The Relationship Between Blinding Integrity and Treatment Efficacy in Randomised-Controlled Trials in Patients with Anxiety Disorders: A Systematic Review and Meta-Analysis

Ruqayyah Haq* [1], Laura Molteni[2,3], Nathan T.M. Huneke* [1,2,3]

1. Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, UK

2. Southern Health National Health Service Foundation Trust, Southampton, UK

3. University Department of Psychiatry, Academic Centre, College Keep, 4-12 Terminus Terrace, Southampton, SO14 3DT, UK

*Contributed equally

Corresponding author:
Nathan T.M. Huneke
Academic Centre, College Keep, 4-12 Terminus Terrace, Southampton, SO14 3DT, UK
n.huneke@soton.ac.uk

1. Abstract

Background: Blinding of patients and assessors is thought to minimise expectancy effects and biases in double-blind randomised-controlled trials (RCT's). However, whether blinding integrity should be assessed and reported remains debated. Furthermore, it is unknown whether blinding failure influences the outcome of RCT's in anxiety disorders. We aimed to understand whether blinding integrity is assessed and reported in anxiolytic RCT's. A secondary aim was to explore whether blinding integrity is associated with anxiolytic treatment efficacy.

Methods: We systematically searched for placebo-controlled randomised trials in adults with generalised and social anxiety disorders, and in panic disorder, from 1980 to present day. We extracted data regarding blinding integrity and treatment efficacy. Where assessments of blinding integrity were not reported, attempts were made to acquire them from authors. Where possible, we subsequently calculated Bang's Blinding Index, and assessed the association between blinding integrity and effect size of treatment compared with placebo through meta-regression.

Results: Of the 248 RCT's that met inclusion criteria, we were able to obtain assessments of blinding integrity from nine (3.63%). Overall, blinding failed in five of these trials (55.56%), but blinding was intact in 80% of placebo arms. We found a significant association between reduced blinding integrity among assessors and increased treatment effect size (beta's < -6.50, p's < 0.001), but this analysis involved only four studies. In patients, we saw a non-significant trend where reduced blinding integrity in the placebo groups was associated with *increased* treatment efficacy, which was not present in active medication arms.

Conclusions: Consistent with work in other psychiatric disorders, blinding integrity is rarely reported in anxiolytic RCT's. Where it is reported, blinding appears to often fail. We found signals that suggest unblinding of clinician assessors and of patients in placebo arms might be associated with larger treatment effect sizes. These analyses were based on limited data. We recommend that data regarding blinding integrity, along with the reasons patients and assessors offer for their beliefs regarding group allocation, are systematically collected in RCT's of anxiolytic treatments.

2. Introduction

Anxiety disorders are common and burdensome. Globally, anxiety disorders are the sixth largest cause of non-fatal health loss (World Health Organization, 2017); and due to impacts on social and occupational functioning, anxiety disorders result in marked socioeconomic and healthcare costs (Olesen et al., 2012; Hendriks et al., 2016). Considering this burden, improved treatments for anxiety disorders are needed.

The current 'gold standard' method of assessing a novel treatment is a randomised-controlled trial (RCT). These trials are designed such that sources of bias are minimised as much as possible (Howick, 2011). An important feature of RCT's is withholding information about treatment assignment from patients and assessors, also known as blinding (Anand et al., 2020). Through blinding assessors and patients, it is thought that observer and response biases can be reduced and thus enhance measurement of the efficacy of the intervention under study (Schulz and Grimes, 2002; Wood et al., 2008; Howick, 2011). However, the requirement to report whether blinding was evaluated and whether it was successful was removed in the 2010 CONSORT statement (Schulz et al., 2010a), mainly because it can be difficult to infer the reasons for blinding failure (Schulz et al., 2010b). Nonetheless, there is evidence that blinding reduces bias in RCT's (Fergusson et al., 2004; Hróbjartsson et al., 2007; Wood et al., 2008; Hróbjartsson et al., 2013).

Whether blinding integrity should be assessed and reported in RCT's remains a subject of debate (Moher et al., 2010; Howick et al., 2020; Webster et al., 2021). The main argument against is that it is impossible to separate blinding integrity from efficacy or other supplementary information (Sackett, 2004; Schulz et al., 2010b). In contrast, the main arguments for measuring blinding integrity is to check for possible observer or response biases and to check both arms are balanced for non-specific expectation effects (Fergusson et al., 2004; Howick et al., 2020; Webster et al., 2021). The latter is potentially important in psychiatry where outcomes are subjective and potentially vulnerable to biases and non-specific expectation effects (Faria et al., 2017; Rutherford et al., 2017; Hjorth et al., 2021; Huneke, 2022). Although there is evidence that blinding reduces bias in RCT's, it is unknown whether blinding *failure* might bias measures of treatment efficacy. Indeed, recent meta-

analyses of antidepressant trials are inconclusive on this point (Lin et al., 2022; Scott et al., 2022). To our knowledge, such analyses in anxiolytic trials have yet to be attempted.

In the current study, we systematically reviewed the literature for RCT's in anxiety disorders from 1980 onwards, and extracted data regarding blinding integrity and symptom reduction. Our aims were to ascertain whether blinding integrity is assessed and reported in anxiolytic trials; and whether blinding integrity is associated with measures of efficacy (i.e. whether it influences trial outcome).

3. Method

This systematic review and meta-analysis was carried out according to PRISMA guidelines (Page et al., 2021). Four reviewers (RH, LM, GW, NH) performed the systematic review and data extraction independently in pairs. All discrepancies were resolved by consensus. The protocol was registered prospectively with PROSPERO (CRD42022328750).

