0·20 to 0·20 as successful blinding, 0·21 to 0·40 as slightly broken, 0·41 to 0·60 as moderately broken, 0·61 to 1 as severely broken.[19] The cited reference was used to explain inter-rater reliability, where an inconsistent result was not desirable. However, in blinding assessment, a certain degree of incorrect guesses was acceptable. The negative limit at -0.20 was following the positive window of successful blinding at 0.20, and we assumed that a more extreme negative value indicates an opposite direction of broken blinding. When multiple arms were reported in a single study, we included only relevant antidepressant and placebo arms. If a study reported the blinding assessment results at multiple timepoints, we chose the assessment at end of treatment.

Because the study characteristics and participant inclusion criteria varied in antidepressant research, we applied a random-effects model to pool kappas from the included studies. Kappas were calculated using a prespecified formula (Appendix 1 p.5).[17] Only studies with a placebo comparison were analyzed to ensure the comparability of blinding effects. If an open-choice option design was applied to guess tests, the "don't know" response was assigned proportionally by the number of active-treatment and control responses in a given report. When a report did not provide the exact number of

patients for whom treatments were guessed, we assumed that all the patients staying in the trial at the time of blinding assessment were examined. Heterogeneity between study-specific estimates was measured with I^2 and tau.[2] The heterogeneity was interpreted as being not important when I^2 were between 0–40%; moderate when 30–60%; substantial when 50–90%; considerable when 75–100%.[20] Publication bias was evaluated through visual analysis of the funnel plot, contour-enhanced funnel plot,[21] and Egger's test. The risk of bias tool, which aims at assessing various sources of bias of an RCT with regard to a specific, chosen outcome, was not used in this meta-analysis. This is because we were assessing the blinding successfulness, one of the sources of bias, rather than the overall bias for an RCT.

To assess the relationship between blinding success and study effect size, we extracted depression scores at treatment endpoint and calculated the treatment effect using Cohen's d, a standardized mean difference (SMD), with its 95% confidence interval (CI). A negative value meant that the depression scores decreased more in the active treatment group and that the effect was superior to placebo. We prioritized endpoint depression scores. If endpoint scores were not reported, we used change scores. If a trial failed to report SDs, we imputed the SDs from other studies in our meta-analysis using the same rating scales.[22] We combined the summary depression effect sizes of two active treatment groups if the two treatment arms were both classified as active treatment in the blinding assessment. Pearson's *r* correlation coefficient was used to assess the relationship between effect sizes (SMD) and the degree of blinding success (κ). An absolute Person's r of 0·1 to 0·4, 0·4 to 0·6, and above 0·6 indicated small, medium, and strong correlation. A positive Pearson's *r* meant more effective blinding was associated with larger effect sizes in favor of antidepressants. We also applied a mixed-effects meta-regression to investigate the relationship of effect sizes and blinding success.

Post-hoc sensitivity analyses were conducted by excluding studies without clear information about assessor blinding, studies which did not report the exact number of patients tested with blinding assessment, and studies reporting change scores for depression, and summarizing kappas for blinding assessment conducted during the trial. All tests were considered statistically significant for p-values less than 0·05. All analyses and the corresponding graphical visualization of forest and funnel plots were performed in R 3.6.0 using the *metafor* Version 3.0.2 package.[23,24]

Role of the funding source

There was no funding source for this study. Y.H.L., E.S., and T.A.F had full access to the data. Y.H.L., E. S., and T.A.F critically revised and edited drafts of the paper and were responsible for the decision to submit for publication.

Results

Figure 1 shows the procedure of study selection and screening. Overall, 154 eligible antidepressant studies were identified. Inter-rater agreement of screening among investigators was 92% (κ=0·65, 95% CI 0·60 to 0·70) in title and abstract screening and 84% (κ=0·65, 95% CI 0·57 to 0·73) in full-text screening.

Proportion of trials with blinding assessment 2010-2020

Among 154 eligible studies, 11 (7·1%) contained information on blinding assessment from 2010 to 2020.[25-35] Across the decade considered, we found no temporal trend toward increasing or decreasing annual rates of blinding documentation.

Five studies were added from the previous meta-analysis[13,36-40] Characteristics of publications reporting on blinding assessment were analyzed in a total of 16 studies (Table 1). Descriptive characteristics of the studies, patients, and treatments are summarized in Appendix 1 p.6-12 (Table S1). Blinding of patients were reported in all 16 studies. However, blinding of assessors was not clearly stated in two studies (12·5%), although both studies claimed that they used double-blinded methods. The blinding assessment was all assessed by asking the personnel, patients or raters, to guess the treatment patients received. Patients were assessed in 75% and assessors in 56·3% of the included studies. Blinding assessment was conducted during the trial in six studies and at the end of treatment in nine. Two (12·5%) did not report any result regarding the test, and only six (37·5%) reported complete data relevant to assessment of blinding, including guessing options, the number of patients tested, and the number or proportion of guesses.

Degree of blinding successfulness

Since the two articles on old generation antidepressants (tricyclics and monoamine oxidase inhibitors) did not report compatible outcomes regarding blinding assessment, one of which provided only proportions of correct guesses without the number of patients assessed,[40] and the other used an active control as a comparison,[35] we could not compare the blinding successfulness between old and new generation antidepressants. Thus, we only estimated the summary blinding effects of new generation antidepressants.

