Psychiatry Research 307 (2022) 114297

Contents lists available at ScienceDirect

Psychiatry Research

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

# A systematic review and meta-analysis of the success of blinding in antidepressant RCTs

Amelia J Scott [a,b,*], Louise Sharpe [a], Ben Colagiuri [a]

[a] School of Psychology, University of Sydney, Sydney, NSW, Australia
[b] Department of Psychology, eCentreClinic, Macquarie University, Sydney, NSW, Australia

ARTICLE INFO

*Keywords:*
Randomized controlled trials
Blinding
Depression
Antidepressants

ABSTRACT

Successful blinding in double-blind RCTs is crucial for minimizing bias, however studies rarely report information about blinding. Among RCTs for depression, the rates of testing and success of blinding is unknown. We conducted a systematic review and meta-analysis of the rates of testing, predictors, and success of blinding in RCTs of antidepressants for depression. Following systematic search, further information about blinding assessment was requested from corresponding authors of the included studies. We reported the frequency of blinding assessment across all RCTs, and conducted logistic regression analyses to assess predictors of blinding reporting. Participant and/or investigator guesses about treatment allocation were used to calculate Bang's Blinding Index (BI). The BI between RCT arms was compared using meta-analysis. Across the 295 included trials, only 4.7% of studies assessed blinding. Pharmaceutical company sponsorship predicted blinding assessment; unsponsored trials were more likely to assess blinding. Meta-analysis suggested that blinding was unsuccessful among participants and investigators. Results suggest that blinding is rarely assessed, and often fails, among RCTs of antidepressants. This is concerning considering controversy around the efficacy of antidepressant medication. Blinding should be routinely assessed and reported in RCTs of antidepressants, and trial outcomes should be considered in light of blinding success or failure.

## 1. Introduction

Depression is a prevalent and burdensome condition that affects 1 in 10 adults (Kessler et al., 2005). Double-blind randomized placebo-controlled trials (RCTs) are considered the gold standard for identifying safe and efficacious treatments to manage depression. However, recent research in other areas (e.g. pain, (Colagiuri et al., 2019) general medicine (Hróbjartsson et al., 2007), orthopaedics (Karanicolas et al., 2008)) has called into question the validity of these highly relied upon trials based on evidence that patient blinding often fails. Understanding the success of blinding in trials of antidepressant medications is critical for validly evaluating the safety and efficacy of these widely-used medications, particularly given claims regarding large placebo responses in these trials (Khan et al., 2005).

Double-blinding is a key methodological feature of placebo-controlled RCTs. It involves concealing patients' treatment allocation (i.e. active or placebo) from both patients and investigators. Successful blinding occurs when patients and investigators cannot guess which treatment patients have been allocated to. This ensures that patients' and investigators' beliefs about treatment allocation are evenly distributed across trial arms (Colagiuri, 2010). As such, successful blinding ensures that treatment effects are not unduly influenced by patient or investigator biases, such as, placebo effects (Kirsch and Sapirstein, 1998), experimenter effects (Hróbjartsson et al., 2012), demand characteristics (McCambridge et al., 2014), and self-report or rater biases (Hróbjartsson et al., 2012). By contrast, failed blinding occurs when patients and/or investigators can guess the patients' treatment allocation at rate better than chance, and may lead to overestimation of treatment efficacy due to the above biases.

Concerningly, numerous recent reviews in trials for other conditions indicate that blinding is rarely assessed, and often unsuccessful (Fergusson et al., 2004; Karanicolas et al., 2008). A review of RCTs published in leading medical and psychiatry journals identified that blinding assessments (i.e. where investigators systematically collect participants' and/or investigators' guesses as to their treatment allocations) were reported in only 7% of studies (Fergusson et al., 2004). Similarly, Hrojbartsson et al. (2007) examined the frequency of blinding assessment among RCTs registered on the Cochrane Central Register of

* Corresponding author at: Department of Psychology, Macquarie University, Balaclava Rd, Macquarie Park NSW Australia.
*E-mail address:* amelia.scott@mq.edu.au (A.J. Scott).

https://doi.org/10.1016/j.psychres.2021.114297
Received 13 October 2021; Received in revised form 14 November 2021; Accepted 18 November 2021
Available online 24 November 2021
0165-1781/© 2021 Elsevier B.V. All rights reserved.

Controlled Trials, and identified that 2% of RCTs reported a blinding assessment. A recent systematic review and meta-analysis within chronic pain yielded similarly low rates of blinding assessment (5.6%; Colagiuri et al., 2019). As reporting on blinding success is not mandated by reporting guidelines like the CONSORT statement, it may be that blinding is assessed more often than it is reported. However, some studies have sought to examine this by contacting study authors for further information (Colagiuri et al., 2019; Hróbjartsson et al., 2007). and the rates of blinding assessment increase marginally, but remain low, e.g. 5.6% (Colagiuri et al., 2019) to 12% (Hróbjartsson et al., 2007), suggesting it is not merely a lack of reporting, but in fact a lack of testing.

Evidence from similar reviews also suggests that, when blinding is assessed, it often fails. In their review of chronic pain RCTs, Colagiuri et al. (2019) found that, overall, blinding was unsuccessful. Specifically, participants allocated to active treatment were significantly more likely to guess they had been given active treatment, compared to those receiving placebo. Similarly, among the 31 trials on the Cochrane register that assessed blinding, Hrojbartsson et al. (2007) reported that blinding was unclear or unsuccessful in more than half (55%) the trials.