3.1. Literature Search and Selection

Our full search strategy is reported in the appendix. We searched eight databases (PubMed, Cochrane Central Register of Controlled Trials, PsycINFO, Embase + Embase classic, OVID MEDLINE, Google Scholar (first 100 pages), CINAHL, and Web of Science) for RCT's in generalised anxiety disorder, panic disorder, and social anxiety disorder published since 1980, with no language restrictions, on 05/05/2022 and updated on 16/08/2023. We additionally searched clinicaltrials.gov for unpublished trials.

At least two reviewers (RH, GW, or LM) screened all titles and abstracts against the following inclusion criteria: the study was a randomised trial involving a medication and placebo intervention; patients were adults aged 18 or older diagnosed with an anxiety disorder, and change in anxiety symptoms was an outcome measure. We obtained full texts for potentially eligible articles, which were then screened by at least two reviewers (RH, GW, or LM).

3.2. Data Extraction

One reviewer (RH) extracted data through the use of a piloted form. All extracted data were checked independently by a second reviewer (LM or NH). We extracted data regarding the patient population, study design, study findings, and recorded whether blinding integrity was assessed. Where blinding integrity was not reported, we contacted authors via email to inquire if they had conducted an assessment of blinding integrity. Authors who did not respond were sent two reminders at two-week intervals. If authors did not respond following either of these reminders then we recorded a non-response.

3.3. Quality Assessment

For studies with blinding integrity information, we assessed for risk of bias with the Cochrane Collaboration's risk of bias 2 tool for randomised trials (Sterne et al., 2019). One reviewer recorded risk of bias for each record using a standardised form (RH), and these assessments were independently checked by a second reviewer (NH). We assessed the risk of bias due to randomisation, deviations from the intended intervention, missing data, outcome measurement, and selective reporting.

3.4. Data Synthesis

We calculated the frequency of assessment of blinding integrity across all included RCT's. In those reporting an assessment of blinding, we quantified blinding integrity with Bang's blinding index (BI) (Bang et al., 2004) where data were available to do so:

$$BI = \frac{(Correct\ Guesses - Incorrect\ Guesses)}{Total\ Guesses}$$

BI closer to zero suggests blinding integrity, while BI closer to one suggests complete unblinding. We calculated BI for patients and assessors, for placebo and active medication groups, separately. We interpreted BI scores between -0.2 and 0.2 as intact blinding (Bang et al., 2004).

We carried out quantitative meta-analyses using the meta (Balduzzi et al., 2019) and metafor R packages (Viechtbauer, 2010). We initially calculated the between-group

standardised mean difference (Hedges' g) for the primary anxiety symptom-related outcome at the end of treatment for each study. We next conducted random effects meta-analysis using a restricted maximum likelihood (REML) estimator. Heterogeneity was assessed through the I^2, τ^2, and Q statistics. Next, to explore whether there was a relationship between blinding integrity and medication efficacy, we conducted meta-regressions with BI as a predictor and between-group effect size as the outcome variable.

4. Results

Our article selection process is summarized in Figure 1. Of 10,448 initial records (plus 1,096 from the updated search) we identified 248 RCT's that involved an active medication arm and a placebo arm in anxiety disorders (see appendix for full list of studies). Of the 248 RCT's we included, six (2.42%) reported an assessment of blinding integrity (Hoffart et al., 1993; Başoğlu et al., 1997; Bakker et al., 1999; Oosterbaan et al., 2001; Guastella et al., 2009; Lenze et al., 2009). We additionally contacted corresponding authors, where possible, to gather additional unreported information regarding blinding integrity. We were able to obtain contact details for the authors of 160 trials (64.52%) (see appendix). Of these, 46 (28.75%) responded, and reported that either blinding integrity was not assessed (n = 34), that the data were no longer accessible (n = 9), or supplied additional information regarding blinding integrity (n = 3) (Van Vliet et al., 1992; van Vliet et al., 1993; Kampman et al., 2002). Therefore, we were able to obtain blinding integrity data for 9 of 248 RCT's (3.63%) in anxiety disorders since 1980. The characteristics of these trials are summarised in Table 1.

Studies for which we obtained a blinding integrity assessment were published between 1992 and 2009. Five of the nine trials were in patients with panic disorder (Hoffart et al., 1993; van Vliet et al., 1993; Başoğlu et al., 1997; Bakker et al., 1999; Kampman et al., 2002), three were in patients with social anxiety disorder (Van Vliet et al., 1992; Oosterbaan et al., 2001; Guastella et al., 2009), and one was in patients with generalised anxiety disorder (Lenze et al., 2009). The total number of patients in these studies was n = 627: sample sizes ranged from 17 to 177 patients. More than half of the trials (55.56%) were carried out with samples fewer than 50 patients (Van Vliet et al., 1992; Hoffart et al., 1993; van Vliet et al., 1993; Kampman et al., 2002; Guastella et al., 2009).