After inquires with the original authors, a total of nine studies provided usable data for the meta-analysis. The nine studies for patients blinding included 1177 patients (564 were allocated to placebo and 613 to active

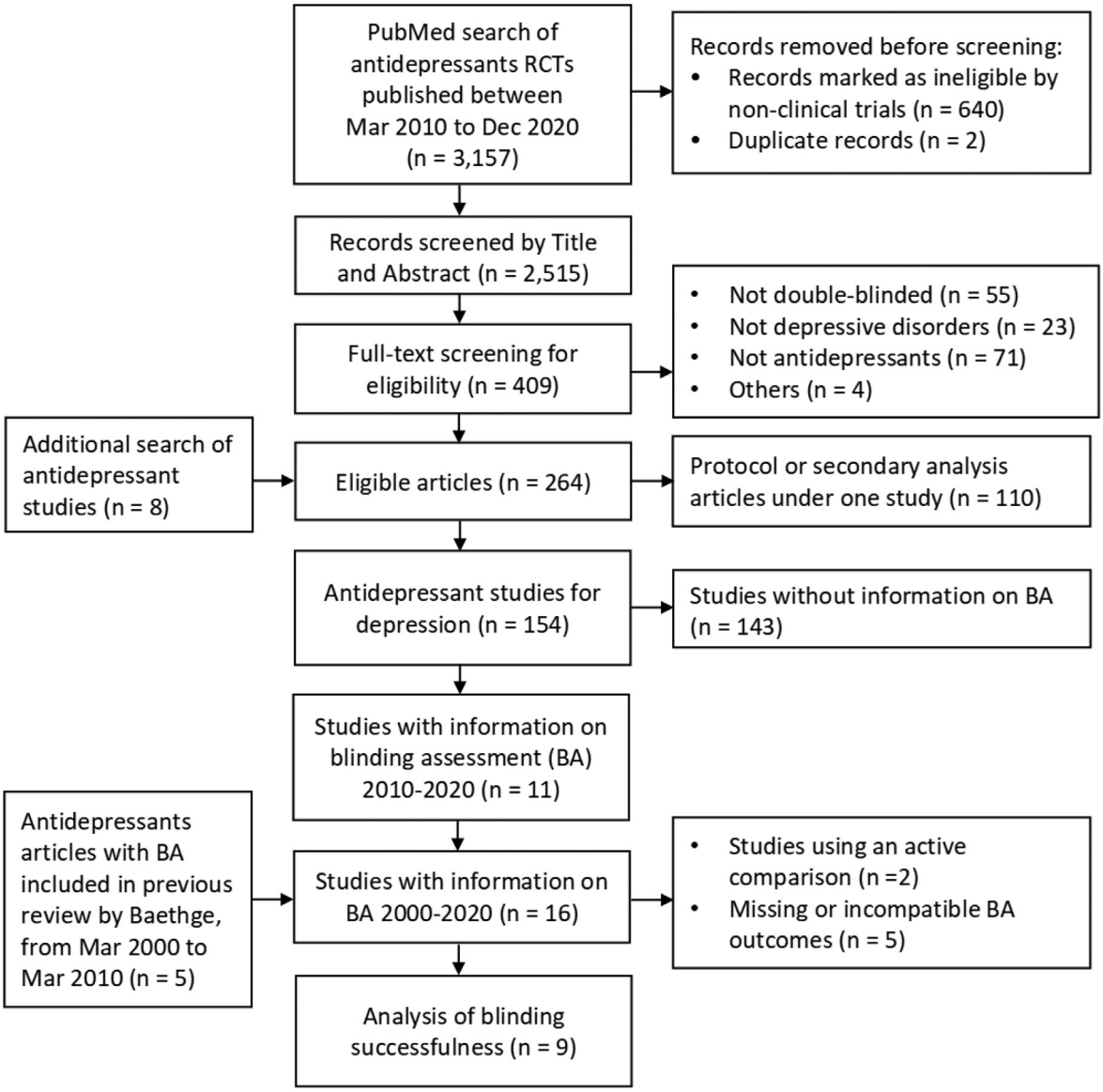

Figure 1. Study selection process. BA: blinding assessment.

drug), and the four studies for assessors blinding included 269 patients (147 were allocated to placebo and 122 to active drug). The active antidepressants were fluoxetine in three studies, sertraline in three, escitalopram in one, venlafaxine in one, and venlafaxine and paroxetine in one. The kappa of individual studies ranged from -0·14 to 0·38, and the proportion of correct guesses ranged from 45% to 71% (Appendix 1 p.13, Table S2). The summary kappa representing blinding effects among patients was 0·21 (95% CI 0·14 to 0·28, I^2=23·8%; Figure 2A). One study in 1997 was included because the secondary analysis on blinding assessment was published in 2000.[39,41] For assessors, only four studies provided enough information to calculate a kappa. The agreement between the assessors' guesses and the true allocations was a kappa of 0·17 (95% CI 0·05 to 0·30, I^2<1%; Figure 2B). We did not observe significant heterogeneity among studies. The results remained similar in all sensitivity analyses, including well-defined double-blinded studies (n=8 for patients), studies with complete blinding information (n=6 for patients and n=2 for assessors), and studies assessing blinding successfulness during trials (n=2 for patients; Appendix 1 p.14, Table S3). Among the four studies claiming to have conducted blinding assessment during the trial, only two provided complete data, and the kappas at different assessed time points are shown in Appendix 1 p.15 (Table S4). Studies were symmetrically