These findings warrant significant cause for concern due to the potential bias introduced by failed blinding. Experimental studies (Colagiuri and Boakes, 2010) and retrospective analysis of clinical trials (Bausell et al., 2005; Colagiuri et al., 2009) repeatedly show that participants who believe they are receiving active treatment experience greater improvement than those who believe they are receiving a placebo, irrespective of their actual treatment. In fact, a recent meta-analysis estimated that unblinding exaggerates the between-group treatment effect for patient-reported outcomes by a standardized mean difference (Cohen's *d*) of 0.56 (Hróbjartsson et al., 2014). Notably, that effect size is greater than the drug-placebo difference among antidepressant trials, which was estimated in a recent individual patient data meta-analysis to be $d = 0.20$ (Fournier et al., 2010). As such, evaluating and understanding blinding is especially important in RCTs for depressive disorders, which are among the most prone to placebo effects (Khan et al., 2005).

Despite its importance, there have been no systematic attempts to evaluate blinding in RCTs of antidepressants to date. The closest to this is Baethge et al. (2013) who examined blinding assessment practices across RCTs within schizophrenia and affective disorders between the years 2000 and 2010. Their review also covered a range of treatments (e.g. brain stimulation, psychotherapy, drug treatment). Overall, the rate of blinding assessment was estimated at 2.7%, though this differed across conditions and treatment methods. For example, the rate of blinding assessment was lower among trials in schizophrenia (0.7%), though higher among trials of transcranial magnetic stimulation (12%). In a randomly-selected sample of RCTs in leading psychiatry journals, Fergusson et al. (2004) identified that 8 of 94 trials reported a blinding assessment in patients, with only three reporting the results of blinding for investigators as well. Among the 8 trials that reported a blinding assessment, 4 reported potentially compromised blinding. Concerningly, only two of these trials were conducted among individuals with depressive disorders (postpartum depression, and premenstrual dysphoric disorder), and neither of these included participants with major depressive disorder.

Currently, therefore, the available evidence suggests that blinding is assessed in a minority of RCTs, and is often compromised when assessment results are reported. However, there is no large-scale evidence to inform us about the assessment and success of blinding among RCTs of antidepressants for depressive disorders. Furthermore, none of the existing reviews have systematically examined the entire literature, nor contacted study authors for potentially unpublished blinding information. In addition, it is not known whether these earlier reviews have compelled future RCTs to improve their blinding assessment practices over the past 10 years.

Given the exponentially rising popularity of antidepressant prescription and use within the Western world (Olfson and Marcus, 2009), it is crucial that we can accurately appraise the efficacy and safety of antidepressants in light of information about blinding. Therefore, the aim of this systematic review and meta-analysis was three-fold; first, we sought to examine the frequency of blinding assessment in RCTs of antidepressants for depression. Next, we aimed to examine what study factors predict an assessment of blinding. Finally, we sought to estimate the overall success of blinding, and what study characteristics predict failed blinding.

## 2. Methods

This systematic review and meta-analysis was prospectively registered on the PROSPERO register (CRD42018100859). Studies were included if they were placebo-controlled randomized controlled trials that aimed to determine the efficacy of a currently approved and regulated antidepressant medication. Studies were included if participants were adults (aged over 18) meeting criteria for a depressive disorder, according to recognized diagnostic guidelines (e.g. DSM, ICD). Studies were included if blinding occurred in at least the patient group (i.e. single-blind). Post-hoc or long-term analyses of previously published RCTs were excluded, as were pooled analyses or sub-studies. Also excluded were interventions where participants were randomized to non-drug treatments alongside antidepressants that may compromise blinding (e.g. exercise or psychotherapy).

### 2.1. Search strategy and study selection

The electronic databases MEDLINE, Embase, PsycINFO and the Cochrane Register of Controlled Trials were systematically searched from inception up to June 2020. Search terms were related to three stems; depressive disorders, antidepressant medications, and randomized controlled trials. Please see Table e1 for a full description of study search terms. Titles and abstracts were reviewed by a single researcher, with a random 10% reviewed by a second researcher. All full-text articles were independently reviewed by two researchers. Disagreements were resolved through consensus.

### 2.2. Data extraction

Data from each study were extracted into a pre-designed coding sheet. Data extracted included trial and sample characteristics, potential predictors of blinding assessment, and blinding-related information (including whether a blinding assessment was reported, among whom blinding was assessed, the method of assessment, and any results reported such as the distribution of treatment guesses in participants and/or investigators). To accurately appraise the rate of blinding assessment in studies where such information was not reported, all contactable authors were e-mailed with a request to indicate whether blinding was assessed. Follow-up emails were sent on a second occasion, two weeks later.