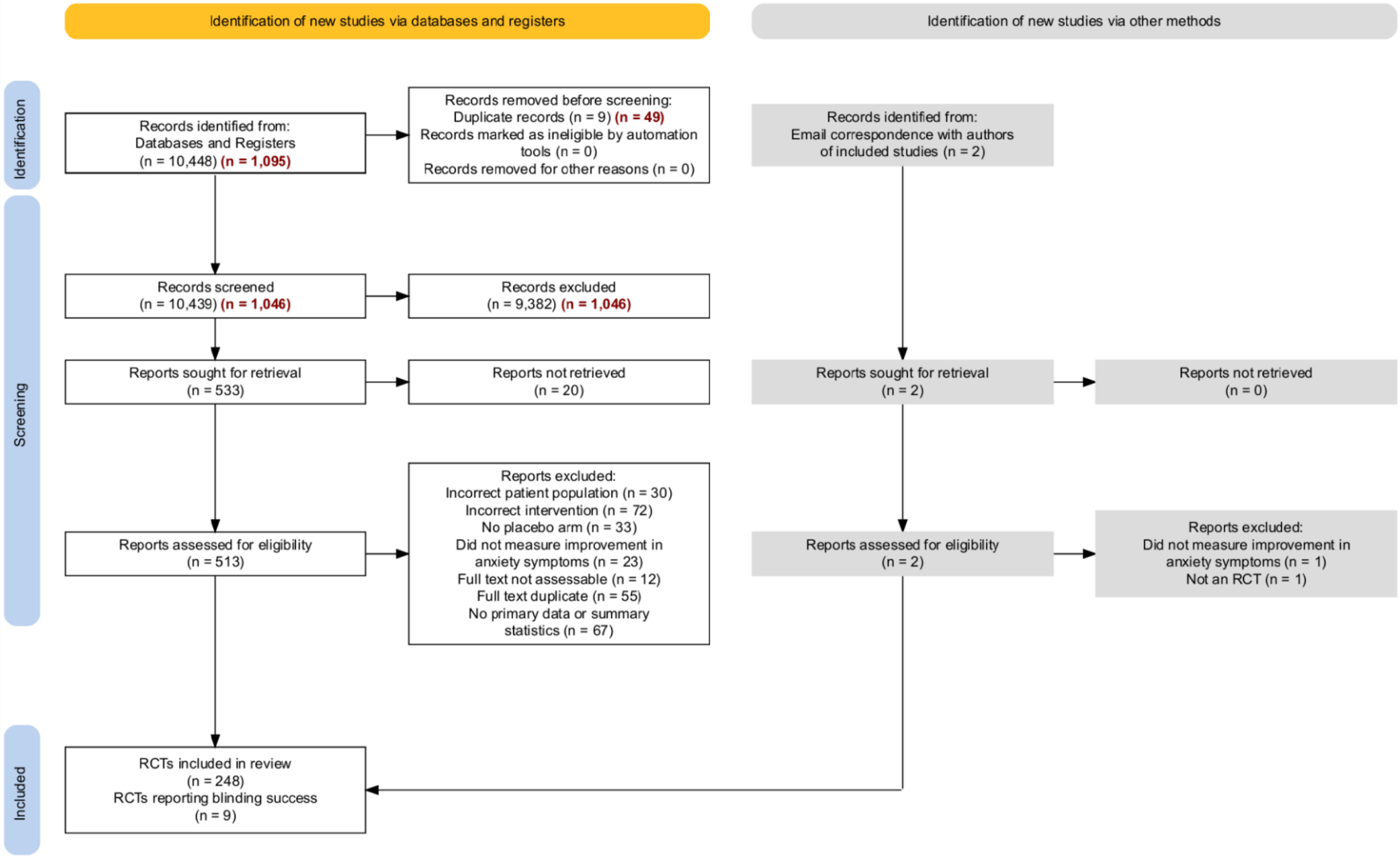

Figure 1 *PRISMA flow diagram. Updated search results in red.*

Table 1 Characteristics of trials reporting an assessment of blinding

Author (Year)	***N***	**Interventions**		**Study Design**	**Study Duration**	**Population**		**Group allocation guesses correct (%)**	**Study Results**
		Experimental	Comparison			Diagnosis	Criteria		
Bakker et al. (1999)	95	Paroxetine, Clomipramine	Placebo	Parallel group	12 weeks	PD± Agoraphobia	DSM-III-R	Patients T 92% P 47% Clinician T 87% P 62%	Paroxetine and clomipramine were both superior to placebo
Başoğlu et al. (1997)	129	Alprazolam	Placebo	Parallel group	16 weeks	PD+ Agoraphobia	DSM-III	Patients T 75% P 68% Clinician T 82% P 78%	Alprazolam not significantly different from placebo
Guastella et al. (2009)	25	Oxytocin	Placebo	Parallel group	5 weeks	SAD	DSM-IV	Patients T 47% P 49% Clinician – no data	No significant difference between oxytocin and placebo on social anxiety symptoms
Hoffart et al. (1993)	17	Clomipramine	Placebo	Crossover	12 weeks	PD+ Agoraphobia	DSM-III-R	No guesses data-kappa values 0.19 for condition A, 0.38 for condition B	Clomipramine was significantly superior to placebo
Kampman et al. (2002)	43	Paroxetine	Placebo	Parallel group	10 weeks	PD± Agoraphobia	DSM-IV	Patients T 76% P 60% Clinician -no data	Paroxetine augmentation of CBT was significantly superior to placebo augmentation of CBT

Author (Year)	*N*	Interventions		Study Design	Study Duration	Population		Group allocation guesses correct (%)	Study Results
		Experimental	Comparison			Diagnosis	Criteria		
Lenze et al. (2009)	177	Escitalopram	Placebo	Parallel group	12 weeks	GAD	DSM-IV	Patients T 55% P 58% Clinician -no data	Escitalopram was superior to placebo
Oosterbaan et al. (2001)	82	Moclobemide	Placebo	Parallel group	15 weeks	SAD	DSM-III-R	Binomial tests for therapists p=0.88; patients p=0.91. Guesses reported as not different from chance.	Moclobemide was not superior to placebo
Van Vliet et al. (1992)	30	Brofaromine	Placebo	Parallel group	12 weeks	SAD	DSM-III-R	Patients – no data Clinicians T 100% P 100%	Brofaromine was significantly superior to placebo
Van Vliet et al. (1993)	29	Brofaromine	Placebo	Parallel group	12 weeks (+12 weeks for some participants)	PD± Agoraphobia	DSM-III-R	Patients – no data Clinicians T 100% P 100%	Brofaromine was significantly superior to placebo

Abbreviations: PD, panic disorder; GAD, generalised anxiety disorder; SAD, social anxiety disorder; T, Treatment; P, placebo

4.1. Quality Assessment

Our assessment of study quality is summarised in Figure 2 and Figure 3. Overall risk of bias was assessed as 'Some concerns' throughout, mainly owing to the lack of clearly identifiable pre-registered analysis plans. Four other studies were additionally rated as 'Some concerns' due to evidence that patients or raters might not have been fully blind to treatment assignment (Van Vliet et al., 1992; van Vliet et al., 1993; Başoğlu et al., 1997; Bakker et al., 1999). Overall, risk of bias was low to moderate.