Characteristics	Trials with BA (*N* = 16), *n* (%)	Trials included in MA (*N* = 9)*, *n*(%)
Published year		
2000-2009	5 (31·3)	3 (33·3)
2010-2020	11 (68·8)	6 (66·6)
Sponsor		
Industry	1 (6·3)	0 (0)
Non-industry	15 (93·8)	9 (100)
Main depression type		
MDD	10 (62·5)	4 (44·4)
Dysthymic disorder	1 (6·3)	1 (11·1)
Postpartum depression	1 (6·3)	0 (0)
Mixed	4 (25·0)	4 (44·4)
Trial arms included		
Two	10 (62·5)	7 (77·8)
Three	6 (37·5)	2 (22·2)
Control type		
Placebo	14 (87·5)	9 (100)
Active treatment	2 (12·5)	0 (0)
Type of antidepressants		
Old generation	2 (12·5)	0 (0)
New generation	14 (87·5)	9 (100)
Blinding method		
Double	3 (18·8)	1 (11·1)
More	13 (81·3)	8 (88·9)
Persons blinded		
Patients	16 (100)	9 (100)
Assessors	14 (87·5)	8 (88·9)
Caregivers	14 (87·5)	8 (88·9)
Investigators	8 (50·0)	5 (55·6)
Analytical method		
Intention to treat	13 (81·3)	8 (88·9)
Depression measure		
Observer-based	15 (93·8)	7 (77·8)
Patient-reported	1 (6·3)	1 (11·1)
Blinding assessed in		
Patients	12 (75·0)	9 (100)
Assessors	9 (56·3)	4 (44·4)
Timing of BA		
During trial	6 (37·5)	4 (44·4)
End of trial	9 (56·3)	4 (44·4)
Unclear	1 (6·3)	1 (11·1)
Blinding ratings		
Forced choice (active vs. control)	11 (68·8)	8 (88·9)
Allow 'don't know' option	1 (6·25)	1 (11·1)
Unclear	4 (25·0)	0 (0)
Qualitative conclusions of BA for patients		
Reported as successful	5 (31·3)	4 (44·4)
Reported as unsuccessful	3 (18·8)	2 (22·2)
No conclusion reported	4 (25·0)	3 (33·3)
Qualitative conclusions of BA for assessors		
Reported as successful	5 (31·3)	3 (33·3)
Reported as unsuccessful	2 (12·5)	1 (11·1)
No conclusion reported	2 (12·5)	0 (0)

***Table 1*: Characteristics of 16 antidepressant RCTs with blinding assessment and of 9 studies included in the meta-analysis.**
BA: blinding assessment; MA: meta-analysis; MDD: major depressive disorders.
* Seven studies without compatible outcome data for kappa calculation were excluded from the meta-analysis.

distributed at both sides of the kappa estimates in the funnel plot (Appendix 1 p.16, Figure S1A; $p = 0{\cdot}19$ for Egger's test). Visual inspection of the contour-enhanced funnel plot was not suggestive of selective publication of non-significant kappa values since the studies lied in both significant ($p<0{\cdot}05$) and non-significant ($p>0{\cdot}05$) areas (Appendix 1 p.16, Figure S1B).

Relationship between blinding successfulness and trial effect sizes

The pooled SMD effect size of the nine included trials was -0·64 (95% CI -1·32 to 0·03; individual effect sizes were shown in Appendix 1 p.13, Table S2). There was insufficient evidence to demonstrate the association between blinding successfulness and effect sizes for both patient blinding ($r = 0{\cdot}37$, 95% CI: -0·39 to 0·83, $p = 0{\cdot}32$; Figure 3A) and assessor blinding ($r = 0{\cdot}28$, 95% CI: -0·93 to 0·97, $p = 0{\cdot}72$; Figure 3B). The sensitivity analysis excluding studies using change scores showed a similar result (Appendix 1 p.17, Figure S2). Meta-regression analysis of patient blinding success on effect sizes revealed no significant relationship (β=0·06, $p = 0{\cdot}09$).

Discussion

We identified 16 trials of antidepressants for depression that examined the success of blinding from randomized treatment assignment. In the trials examining new generation antidepressants, the pooled kappa statistics between the true allocations and guesses suggested that blinding was more or less successful, with confidence intervals straddling the border of being successful to slightly broken among patients and assessors, which were 0·21 (95% CI 0·14 to 0·28) and 0·17 (95% CI 0·05 to 0·30), respectively. We did not find a significant relationship between blinding successfulness and antidepressant effect sizes.