### 2.3. Data analysis

First, the proportion of studies that reported a test for blinding was calculated. This outcome variable was coded as '1′ indicating a blinding assessment was reported, or '0′ where blinding assessment was not assessed, or no information was provided. We then conducted separate logistic regression analyses to examine predictors of whether a blinding test was reported, including; year of publication, sample size, trial sponsorship (including fully sponsored, partially sponsored or not sponsored), trial length (in weeks), type of depressive disorder, and antidepressant class. We repeated these analyses including only those studies where blinding assessment was known definitively (i.e. those studies that reported assessing or not assessing blinding). Results were considered statistically significant when $p < .05$. Analyses were conducted in SPSS (Version 25; SPSS, Inc, Chicago, IL).

In order to calculate the success of blinding, we used Bang's Blinding Index (BI; Bang et al., 2004)). Bang's BI is a chance-corrected index that ranges from −1 to 1. A BI of 1 indicates all individuals guessed correctly, and a BI of −1 indicates all individuals guessed incorrectly. As such, a BI of 0 reflects chance guessing (i.e. perfect blinding). For this analysis, the BI was calculated for each arm, and was entered (along with its standard deviation) into Comprehensive Meta-analysis (CMA, V3). We then calculated the standardized mean difference (Hedges'g) between the active BI and placebo BI using random effects models. For this analysis, the BI for the placebo arm was reverse-coded (therefore, a score > 0 suggested participants were more likely to guess they received active medication). As such, the Hedges'g obtained reflected the difference in likelihood of individuals guessing they had been allocated to active treatment. We also examined between-study heterogeneity by examining the Cochrane's Q, which suggests whether statistically significant heterogeneity is present (Higgins et al., 2003), and $I^2$, which estimates the percentage of variance across studies that is not due to chance (Higgins and Thompson, 2002) Potential publication bias was tested by viewing the funnel plot for asymmetry, and by conducting the Egger weighted regression method as a statistical test of funnel plot asymmetry (Egger et al., 1997).

## 3. Results

The initial searches yielded 9941 articles, of which 715 remained after title and abstract review. Following full-text review, 295 studies (representing 71,448 participant) met full eligibility criteria and were retained for data extraction (see Fig. 1). See Table 1 for summary statistics of all included studies.

### 3.1. Rates and predictors of blinding assessment

Following data extraction, information on blinding assessment was only available in 14 papers (3 studies stated that blinding was not assessed while the remaining 11 reported a blinding assessment). Of the remaining 281 studies without blinding information, 91 authors were not contactable, as e-mail address information was both unavailable in the published manuscript (typical of papers published prior to the year 2000) and/or author contact details could not be located online. Of the remaining 190 authors contacted via e-mail, 28 responded with further information. Two authors (representing three further RCTs) reported assessing blinding but no longer had access to blinding data, while 26 authors stated they did not assess blinding. A further 3 authors responded though stated they were unable to provide further information, while the remaining 159 authors did not reply or could not be reached.

### 3.2. Blinding assessment frequency and predictors

Based on the available information, 14 of 295 (4.7%) of studies reported an assessment of blinding of participants or investigators. Regarding predictors of blinding assessment, trial sponsorship significantly predicted whether blinding was assessed – studies with partial sponsorship (OR 4.20, 95% CI 1.01 – 17.47) and no sponsorship (OR 12.85, 95% CI 3.20 – 52.15) were significantly more likely to assess blinding, compared to fully sponsored studies (Table 2). Studies examining the efficacy of antidepressants in trials recruited mixed depressive disorders or other depressive disorders (e.g. seasonal affective disorder) were also more likely to assess blinding (OR 5.66, 95% CI 1.07 – 29.85; see table 2). No other study characteristics emerged as significant