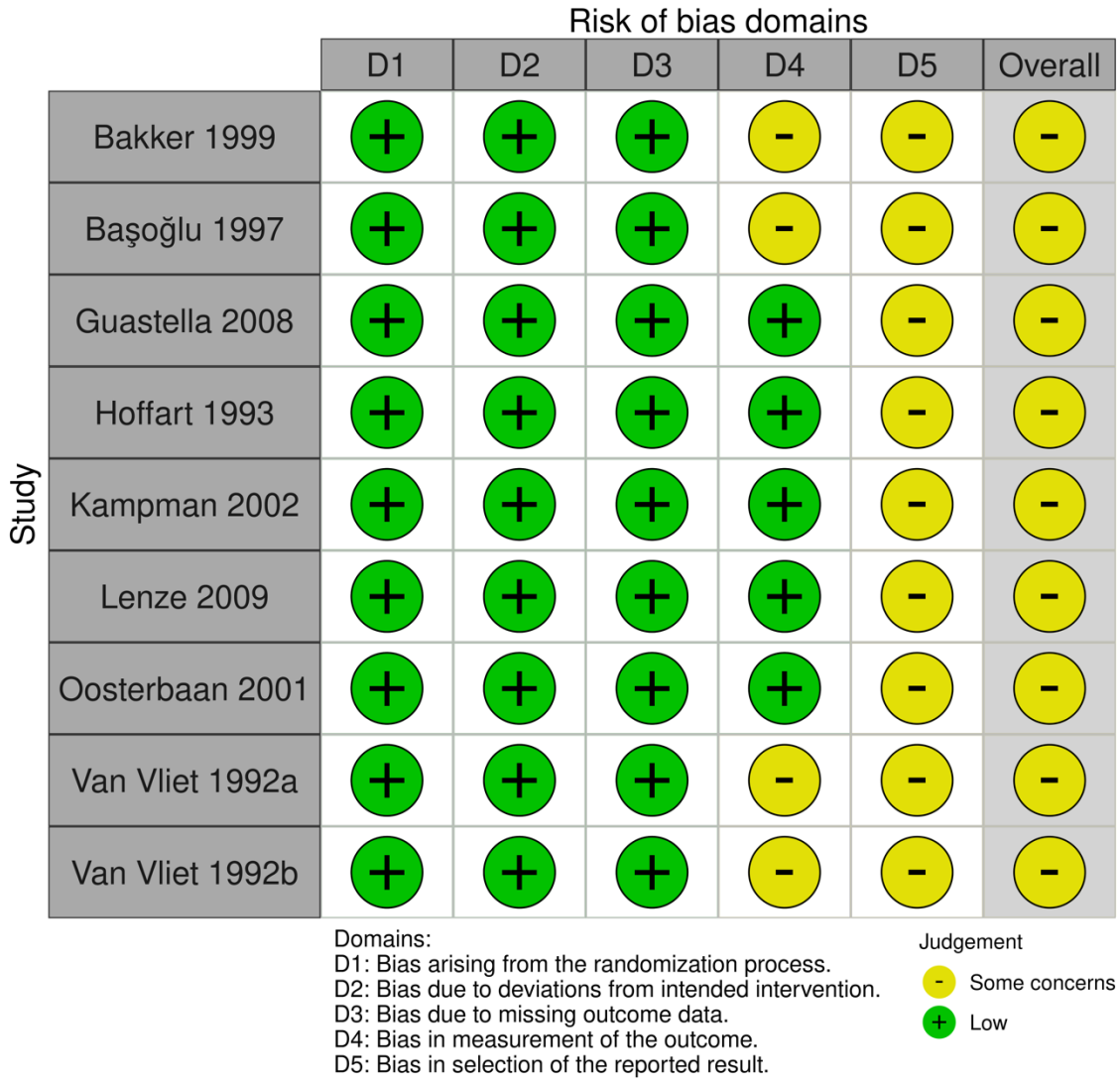

Figure 2 Risk of bias judgment plot generated through Robvis Shinyapp (McGuinness and Higgins, 2020).

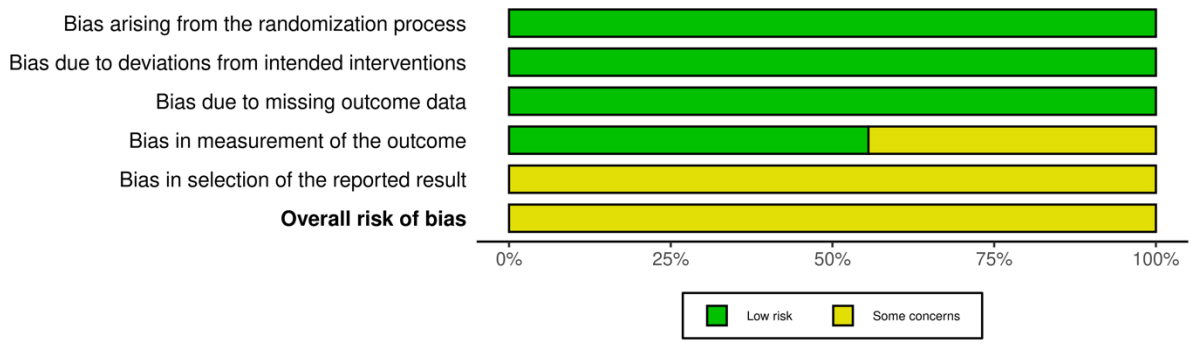

Figure 3 Risk of bias summary plot generated through Robvis Shinyapp (McGuinness and Higgins, 2020).