In our study, between 2010 and 2020, only 7·1% of antidepressant trials examining blinding success. While there was a slight increase in the proportion of studies with blinding examination compared to 1·79% from 2000 to 2010,[13] the proportion was still low. Among the 11 studies, two reported a blinding assessment in the protocol, but it was not reported in the final paper. Therefore, blinding may have been assessed more often than we identified. The low prevalence might be due to a lack of consensus as to appropriate methods of blinding assessment, such as the appropriate assessment time and measurement scales. A guideline of blinding assessment was proposed in 2009, suggesting that blinding assessment should be repeatedly done at baseline, during, and near the end of treatment.[18] Changing answers can be taken into account when assessing blinding quality. From the results of kappas at different time points for studies assessing blinding

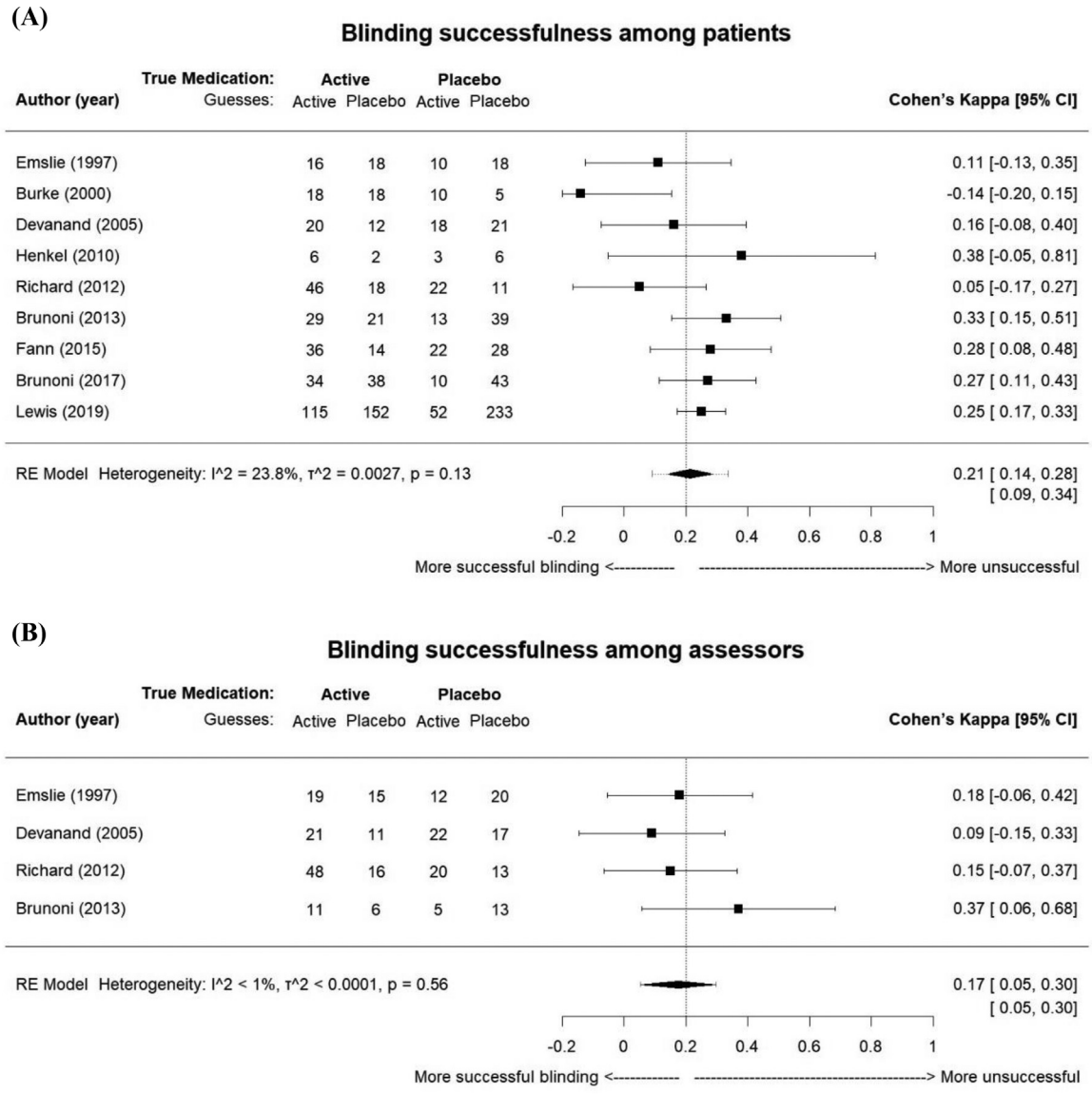

Figure 2. Forest plot of blinding successfulness among (A) patients and (B) assessors in antidepressants trials. κ value: -0.20 to 0.20 as successful blinding, 0.21 to 0.40 as slightly broken, 0.41 to 0.60 as moderately broken, 0.61 to 1 as severely broken. The width of diamond is the 95% confidence interval of the summary kappa, and the dashed line shows the prediction interval. τ^2: tau^2 measure of between-study variance. One study in 1997 was included because the secondary analysis on blinding assessment was published in 2000.