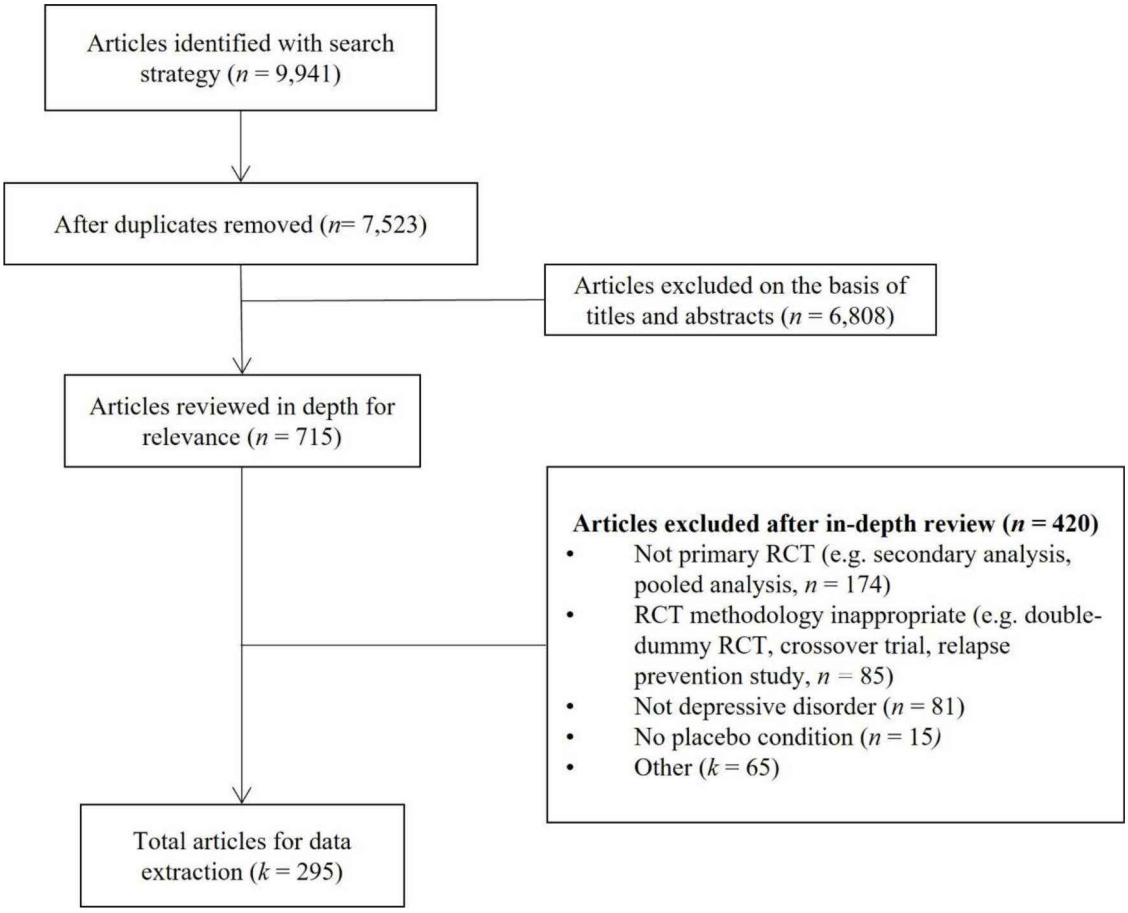


**Fig. 1.** Search and retrieval process.

**Table 1**
Summary of included study characteristics.

| Characteristic | Frequency (%) |
|---|---|
| **Year of Study** | |
| 1970 – 1980 | 4 (1.4) |
| 1981 – 1990 | 40 (13.6) |
| 1991 – 2000 | 96 (32.5) |
| 2001 – 2010 | 86 (29.2) |
| 2011 – 2020 | 69 (23.4) |
| **Depressive Disorder** | |
| Major Depressive Disorder | 261 (88.5) |
| Persistent Depressive Disorder | 10 (3.4) |
| Premenstrual Dysphoric Disorder | 12 (4.1) |
| Mixed / Other depressive disorders | 12 (4.1) |
| **Recruitment Setting** | |
| Inpatient | 14 (4.7) |
| Outpatient | 271 (94.9) |
| Mixed / Not specified | 10 (3.4) |
| **Number of active treatment arms** | |
| One | 144 (48.8) |
| Two | 122 (41.4) |
| Three | 26 (8.8) |
| Four | 3 (1.0) |
| **Primary antidepressant class** | |
| Atypical | 40 (13.6) |
| Azapirone | 2 (0.7) |
| MAOI | 15 (5.1) |
| NRI | 5 (1.7) |
| SARI | 9 (3.1) |
| SNRI | 52 (17.6) |
| SSRI | 116 (39.3) |
| TCA | 47 (15.9) |
| TeCA | 8 (2.7) |

*Note,* MAOI = monoamine oxidase inhibitor, NRI = norepinephrine reuptake inhibitor, SARI = serotonin antagonist and reuptake inhibitor, SNRI = serotonin and norepinephrine reuptake inhibitor, SSRI = selective serotonin reuptake inhibitor, TCA = tricyclic, TeCA = tetracyclic.

**Table 2**
Logistic regression analyses predicting studies' assessment of blinding.

| Predictor | Studies | OR (95% CI) | Wald statistic | *p* value |
|---|---|---|---|---|
| Year of Publication[a] | | .99 (0.94, 1.04) | .141 | .71 |
| Sample Size[a] | | 1.0 (0.99, 1.0) | 3.58 | .06 |
| Pharmaceutical involvement | | | | |
| Fully sponsored (ref) | 185 | | | |
| Partial involvement | 18 | 4.2 (1.01, 17.47) | 3.91 | .04 |
| No sponsorship | 44 | 12.85 (3.2, 52.15) | 12.76 | .000 |
| Trial length (weeks)[a] | | 1.0 (0.94, 1.07) | .00 | 1.0 |
| Depressive Disorder[b] | | | | |
| Major Depressive Disorder (ref) | 261 | | | |
| Persistent Depressive Disorder | 10 | 2.52 (0.29, 21.65) | .71 | .40 |
| Mixed/other | 12 | 5.66 (1.07, 29.85) | 4.17 | .04 |
| Antidepressant class[b] | | | | |
| SSRI (ref) | 116 | | | |
| SNRI | 52 | .26 (0.03, 2.15) | 1.55 | .21 |
| TCA | 47 | .91 (0.23, 3.60) | .01 | .90 |
| Other | 25 | 1.22 (0.24, 6.12) | .06 | .81 |

[a] Results for year of publication, sample size, and trial length indicate the OR change per one-unit increase in these predictors.

[b] Studies among participants with premenstrual dysphoric disorder, and studies of atypical and MAOI antidepressants were removed from logistic regression analyses as no studies reported an assessment of blinding.

predictors of blinding assessment. Sensitivity analyses including only those studies that reported assessing ($n = 14$) and not assessing ($n = 26$) revealed a mostly similar pattern of results. Compared to fully sponsored trials, RCTs with no sponsorship remained significantly more likely to assess blinding (OR 21.00, 95% CI 1.98, 222.00, $p = .01$; table 2). However, there was a non-significant difference between partially sponsored and fully sponsored studies (OR 3.36, 95% CI 0.65, 17.27; table 2). In addition, there was no difference in the likelihood of assessing blinding depending upon the type of depressive disorder. All other predictor variables remained non-significant.