4.2. Results of blinding integrity assessments

We were able to calculate Bang's Blinding Index (BI) for seven trials (Table 2). We were able to calculate BI for patients in five of these trials (Başoğlu et al., 1997; Bakker et al., 1999; Kampman et al., 2002; Guastella et al., 2009; Lenze et al., 2009). Patients in the placebo arm of RCT's were the most blind group, with a median BI of 0.16. Only one trial (20.00%) reported a BI suggesting unsuccessful blinding (0.357) (Başoğlu et al., 1997). Median blinding index for patients in the medication arms was 0.50, with three trials (60.00%) reporting a blinding index consistent with the blind being broken (BI > 0.20) (Başoğlu et al., 1997; Bakker et al., 1999; Kampman et al., 2002). We were able to calculate BI for clinicians in four trials (Van Vliet et al., 1992; Hoffart et al., 1993; van Vliet et al., 1993; Oosterbaan et al., 2001). The BI's were consistent with the blind being broken in 100% of trials in both placebo and medication arms (median BI = 0.78 and 0.87, respectively).

We could not calculate BI for two trials (Hoffart et al., 1993; Oosterbaan et al., 2001). However, the authors of these trials did report statistics regarding blinding integrity. First, in a crossover study of clomipramine versus placebo in patients with panic disorder, kappa values for patient beliefs about treatment assignment was 0.19 and 0.38, suggesting blinding was intact (Hoffart et al., 1993). Second, in a study assessing moclobemide for social anxiety disorder, blinding was reported to be intact based on the results of binomial tests showing correct guess rates were equal to chance (patients, p = 0.91; clinicians, p=0.88) (Oosterbaan et al., 2001). Altogether, we found that blinding failed in five of the nine trials we included (55.56%).

Table 2 Bang's Blinding Index calculated for patients and clinicians

	Bang's Blinding Index			
	Patients		Clinicians	
	Placebo Arm	Medication Arm	Placebo Arm	Medication Arm
Bakker 1999	-0.0625	0.841	0.429	0.642
Basoglu 1997	0.357	0.500	0.567	0.733
Guastella 2008	-0.0769	0		
Kampman 2002	0.200	0.529		
Lenze 2009	0.160	0.0909		
van Vliet 1992			1.00	1.00
van Vliet 1993			1.00	1.00

4.3. Quantitative synthesis

4.3.1. Anxiolytic efficacy

We carried out a meta-analysis to synthesise the data regarding treatment efficacy in trials reporting an assessment of blinding success. One trial was excluded due to insufficient data (Hoffart et al., 1993). The results of this analysis are summarised in Figure 4. There was a trend towards significant superiority of medication compared with placebo (g = 1.10, 95% CI [0.04, 2.23], p = 0.058), with high heterogeneity between studies (I^2 = 99%, τ^2 = 2.64, $Q_{(7)}$ = 592.04, p < 0.001).

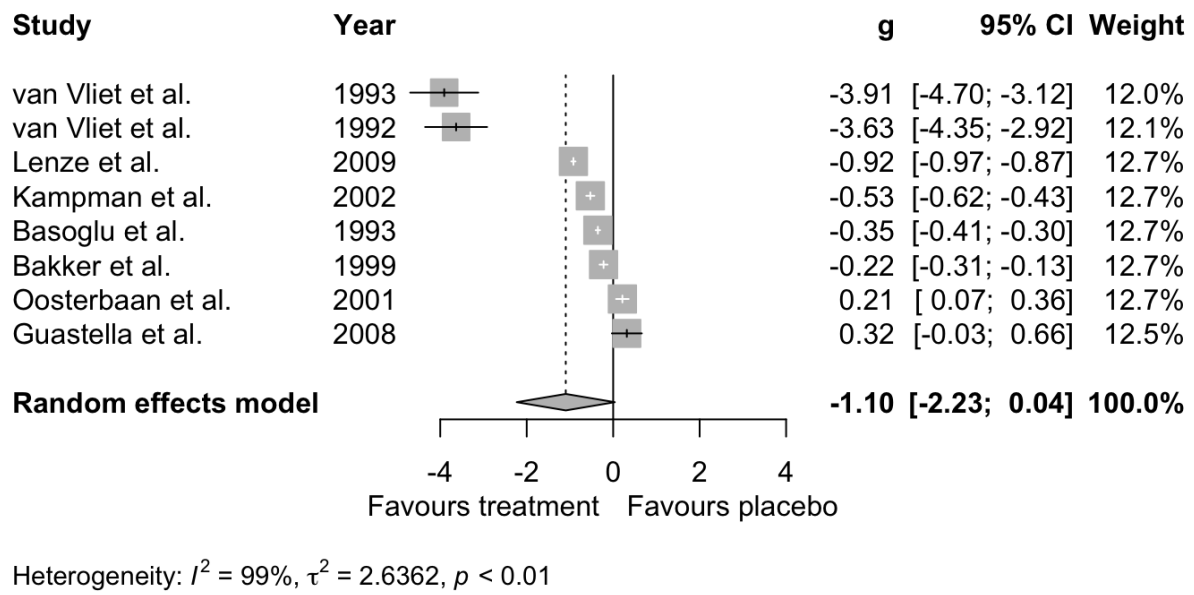

Figure 4 Forest plot of between-group effect sizes comparing active medication and placebo. There was a trend towards significant superiority of medication versus placebo.