successfulness during the trial, kappas may increase as the trials proceeded, which indicated that patients and assessors may be more likely to identify the treatment when the side effects became more apparent. Moreover, a "don't know" option should be included,[42,43] and new indexes, James's BI and Bang's BI,[18,44,45] were developed to handle the "don't know" response. However, some have criticized that when there was a choice with uncertainty, people tend to choose that answer. Therefore, a new five-point scale has been created using response categories of strongly believe treatment, somewhat believe treatment, don't know, somewhat believe control, or strongly believe control.[18,46] This scale accounts for the degree to which respondents believe the response. However, the commonly used blinding index (BI), such as kappa, James's BI and Bang's BI, cannot be measured with three-point and five-point scales. In cases allowing the "don't know" option, studies could simply report a contingency table with descriptive results. Because there are many options for

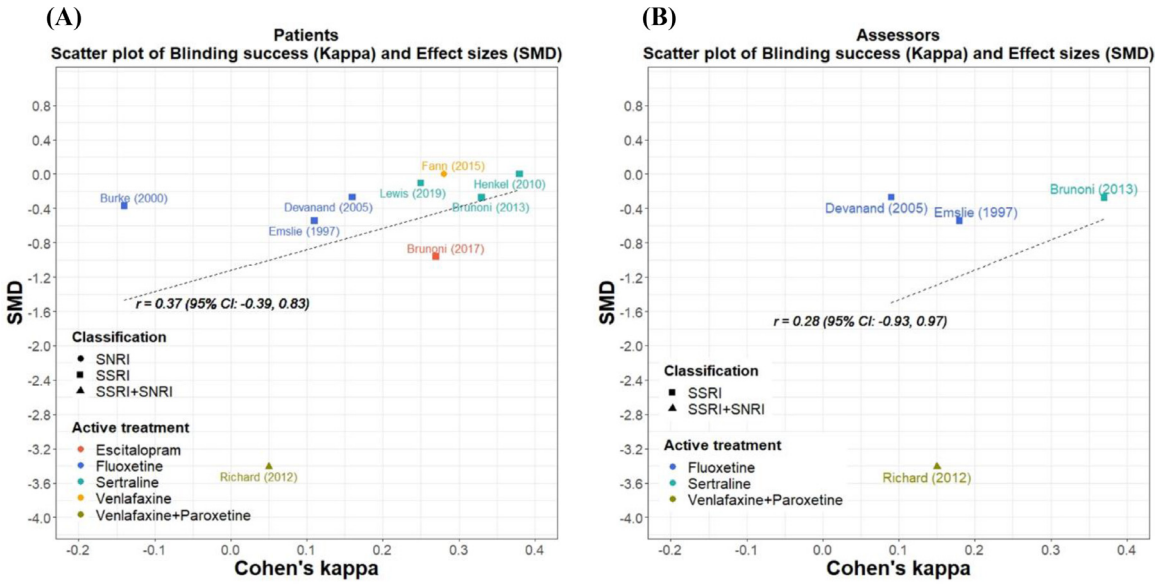

Figure 3. Relationship (Pearson) between effect sizes (SMD) and the degree of blinding successfulness (kappa) of (A) patients in 9 trials (r = 0.37, 95% CI: -0.39 to 0.83, p = 0.32) and (B) assessors in 4 trials (r = 0.28, 95% CI: -0.93 to 0.97, p = 0.72). SMD was measured by Cohen's d-statistic. A negative value indicated that the effect of active antidepressants was superior to placebo.

assessment, and the trustworthiness of blinding successfulness measurement has been called into question, a formal recommendation guideline is needed.

Reporting standards for blinding assessment was once included in the Consolidated Standards of Reporting Trial (CONSORT) but was eliminated from the 2010 guidelines.[47] The CONSORT authors cited interpretation and measurement difficulties as the reason for its removal. However, the internal validity of RCTs with placebo controls is based on the assumption of appropriate blinding.[1] Based on the present findings, we suggest that blinding assessment should be a key feature of RCT reports. When assessing the quality of a study, in addition to checking whether patients or assessors are stated blinded, we should also consider the existence of blinding assessment.

Naudet et al.[48] suggest that blinding quality could be improved by using a placebo mimicking the appearance and taste of active treatments, by keeping raters blinded from the treatment allocation, the study design, and the hypothesis, and by adopting an active control. Further, a four-arm balanced placebo trial design with one factor of intervention (antidepressants or placebo) and one factor of instructions about what they will receive (antidepressants or placebo), was developed to reduce broken blinding; however, due to the ethical concerns, deception has not been applied in antidepressant trials.[49]

The summary agreement between true allocations and guesses was within the range of successful blinding and slightly broken for both patients and assessors in new generation antidepressant trials. Although the summary kappa was on the border of success and slightly broken, we concluded that blinding was maintained because the results of the correlation analysis (r = 0.37 for patients; r = 0.28 for assessors) and meta-regression (β=0·06, p = 0·09) demonstrated that unblinding was not associated with effect size. A recent systematic review came to a conflicting conclusion, stating that blinding failed in antidepressant trials.[16] However, their search may be deficient since they excluded trials applying a double-dummy strategy, which lowers the representativeness of antidepressant trials, with only two placebo-controlled studies identified after 2000. Further, they did not account for loss to follow-up in their blinding assessment. They calculated the number of patients randomized in the final blinding assessment, rather than the number completed. This is misleading, as blinding assessment is done at a midpoint or at the end of the trial.