### 3.3. Success of blinding

#### 3.3.1. Participants

Of the 14 studies that reported an assessment of blinding, data regarding participant guesses was available in seven of these studies (see Table 3). We observed significant heterogeneity across these studies ($Q_6$ 181.03, $I^2$ 96.69). No evidence of publication bias was detected, supported by a non-significant Egger regression test ($p = .20$). The overall Hedges' g derived from comparing active and placebo BIs was 3.07 (95% CI 1.74, 4.41, $p < .0005$). These results suggest that participants allocated to active treatment were significantly more likely to guess they had received active medication, compared to placebo participants. Post-hoc sensitivity analysis was conducted removing one outlier (Rabkin et al., 1986) and also removing trials with more than one active treatment arms, and the Hedges' g remained significant ($p < .0005$ in both cases; see Table 3). Due to the limited number of studies with blinding data available, we could not examine potential predictors of blinding success in participants.

#### 3.3.2. Investigators

Data regarding investigator blinding was available for only four studies (see Table 4). Again, there was significant heterogeneity across these studies ($Q_3$ 240.84, $I^2$ 98.75). The Egger regression test was marginally non-significant ($p = .05$). The overall Hedges' g derived from comparing active and placebo BIs for investigators was 5.20 (95% CI 2.27, 8.13, $p = .001$). After removing Rabkin and colleagues (Rabkin et al., 1986) from the analysis, the Hedges' g obtained was marginally non-significant ($g = 1.93$, 95% CI −0.02, 3.87, $p = .053$). As with participant data, there were insufficient studies to examine potential predictors of blinding success among investigators.

## 4. Discussion

Despite growing concern regarding the assessment and success of blinding among RCTs, there has been a scarcity of evidence about blinding assessment practices and outcomes in trials of antidepressants for depression. This is the first known review of blinding assessment frequency and its success among antidepressant RCTs for depression. We identified 295 double-blind, placebo-controlled RCTs for depression, including over 70,000 participants. Despite this large pool of studies and participants, only 14 RCTs reported a blinding assessment which is fewer than 5%, even after contacting authors where blinding was not reported in the original paper. While these results are consistent with previous reviews in other domains, such reviews have unanimously recommended improvements to blinding assessment and reporting as a method of improving the validity of RCTs. It appears that these calls have received little attention.

Although blinding data could only be obtained from a small number of studies, our results clearly show that participants in active treatment arms were significantly more likely to guess they had received active treatment compared to those receiving placebo. Similarly, study investigators were also more likely to correctly guess the treatment assignment of participants in active treatment arms compared to placebo. This finding raises significant concerns in light of existing evidence regarding the implications of un-blinding among RCTs. As discussed,

**Table 3**
Blinding guess data for participants.

| Study | Assigned to Active | | | Assigned to Placebo | | | Blinding Index | | Hedges' g (95% CI) |
|---|---|---|---|---|---|---|---|---|---|
| | N | Guess active (%) | Guess placebo (%) | N | Guess active (%) | Guess placebo (%) | Active BI (SD) | Placebo BI (SD) | |
| Richard et al., 2012 | 76 | 55 (72) | 22 (28) | 39 | 26 (67) | 13 (33) | .43 (0.10) | .30 (0.15) | 1.08 (0.67, 1.48) |
| Devanand et al., 2005 | 44 | 28 (64) | 16 (36) | 46 | 22 (48) | 24 (52) | .27 (0.15) | .04 (0.15) | 2.05 (1.54, 2.56) |
| Kleber et al., 1983 | 20 | 13 (65) | 7 (35) | 24 | 9 (38) | 15 (63) | .30 (0.21) | .25 (0.20) | 2.64 (1.84, 3.44) |
| Hypericum Depression Trial Study Group, 2002 | 75 * | 37 (49) | 17 (23) | 100 | 24 (24) | 39 (39) | .3 (0.12) | −0.62 (0.24) | 3.69 (3.20. 4.18) |
| Edwards and Goldie, 1993 | 19 | 12 (63) | 7 (37) | 20 | 12 (60) | 8 (40) | .26 (0.22) | −0.41 (0.22) | 2.99 (2.09, 3.90) |
| Rabkin et al., 1986 | 63 * | 56 (89) | 7 (11) | 37 | 15 (41) | 22 (59) | .78 (0.07) | −0.19 (0.16) | 8.61 (7.35, 9.87) |
| Mischoulon et al., 2013 | 23 * | 7 (30) | 4 (17) | 25 | 8 (32) | 7 (28) | .27 (0.29) | −0.07 (0.26) | .71 (−0.07, 1.49) |

* These studies contained more than one active treatment arm, where the other active comparator was not an approved/regulated antidepressant. A sensitivity analysis was conducted removing these trials and the pattern of results remained unchanged (Hedges' G 1.77, 95% CI 1.48, 2.05).