4.3.2. Relationship between efficacy and blinding integrity

We next explored the relationship between clinician- and patient-rated Bang's blinding index and the between-group effect size of medication versus placebo through meta-regression (Figure 5). Blinding integrity in clinicians was significantly associated with effect sizes in both active (beta = -10.62, 95% CI [-14.02, -7.22], Z = -6.11, p < 0.001) and placebo arms (beta = -6.63, 95% CI [-8.66, -4.61], Z = -6.42, p < 0.001). These results were likely to have been driven by the two studies by van Vliet and colleagues (1992; 1993), in which clinicians were 100% correct in allocation guesses and effect sizes were very large (g's > 3.50). Regarding blinding integrity in patients, there was a non-significant relationship between increased BI in the placebo arms and increased between-group effect size (beta = -1.32, 95% CI [-3.58, 0.93], Z = -1.15, p = 0.25) that was not present in the active medication arms (beta = -0.06, 95% CI [-1.50, 1.38], Z = -0.08, p = 0.93).

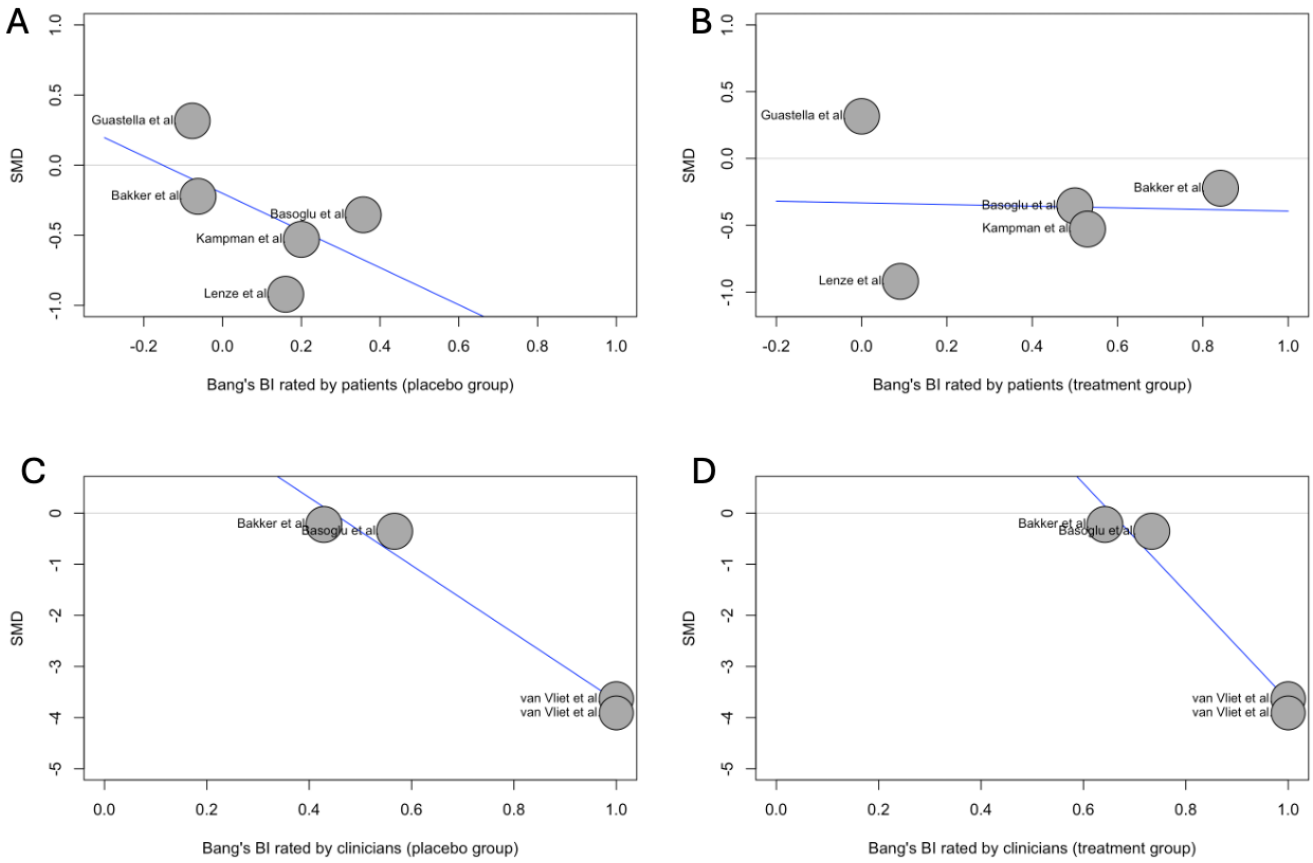

Figure 5 Graphs showing associations between Bang's Blinding Index (BI) and between-group effect size at end of treatment (SMD). Negative effect size favours treatment over placebo. A: Patients in placebo groups; B: Patients in medication groups; C: Clinicians rating placebo groups; D: Clinicians rating treatment groups.

5. Discussion

We conducted a systematic review to identify the number of anxiolytic RCT's conducted since 1980 that carried out and reported an assessment of blinding integrity. We were able to obtain data pertaining to blinding integrity in only 9 of 248 RCT's (3.63%) found by our systematic search. The results of these assessments suggested that blinding was successful in patients in the placebo groups only. Patients in medication arms, and clinicians rating outcomes in both arms, on average guessed group allocation correctly at a rate higher than chance. Regards our secondary aim to explore whether blinding integrity is associated treatment efficacy, we saw an apparently strong relationship between clinicians' blinding index and efficacy, which was likely driven by two studies with very large effect sizes and

clinicians guessing group allocation correctly 100% of the time (Van Vliet et al., 1992; van Vliet et al., 1993). In patients, we saw a non-significant trend where reduced blinding integrity in the placebo groups *only* was associated with *increased* treatment efficacy.