Our summary treatment effect sizes between antidepressants and placebo (-0·64, 95% CI -1·32 to 0·03) was larger than an overall effect estimate with a SMD of 0·3 reported by Cipriani et al.[15] The effect size analysis in this study was skewed by one study,[33] which used a change score to measure the effect sizes and had a much smaller standard deviation. Excluding this study, the effect sizes were similar to previous evidence (-0·32, 95% CI -0·56, -0·07).

There was insufficient evidence to conclude an association between blinding success and antidepressant effect size. We found a positive correlation meaning that the lower the blinding success, the smaller the treatment effect size. The direction of association in the present study was in opposition to those from previous

findings, which show that greater blinding success would lead to smaller effect sizes.[3,4,13,50] Yet, the correlation coefficient was small and not statistically significant. This finding may be because there were no moderately or severely unblinded studies found in our review. Thus, slightly broken blinding did not have an appreciable effect on depression effect sizes. Since there were a small number of studies assessing the blinding assessment, the power in this study was low. Thus, although the present point estimate is suggestive of a limited effect, the current confidence interval allowed for substantial uncertainty. Taken together, the counter-intuitive directionality of the relationship, small coefficient, large confidence intervals, and low power make interpretation difficult, and more work with more blinding data from studies is necessary.

The present findings of no association between blinding effects and antidepressant effect sizes support the notion that new generation antidepressant side effects may pose low risks of assessment, performance, and attrition bias due to broken blinding. Past research has cast doubt on the efficacy of antidepressant benefits because of the poor quality of evidence.[8,51,52] However, our results supports an alternative view that the bias related to blinding does not pose a serious risk to effect estimation and reinforces the idea that performance and assessment bias would not be a major threat when considering the overall quality of trials. Further, our results extend the finding from a comprehensive review that unblinding was less likely to occur in SSRI trials compared with tricyclic antidepressant (TCA) trials since the expected side-effect symptom rates in the drug groups are lower in SSRIs.[9] Newer antidepressants have less side effects[15] that could ease the identifiability of treatments. Overall, the successful blinding of new antidepressants trials can enhance the validity of the effect size estimates of new antidepressant, e.g., SSRIs, SNRIs, etc., reported in the trials.[14,15]

This study has limitations. First, this study did not find enough studies to compare the blinding successfulness between older and newer antidepressants. However, we were able to estimate the summary kappa of new generation antidepressant trials, and it showed that blinding was well maintained for new generation antidepressants. Second, the summary statistic of blinding successfulness was based on only nine studies providing sufficient information. The scarce number of studies limits the generalizability. Importantly, this is also a serious limitation of the evidence base, bringing uncertainty to the efficacy of antidepressants. Thus, we suggest that blinding assessment should be conducted and reported regularly to enhance validity and the future evidence base. Third, the studies collected were somewhat similar, most being SSRIs, and subgroup analysis considering mechanism of action was not possible. For newer antidepressants included in this study, i.e., SSRIs and SNRIs, the side effect profiles do not substantially differ. Thus, the blinding successfulness was expected to be similar.[17] Fourth, we assumed that studies not reporting blinding assessment did not assess the blinding, which may underestimate the number of studies examining the blinding. However, we included protocols during the study selection. If results of blinding assessment were not reported in the final publications but the assessment was mentioned in the protocol, we contacted authors for relevant information, which minimized the missing of studies with blinding assessment. Fifth, the present study did not search for unpublished papers because we were updating the previous systematic review by Baethge,[13] following their search strategy. Further, we actually found more studies than Scott et al.,[16] which did include a search for unpublished studies.

In conclusion, blinding assessment was reported in less than 10% of recent antidepressant trials, and the results of blinding assessments in new generation antidepressant trials indicated that patients or assessors were unlikely to judge which treatment the patients were on. Currently, the available evidence suggests that efficacy of new antidepressants is probably not overestimated due to broken blinding. However, there are only a very limited number of studies that report blinding success and so more accountability and transparency is needed among clinical trials. Blinding and reporting of blinding assessment can lead to more accurate effect estimates and greater confidence in the drugs prescribed to depressed patients. More rigorous guidance on how to assess and report blinding success in psychiatric trials is warranted.

Contributors

TAF, ES and YHL conceived and designed the study. YHL, ES, KS, MI, NH and ML reviwed the articles and extracted the data. YHL performed the analysis and the visualization. TAF, ES, YHL, CB, AC and AT contrubuted to the intepretation of the findings. TAF, ES and YHL wrote the first draft of the manuscript. All authors contributed to the critical revision of the manuscripts and approved the final version for submission.

Data sharing statement

Data collected for this study and the code used to analyze these data is available upon request from the corresponding author, for purposes of reproducing the results.

Declaration of interests

TAF reports grants and personal fees from Mitsubishi-Tanabe, personal fees from SONY, grants and personal fees from Shionogi, outside the submitted work; In addition, TAF has a patent 2020-548587 concerning

smartphone CBT apps pending, and intellectual properties for Kokoro-app licensed to Mitsubishi-Tanabe. AC declares funds from Angelini Pharma for a research project on telepsychiatry during COVID-19; AC also declares payment and travel support form Angelini Pharma for a workshop in Rome about digital psychiatry. AT has received funds from Angelini Pharma. All the other authors declared no conflict of interest.