**Table 4**
Blinding guess data for investigators.

| Study | Assigned to Active | | | Assigned to Placebo | | | Blinding Index | | Hedges' g (95% CI) |
|---|---|---|---|---|---|---|---|---|---|
| | N | Guess active (%) | Guess placebo (%) | N | Guess active (%) | Guess placebo (%) | Active BI (SD) | Placebo BI (SD) | |
| Richard et al., 2012 | 76 | 57 (75) | 19 (25) | 39 | 24 (62) | 15 (38) | .50 (0.1) | .23 (0.15) | 2.3 (1.8, 2.7) |
| Devanand et al., 2005 | 44 | 28 (64) | 16 (36) | 46 | 29 (63) | 17 (37) | .27 (0.15) | .26 (0.14) | .06 (−0.35, 0.47) |
| Kleber et al., 1983 | 31 | 17 (55) | 14 (45) | 25 | 6 (24) | 19 (76) | .10 (0.18) | −0.52 (0.17) | 3.5 (2.7, 4.3) |
| Rabkin et al., 1986 | 86 | 77 (90) | 9 (10) | 51 | 9 (18) | 42 (82) | .79 (0.07) | −0.65 (0.11) | 16.5 (14.1, 18.8) |

available evidence from RCTs suggests that patient outcomes among unblinded trials are exaggerated by $d = 0.56$ on average, and between 0.41 and 0.71 (Hróbjartsson et al., 2014). This substantial difference is likely due to the well-documented effect of treatment expectancies (i.e. participants' beliefs about the efficacy of treatment, along with their perceived treatment assignment) on symptom improvement (Colagiuri, 2010). In addition to other non-specific factors like spontaneous improvement and regression to the mean, the use of blinded placebo controls in RCTs is intended to control for the confounding effect of expectancy on treatment outcomes. However, this practice is only successful when participant expectancies are actually evenly distributed between treatment arms, i.e. when blinding is actually achieved. It is alarming that so few trials seek to confirm that blinding has been successful, and that in those that do, it is often found to have failed.

We also examined whether any study or participant characteristics predicted whether or not blinding was assessed. Our results suggest that industry-sponsored studies were significantly less likely to assess blinding, compared to unsponsored studies. This association remained significant among the subsample of studies for which information about blinding assessment was certain (i.e. 'yes' or 'no'). It is unclear why this is the case and whether it reflects biases in pharmaceutical trials.

It is important to mention the role and potential influence of the CONSORT group on the low rates of blinding assessments. The CONSORT statement is intended to improve reporting of RCTs, and was first released in 2001. Since its introduction, the conduct and reporting of RCTs has greatly improved (Plint et al., 2006). However, the CONSORT group (surprisingly to our minds) removed the recommendation in their 2010 statement that researchers evaluate and report upon the success of blinding (Schulz et al., 2010). This was on the basis of an argument that blinding assessment merely (and only) reflects treatment efficacy. However, this is a questionable argument. For example, one included RCT identified that the symptom improvements reported among those receiving active treatment (sertraline) did not differ between those who guessed they were vs. were not in the active treatment arm (mean change was −11.6 vs. −11.9, respectively; Hypericum Depression Trial Study Group, 2002). In addition, Laferton et al. (2018) found that perceived treatment assignment (rather than *actual* treatment assignment) was prospectively associated with symptom improvement at the end of treatment in their RCT within depression. There is also evidence that other cues, such as medication side effects, may be associated with unsuccessful blinding (Altman et al., 2001). For example, Colagiuri et al. (2019)'s review of blinding success in chronic pain identified that adverse effects was the most robust predictor of unsuccessful blinding. In addition, the Cochrane Risk of Bias Tool suggests that, if blinded participants experience side effects or toxicities known to be specific to the intervention, then blinding could be considered compromised (Sterne et al., 2019). On the basis of these concerns, there is growing interest in and use of active placebos, which mimic the side effects of the treatment under investigation, and therefore reduce the risk of bias associated with unblinding (Laursen et al., 2020).

While these findings suggest that failed blinding may be due to features other than treatment efficacy, ultimately the lack of existing data prevents any empirical investigation of both the causes of failed blinding, and the consequences of failed blinding on treatment outcomes. In order to properly investigate the causes and consequences of failed blinding, it is crucial that future RCTs evaluate participant and investigator guesses regarding treatment allocation at multiple timepoints during the trial. We also recommend that researchers assess participant expectancies alongside their guesses about treatment allocation, in order to understand the contribution of both factors on treatment outcomes. In addition, the reporting of blinding information in RCTs should be considered an indicator of quality and risk of bias, and incorporated into widely used reporting guidelines, such as the CONSORT statement and the Cochrane Risk of Bias tool (Schulz et al., 2010; Sterne et al., 2019). We acknowledge these practices reflect a small additional burden on participants, and/or resource requirements for investigators. However, double-blind placebo-controlled RCTs reflect the cornerstone of evidence-based medicine, and their outcomes are used to determine the prescribing practices of healthcare providers worldwide. Thus, the introduction of such measures reflects a small

burden relative to their value in helping researchers to better interpret the evidence of placebo-controlled RCTs.