We found that the minority of anxiolytic RCT's since 1980 reported an assessment of blinding integrity. Despite contacting authors for unpublished data, we were only able to obtain blinding integrity assessments for 3.63% of trials. Our findings are in line with previous studies showing that less than 10% of trials report success of blinding in general medicine and in psychiatry (Fergusson et al., 2004; Hróbjartsson et al., 2007; Baethge et al., 2013; Colagiuri et al., 2019; Lin et al., 2022; Scott et al., 2022). Taken as a whole, we found that blinding failed in five of nine trials reporting these data (i.e. more than half; 55.56%). These are similar proportions to those reported in clinical trials across medicine (Fergusson et al., 2004; Boutron et al., 2005; Hróbjartsson et al., 2007; Colagiuri et al., 2019). However, when considering placebo groups only we found that blinding failed in just 20% of anxiolytic trials. This suggests that blinding success is not uniform across intervention groups and raters.

Our secondary aim was to explore whether blinding integrity is associated with treatment efficacy in the context of an RCT. Although we found a significant relationship between BI and treatment effect size among clinicians, this result was likely skewed by two trials in which effect sizes were very large (g's > 3.00) and clinicians guessed treatment allocation 100% correctly (Van Vliet et al., 1992; van Vliet et al., 1993). These trials accounted for 50% of the data entered into these meta-regressions. This finding is challenging to interpret as a result. A previous meta-analysis did not find a relationship between blinding success among clinicians and treatment efficacy in antidepressant trials (Lin et al., 2022) but, as with our study, only four trials were included in their analysis. Nonetheless, there is evidence that non-blind assessors of scale outcomes might over-estimate effect size of treatment by 68% compared with blinded assessors of the same outcome in the same trial (Hróbjartsson et al., 2013). Such observer bias could potentially explain the relationship we found between BI and effect size. However, it remains unknown whether these biases are also applicable to trials that are double-blind by design, but where assessors become unblind during the trial. Assessors might become unblind due to observing symptom improvements or adverse effects of treatment (van der Ende et al., 2023). The former would reflect the efficacy of the

treatment and therefore constitute benign unblinding, while the latter could result in observer bias and constitute malicious unblinding (Szigeti and Heifets, 2024). More data are needed regarding the reasoning behind the guesses clinicians make about treatment allocation, and how this impacts the measurement of treatment efficacy.

We additionally found that blinding was generally intact in patients in the placebo arms of anxiolytic RCT's, but in the treatment arms blinding generally failed. There is evidence in patients with depression that perceived allocation to active treatment is associated with larger improvements in depressive symptoms regardless of actual assignment (Laferton et al., 2018; Lii et al., 2023). Similarly, open-label escitalopram is superior to blinded escitalopram in patients with social anxiety disorder (Faria et al., 2017; Hjorth et al., 2021). Yet, we did not find a relationship between patients' BI and efficacy in the active treatment arms, despite evidence of unblinding in these arms. A meta-analysis of antidepressant trials likewise found no significant relationship between blinding integrity among patients and treatment efficacy (Lin et al., 2022). Further, it appears that there is little difference in estimated treatment effect in patient-reported outcomes in non-blind compared with blinded trials across medicine (Moustgaard et al., 2020). However, in this meta-analysis we have uniquely assessed this relationship separately for placebo and treatment groups. We found a non-significant trend towards reduced blinding integrity in placebo arms being associated with *increased* treatment efficacy. This trend was not present in active treatment arms. This result needs interpreting cautiously as the analysis involved only five datapoints. Nonetheless, this possibly highlights an important issue. There is potential for patient unblinding to cause greater or lesser symptom improvements via response biases or changes in expectations (Schulz and Grimes, 2002; Fergusson et al., 2004; Howick and Hoffmann, 2018; Webster et al., 2021; Szigeti and Heifets, 2024). For example, patients whose symptoms do not improve, or who do not experience adverse effects, might deduce that they have been randomised to placebo, leading to feelings of 'disappointment', or conferring lower expectations of benefit (Schulz and Grimes, 2002; Howick and Hoffmann, 2018). In this scenario, such unblinding might lead to reductions in symptom improvement in the placebo arm, resulting in overestimates of treatment efficacy. Although such phenomena might affect both trial arms, we cautiously propose that our data suggest the

effect might be larger in placebo arms. More data are needed to understand whether this is indeed the case.

In this review, we found signals that suggest patient or assessor unblinding might affect inferences about treatment efficacy. However, these signals result from analyses of limited data. Therefore, we would argue that blinding integrity should be assessed and reported in RCT's for anxiety disorders for two broad reasons: 1) to ensure that potential for response or observer biases are limited as much as possible, and 2) to gather data about the extent of unblinding among patients and assessors, the reasons unblinding occurs, and how much such unblinding affects the estimates of treatment efficacy. Scales have been developed to capture information relevant to the second, such as the 'Guess of Treatment Questionnaire' (Szigeti et al., 2023). We argue that such measures should be taken routinely in RCT's for anxiety disorders.

Our results should be considered in light of possible limitations. First, as with all systematic reviews, we are limited by the quality of the included component studies. However, risk of bias in these studies was rated as low to moderate, suggesting study quality was reasonably high. Second, we were constrained by the limited availability of data regarding blinding integrity. We could only identify six RCT's with this information from 248. Only 29% of authors responded to emails for further information. Therefore, there is potentially data that we have not been able obtain. This highlights the need for systematic collection of data regarding blinding integrity. Finally, we chose to focus solely on the primary outcome of component studies. It is possible that a relationship might have been seen between blinding integrity and treatment efficacy in other outcomes, although this is unlikely given the paucity of data.

6. Conclusion

In summary, this study is the first to our knowledge to assess the frequency and reporting of blinding integrity assessments in RCT's in patients with anxiety disorders. In line with work in other psychiatric disorders, and medicine generally, blinding integrity is rarely reported. Where it is reported, blinding appears to often fail. The potential impacts of this on inferences regarding treatment efficacy remain unclear; but, we found signals that suggest

unblinding of clinician assessors and of patients in placebo arms might be associated with larger treatment effect sizes. We recommend that data regarding blinding integrity, along with the reasons patients and assessors offer for their beliefs regarding group allocation, are systematically collected in RCT's of anxiolytic treatments.