Funding

There was no funding for this study.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.eclinm.2022.101505.

References

1 Higgins JPT, Savović J, Page MJ, Elbers RG, Sterne JAC. Chapter 8: Assessing risk of bias in a randomized trial. In: Higgins JPT, Thomas J, Chandler J, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 62 (updated February 2021). Cochrane. 2021. Available from: www.training.cochrane.org/handbook.
2 Karanicolas PJ, Farrokhyar F, Bhandari M. Practical tips for surgical research: blinding: who, what, when, why, how? *Can J Surg*. 2010;53(5):345–348.
3 Savovic J, Turner RM, Mawdsley D, et al. Association between risk-of-bias assessments and results of randomized trials in cochrane reviews: the ROBES meta-epidemiologic study. *Am J Epidemiol*. 2018;187(5):1113–1122.
4 Leucht S, Corves C, Arbter D, Engel RR, Li C, Davis JM. Second-generation versus first-generation antipsychotic drugs for schizophrenia: a meta-analysis. *Lancet*. 2009;373(9657):31–41.
5 Wood L, Egger M, Gluud LL, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. *BMJ*. 2008;336(7644):601.
6 GBD Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet*. 2017;390(10100):1211–1259.
7 OECD. *Health at a Glance 2019: OECD Indicators*. Paris: OECD Publishing; 2019.
8 Gaudiano BA, Herbert JD. Methodological issues in clinical trials of antidepressant medications: perspectives from psychotherapy outcome research. *Psychother Psychosom*. 2005;74(1):17–25.
9 Mora MS, Nestoriuc Y, Rief W. Lessons learned from placebo groups in antidepressant trials. *Philos Trans R*. 2011;366(1572):1879–1888.
10 Moncrieff J. Are antidepressants overrated? A review of methodological problems in antidepressant trials. *J Nervous Ment Dis*. 2001;189(5):288–295.
11 Boutron I, Estellat C, Guittet L, et al. Methods of blinding in reports of randomized controlled trials assessing pharmacologic treatments: a systematic review. *PLoS Med*. 2006;3(10):e425-e.
12 Hróbjartsson A, Forfang E, Haahr MT, Als-Nielsen B, Brorson S. Blinded trials taken to the test: an analysis of randomized clinical trials that report tests for the success of blinding. *Int J Epidemiol*. 2007;36(3):654–663.
13 Baethge C, Assall OP, Baldessarini RJ. Systematic review of blinding assessment in randomized controlled trials in schizophrenia and affective disorders 2000-2010. *Psychother Psychosom*. 2013;82(3):152–160.
14 Santarsieri D, Schwartz TL. Antidepressant efficacy and side-effect burden: a quick guide for clinicians. *Drugs Context*. 2015;4:212290.
15 Cipriani A, Furukawa TA, Salanti G, 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 North Am Ed*. 2018;391(10128):1357–1366.
16 Scott AJ, Sharpe L, Colagiuri B. A systematic review and meta-analysis of the success of blinding in antidepressant RCTs. *Psychiatry Res*. 2022;307: 114297.
17 Sun S. Meta-analysis of Cohen's kappa. Health Serv Outcomes Res Methodol. 2011;11(3):145–163.
18 Kolahi J, Heejung B, Park J. Towards a proposal for assessment of blinding success in clinical trials: up-to-date review. Commun Dent Oral Epidemiol. 2009;37(6):477–484.
19 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174.
20 Deeks JJ, Higgins JPT, Altman DG. Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins JPTTJ, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, eds. Cochrane Handbook for Systematic Reviews of Interventions version 62 (updated February 2021). Cochrane. 2021. Available from: www.training.cochrane.org/handbook.
21 Peters JL, Sutton AJ, Jones DR, Abrams KR, Rushton L. Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *J Clin Epidemiol*. 2008;61(10):991–996.
22 Furukawa TA, Barbui C, Cipriani A, Brambilla P, Watanabe N. Imputing missing standard deviations in meta-analyses can provide accurate results. *J Clin Epidemiol*. 2006;59(1):7–10.
23 Viechtbauer W. Conducting meta-analyses in R with the metafor package. *J Stat Softw*. 2010;1(3):2010.
24 Core Team R. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2020.
25 Bloch M, Meiboom H, Lorberblatt M, Bluvstein I, Aharonov I, Schreiber S. The effect of sertraline add-on to brief dynamic psychotherapy for the treatment of postpartum depression: a randomized, double-blind, Placebo-controlled study. *J Clin Psychiatry*. 2012;73(2):235–241.
26 Brunoni AR, Moffa AH, Sampaio-Junior B, et al. Trial of electrical direct-current therapy versus escitalopram for depression. *N Engl J Med*. 2017;376(26):2523–2533.
27 Brunoni AR, Valiengo L, Baccaro A, et al. The sertraline vs. electrical current therapy for treating depression clinical study: results from a factorial, randomized, controlled trial. *JAMA Psychiatry*. 2013;70(4):383–391.
28 Fann JR, Bombardier CH, Richards JS, et al. Venlafaxine extended-release for depression following spinal cord injury: a randomized clinical trial. *JAMA Psychiatry*. 2015;72(3):247–258.
29 Henkel V, Mergl R, Allgaier AK, et al. Treatment of atypical depression: post-hoc analysis of a randomized controlled study testing the efficacy of sertraline and cognitive behavioural therapy in mildly depressed outpatients. *Eur Psychiatry*. 2010;25(8):491–498.
30 Komulainen E, Heikkilä R, Nummenmaa L, et al. Short-term escitalopram treatment normalizes aberrant self-referential processing in major depressive disorder. *J Affect Disord*. 2018;236:222–229.
31 Lavretsky H, Reinlieb M, St Cyr N, Siddarth P, Ercoli LM, Senturk D. Citalopram, methylphenidate, or their combination in geriatric depression: a randomized, double-blind, placebo-controlled trial. *Am J Psychiatry*. 2015;172(6):561–569.
32 Lewis G, Duffy L, Ades A, et al. The clinical effectiveness of sertraline in primary care and the role of depression severity and duration (PANDA): a pragmatic, double-blind, placebo-controlled randomised trial. *Lancet Psychiatry*. 2019;6(11):903–914.
33 Richard IH, McDermott MP, Kurlan R, et al. A randomized, double-blind, placebo-controlled trial of antidepressants in Parkinson disease. *Neurology*. 2012;78(16):1229–1236.
34 Vermeiden M, Mulder PG, van den Broek WW, Bruijn JA, Birkenhäger TK. A double-blind randomized study comparing plasma level-targeted dose imipramine and high-dose venlafaxine in depressed inpatients. *J Psychiatr Res*. 2013;47(10):1337–1342.
35 Wijkstra J, Burger H, van den Broek WW, et al. Treatment of unipolar psychotic depression: a randomized, double-blind study comparing imipramine, venlafaxine, and venlafaxine plus quetiapine. *Acta Psychiatr Scand*. 2010;121(3):190–200.
36 Burke WJ, Hendricks SE, McArthur-Miller D, et al. Weekly dosing of fluoxetine for the continuation phase of treatment of major depression: results of a placebo-controlled, randomized clinical trial. *J Clin Psychopharmacol*. 2000;20(4):423–427.