The results of this review represent an exhaustive retrieval and synthesis of blinding assessments within antidepressant RCTs for depressive disorders. However, it should be noted that the available data analyzed with regard to the success of blinding reflects a small proportion of the total number of studies and participants in the review. As noted, there was therefore insufficient data to examine what factors are associated with unsuccessful blinding. An associated limitation relates to the accessibility of further information and data from study authors. The majority (86%) of authors contacted either could not be reached or did not respond with information. It may be that the true rate of blinding differs from our results, though this appears unlikely considering our findings are consistent with similarly-conducted reviews in other areas (Fergusson et al., 2004; Hróbjartsson et al., 2007; Karanicolas et al., 2008). This limitation reflects a broader issue in regarding the availability of further information and study data after completion and publication. While it is encouraging to observe that efforts to increase reporting transparency and data availability post-publication are becoming more popular,(Gewin, 2016) there remain other barriers to obtaining necessary information that is not reported in original articles (e.g. individuals changing institution, reluctance to provide data).

In summary, this is the first known review of the rates and success of blinding assessment among antidepressant RCTs within depression. Our findings suggest that blinding is not assessed in the large majority of RCTs, and when blinding is assessed, it often fails. In the absence of alternative evidence, our results suggest that participant expectancies are not successfully controlled for among antidepressant RCTs. This represents a substantial threat to the validity of this large and important evidence base. However, the exact influence of failed blinding on treatment outcomes is difficult to appraise, given how rarely it is reported. It is crucial that future RCTs assess and report on the results of participant and investigator blinding in order to protect the validity of this body of research and improve our confidence in the safety and efficacy of antidepressant medication.

## Author statements

This research was supported by Australian Research Council Discovery Project Grant (DP150104026) awarded to BC and LS. The authors declare no financial relationships with any organisations that might have an interest in the submitted work, and no other relationships or activities that could appear to have influenced the submitted work.

## Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

## CRediT authorship contribution statement

**Amelia J Scott:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Writing – original draft, Writing – review & editing. **Louise Sharpe:** Conceptualization, Supervision, Writing – review & editing. **Ben Colagiuri:** Conceptualization, Formal analysis, Investigation, Project administration, Supervision, Writing – review & editing.

## Declaration of Competing Interest

BC and LS are supported by an Australian Research Council Discovery Project Grant (DP150104026). The authors declare no financial relationships with any organisations that might have an interest in the submitted work, and no other relationships or activities that could appear to have influenced the submitted work.

## Acknowledgements

We thank Dr Shannon Web for assistance with title and abstract, and full-text screening.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.psychres.2021.114297.

## References

Altman, D.G., Schulz, K.F., Moher, D., Egger, M., Davidoff, F., Elbourne, D., Gøtzsche, P. C., Lang, T., 2001. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. Ann. Intern. Med. 134 (8), 663–694.

Baethge, C., Assall, O.P., Baldessarini, R.J., 2013. Systematic review of blinding assessment in randomized controlled trials in schizophrenia and affective disorders 2000-2010. Psychother. Psychosom. 82 (3), 152–160.

Bang, H., Ni, L., Davis, C.E., 2004. Assessment of blinding in clinical trials. Control. Clin. Trials 25 (2), 143–156.

Bausell, R.B., Lao, L., Bergman, S., Lee, W.-L., Berman, B.M., 2005. Is acupuncture analgesia an expectancy effect? Preliminary evidence based on participants' perceived assignments in two placebo-controlled trials. Eval. Health Prof. 28 (1), 9–26.

Colagiuri, B., 2010. Participant expectancies in double-blind randomized placebo-controlled trials: potential limitations to trial validity. Clin. Trials 7 (3), 246–255.

Colagiuri, B., Boakes, R.A., 2010. Perceived treatment, feedback, and placebo effects in double-blind RCTs: an experimental analysis. Psychopharmacology 208 (3), 433–441.

Colagiuri, B., Morley, K., Boakes, R., Haber, P., 2009. Expectancy in double-blind placebo-controlled trials: an example from alcohol dependence. Psychother. Psychosom. 78 (3), 167–171.

Colagiuri, B., Sharpe, L., Scott, A., 2019. The blind leading the not-so-blind: a meta-analysis of blinding in pharmacological trials for chronic pain. J. Pain 20 (5), 489–500.

Devanand, D., Nobler, M.S., Cheng, J., Turret, N., Pelton, G.H., Roose, S.P., Sackeim, H. A., 2005. Randomized, double-blind, placebo-controlled trial of fluoxetine treatment for elderly patients with dysthymic disorder. Am. J. Geriatr. Psychiatry 13 (1), 59–68.

Edwards, J.G., Goldie, A., 1993. Placebo-controlled trial of paroxetine in depressive illness. Hum. Psychopharmacol. 8 (3), 203–209.

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.

Fergusson, D., Glass, K.C., Waring, D., Shapiro, S., 2004. Turning a blind eye: the success of blinding reported in a random sample of randomised, placebo controlled trials. BMJ 328 (7437), 432.

Fournier, J.C., DeRubeis, R.J., Hollon, S.D., Dimidjian, S., Amsterdam, J.D., Shelton, R. C., Fawcett, J., 2010. Antidepressant drug effects and depression severity: a patient-level meta-analysis. JAMA 303 (1), 47–53.

Gewin, V., 2016. Data sharing: an open mind on open data. Nature 529 (7584), 117–119.

Higgins, J.P., Thompson, S.G., 2002. Quantifying heterogeneity in a meta-analysis. Stat. Med. 21 (11), 1539–1558.

Higgins, J.P., Thompson, S.G., Deeks, J.J., Altman, D.G., 2003. Measuring inconsistency in meta-analyses. BMJ 327 (7414), 557–560.

Hróbjartsson, A., Emanuelsson, F., Skou Thomsen, A.S., Hilden, J., Brorson, S., 2014. Bias due to lack of patient blinding in clinical trials. A systematic review of trials randomizing patients to blind and nonblind sub-studies. Int. J. Epidemiol. 43 (4), 1272–1283.

Hróbjartsson, A., Forfang, E., Haahr, M., Als-Nielsen, B., Brorson, S., 2007. Blinded trials taken to the test: an analysis of randomized clinical trials that report tests for the success of blinding. Int. J. Epidemiol. 36 (3), 654–663.

Hróbjartsson, A., Thomsen, A.S.S., Emanuelsson, F., Tendal, B., Hilden, J., Boutron, I., Ravaud, P., Brorson, S., 2012. Observer bias in randomised clinical trials with binary outcomes: systematic review of trials with both blinded and non-blinded outcome assessors. BMJ 344.

Hypericum Depression Trial Study Group, 2002. Effect of Hypericum perforatum (St John's wort) in major depressive disorder: a randomized controlled trial. JAMA 287 (14), 1807–1814.

Karanicolas, P.J., Bhandari, M., Taromi, B., Akl, E.A., Bassler, D., Alonso-Coello, P., Rigau, D., Bryant, D., Smith, S.E., Walter, S.D., 2008. Blinding of outcomes in trials of orthopaedic trauma: an opportunity to enhance the validity of clinical trials. JBJS 90 (5), 1026–1033.

Kessler, R.C., Berglund, P., Demler, O., Jin, R., Merikangas, K.R., Walters, E.E., 2005. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch. Gen. Psychiatry 62 (6), 593–602.

Khan, A., Kolts, R.L., Rapaport, M.H., Rama Krishnan, K.R., Brodhead, A.E., Brown, W.A., 2005. Magnitude of placebo response and drug-placebo differences across psychiatric disorders. Psychol. Med. 35 (5), 743–749.

Kirsch, I., Sapirstein, G., 1998. Listening to Prozac but hearing placebo: a meta-analysis of antidepressant medication. Prevent. Treatment 1 (2), 2a.

Kleber, H.D., Weissman, M.M., Rounsaville, B.J., Wilber, C.H., Prusoff, B.A., Riordan, C. E., 1983. Imipramine as treatment for depression in addicts. Arch. Gen. Psychiatry 40 (6), 649–653.
Laferton, J.A., Vijapura, S., Baer, L., Clain, A.J., Cooper, A., Papakostas, G., Price, L.H., Carpenter, L.L., Tyrka, A.R., Fava, M., 2018. Mechanisms of perceived treatment assignment and subsequent expectancy effects in a double blind placebo controlled RCT of major depression. Front Psychiatry 9, 424.
Laursen, D., Hansen, C., Paludan-Müller, A., Hróbjartsson, A., 2020. Active placebo versus standard placebo control interventions in pharmacological randomised trials. Cochrane Database Syst. Rev. 2020 (7).
McCambridge, J., Kypri, K., Elbourne, D., 2014. Research participation effects: a skeleton in the methodological cupboard. J. Clin. Epidemiol. 67 (8), 845–849.
Mischoulon, D., Price, L.H., Carpenter, L.L., Tyrka, A.R., Papakostas, G.I., Baer, L., Dording, C.M., Clain, A.J., Durham, K., Walker, R., 2013. A double-blind, randomized, placebo-controlled clinical trial of S-adenosyl-L-methionine (SAMe) versus escitalopram in major depressive disorder. J. Clin. Psychiatry 74 (4), 0-0.
Olfson, M., Marcus, S.C., 2009. National patterns in antidepressant medication treatment. Arch. Gen. Psychiatry 66 (8), 848–856.
Plint, A.C., Moher, D., Morrison, A., Schulz, K., Altman, D.G., Hill, C., Gaboury, I., 2006. Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. Med. J. Aust. 185 (5), 263–267.
Rabkin, J.G., Markowitz, J.S., Stewart, J., McGrath, P., Harrison, W., Quitkin, F.M., Klein, D.F., 1986. How blind is blind? Assessment of patient and doctor medication guesses in a placebo-controlled trial of imipramine and phenelzine. Psychiatry Res. 19 (1), 75–86.
Richard, I., McDermott, M., Kurlan, R., Lyness, J., Como, P., Pearson, N., Factor, S., Juncos, J., Ramos, C.S., Brodsky, M., 2012. A randomized, double-blind, placebo-controlled trial of antidepressants in Parkinson disease. Neurology 78 (16), 1229–1236.
Schulz, K.F., Altman, D.G., Moher, D., 2010. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. Trials 11 (1), 1–8.
Sterne, J.A., Savović, J., Page, M.J., Elbers, R.G., Blencowe, N.S., Boutron, I., Cates, C.J., Cheng, H.-Y., Corbett, M.S., Eldridge, S.M., 2019. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 366.