37 Davidson JR, Gadde KM, Fairbank JA, et al. Effect of Hypericum perforatum (St John's wort) in major depressive disorder: a randomized controlled trial. *JAMA*. 2002;287(14):1807–1814.
38 Devanand DP, Nobler MS, Cheng J, et al. Randomized, double-blind, placebo-controlled trial of fluoxetine treatment for elderly patients with dysthymic disorder. *Am J Geriatr Psychiatry*. 2005;13(1):59–68.
39 Emslie GJ, Rush AJ, Weinberg WA, et al. A double-blind, randomized, placebo-controlled trial of fluoxetine in children and adolescents with depression. *Arch Gen Psychiatry*. 1997;54 (11):1031–1037.
40 Sackeim HA, Haskett RF, Mulsant BH, et al. Continuation pharmacotherapy in the prevention of relapse following electroconvulsive therapy: a randomized controlled trial. *JAMA*. 2001;285 (10):1299–1307.
41 Hughes CW, Emslie G, Kowatch R, Weinberg W, Rintelmann J, Rush AJ. Clinician, parent, and child prediction of medication or placebo in double-blind depression study. *Neuropsychopharmacology*. 2000;23(5):591–594.
42 Sharpe L, Ryan B, Allard S, Sensky T. Testing for the integrity of blinding in clinical trials: how valid are forced choice paradigms? *Psychother Psychosom*. 2003;72(3):128–131.
43 Bang H, Ni L, Davis CE. Assessment of blinding in clinical trials. *Control Clin Trials*. 2004;25(2):143–156.
44 Freed B, Assall OP, Panagiotakis G, et al. Assessing blinding in trials of psychiatric disorders: a meta-analysis based on blinding index. *Psychiatry Res*. 2014;219(2):241–247.
45 Bang H, Flaherty S, Kolahi J, Park J. Blinding assessment in clinical trials: A review of statistical methods and a proposal of blinding assessment protocol. *Clinic Res Regulat Aff*. 2010;27:42–51.
46 Waite JC. *Assessing Blinding in Randomized Clinical Trials [Masters Thesis]*. Pomona: California State Polytechnic University; 2017.
47 Kolahi J, Heejung B, Jongbae P, Desbiens N. CONSORT 2010 and controversies regarding assessment of blindness in RCTs. *Dent Hypotheses*. 2010;1:99–105.
48 Naudet F, Millet B, Reymann JM, Falissard B. Improving study design for antidepressant effectiveness assessment. *Int J Methods Psychiatr Res*. 2013;22(3):217–231.
49 Muthukumaraswamy SD, Forsyth A, Lumley T. Blinding and expectancy confounds in psychedelic randomized controlled trials. *Expert Rev Clin Pharmacol*. 2021;14(9):1133–1152.
50 Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*. 1995;273 (5):408–412.
51 Furukawa TA, Kessler RC. Why has prevalence of mental disorders not decreased as treatment has increased? *Austral New Zeal J Psychiatry*. 2019;53(12):1143–1144.
52 Middleton H, Moncrieff J. 'They won't do any harm and might do some good': time to think again on the use of antidepressants? *Br J Gen Pract*. 2011;61(582):47.