7. Acknowledgements

We would like to thank Dr Gemma Westcott for her support with the literature search and record screening.

8. Conflicts of Interest

Dr Huneke is an NIHR Academic Clinical Lecturer. The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR, NHS or the UK Department of Health and Social Care. All other authors declare no conflicts of interest.

9. Funding

None.

10. Data Availability

Data and code will be made available on the Open Science Framework following peer review and publication.

11. References

Anand R, Norrie J, Bradley JM, McAuley DF, Clarke M (2020) Fool's gold? Why blinded trials are not always best. BMJ 368.

Baethge C, Assall OP, Baldessarini RJ (2013) Systematic review of blinding assessment in randomized controlled trials in schizophrenia and affective disorders 2000-2010. Psychother Psychosom 82:152-160.

Bakker A, Spinhoven P, van Balkom AJ, Matser D, van Dyck R (1999) Double-blindness procedure did not mask giving of medication in panic disorder. J Affect Disord 54:189-192.

Balduzzi S, Rücker G, Schwarzer G (2019) How to perform a meta-analysis with R: a practical tutorial. BMJ Ment Health 22:153-160.

Bang H, Ni L, Davis CE (2004) Assessment of blinding in clinical trials. Controlled clinical trials 25:143-156.

Başoğlu M, Marks I, Livanou M, Swinson R (1997) Double-blindness procedures, rater blindness, and ratings of outcome. Observations from a controlled trial. Arch Gen Psychiatry 54:744-748.

Boutron I, Estellat C, Ravaud P (2005) A review of blinding in randomized controlled trials found results inconsistent and questionable. J Clin Epidemiol 58:1220-1226.

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. The Journal of Pain 20:489-500.

Faria V, Gingnell M, Hoppe JM, Hjorth O, Alaie I, Frick A, Hultberg S, Wahlstedt K, Engman J, Månsson KNT, Carlbring P, Andersson G, Reis M, Larsson E-M, Fredrikson M, Furmark T (2017) Do You Believe It? Verbal Suggestions Influence the Clinical and Neural Effects of Escitalopram in Social Anxiety Disorder: A Randomized Trial. EBioMedicine 24:179-188.

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

Guastella AJ, Howard AL, Dadds MR, Mitchell P, Carson DS (2009) A randomized controlled trial of intranasal oxytocin as an adjunct to exposure therapy for social anxiety disorder. Psychoneuroendocrinology 34:917-923.

Hendriks SM, Spijker J, Licht CMM, Hardeveld F, De Graaf R, Batelaan NM, Penninx BWJH, Beekman ATF (2016) Long-term disability in anxiety disorders. BMC Psychiatry 16.

Hjorth OR, Frick A, Gingnell M, Hoppe JM, Faria V, Hultberg S, Alaie I, Månsson KNT, Rosén J, Reis M, Wahlstedt K, Jonasson M, Lubberink M, Antoni G, Fredrikson M, Furmark T (2021) Expectancy effects on serotonin and dopamine transporters during SSRI treatment of social anxiety disorder: a randomized clinical trial. Translational Psychiatry 11.

Hoffart A, Due-Madsen J, Lande B, Gude T, Bille H, Torgersen S (1993) Clomipramine in the treatment of agoraphobic inpatients resistant to behavioral therapy. J Clin Psychiatry 54:481-487.

Howick J, Hoffmann T (2018) How placebo characteristics can influence estimates of intervention effects in trials. CMAJ 190:E908-E911.

Howick J, Webster RK, Rees JL, Turner R, Macdonald H, Price A, Evers AW, Bishop F, Collins GS, Bokelmann K (2020) TIDieR-Placebo: a guide and checklist for reporting placebo and sham controls. PLoS Med 17:e1003294.

Howick JH (2011) The philosophy of evidence-based medicine: John Wiley & Sons.

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 36:654-663.

Hróbjartsson A, Thomsen ASS, Emanuelsson F, Tendal B, Hilden J, Boutron I, Ravaud P, Brorson S (2013) Observer bias in randomized clinical trials with measurement scale outcomes: a systematic review of trials with both blinded and nonblinded assessors. CMAJ 185:E201-E211.

Huneke NTM (2022) Is superiority to placebo the most appropriate measure of efficacy in trials of novel psychotropic medications? European neuropsychopharmacology: the journal of the European College of Neuropsychopharmacology 62:7-9.

Kampman M, Keijsers GP, Hoogduin CA, Hendriks G-J (2002) A randomized, double-blind, placebo-controlled study of the effects of adjunctive paroxetine in panic disorder patients unsuccessfully treated with cognitive-behavioral therapy alone. Journal of Clinical Psychiatry 63:772-777.

Laferton JA, Vijapura S, Baer L, Mischoulon D (2018) Mechanisms of perceived treatment assignment and subsequent expectancy effects in a double blind placebo controlled RCT of major depression. Frontiers in Psychiatry 9:410069.

Lenze EJ, Rollman BL, Shear MK, Dew MA, Pollock BG, Ciliberti C, Constantino M, Snyder S, Shi P, Spitznagel E, Andreescu C, Butters MA, Reynolds CF, III (2009) Escitalopram for older adults with generalized anxiety disorder: A randomized controlled trial. JAMA: Journal of the American Medical Association 301:295-303.

Lii TR, Smith AE, Flohr JR, Okada RL, Nyongesa CA, Cianfichi LJ, Hack LM, Schatzberg AF, Heifets BD (2023) Randomized trial of ketamine masked by surgical anesthesia in patients with depression. Nature Mental Health 1:876-886.

Lin Y-H, Sahker E, Shinohara K, Horinouchi N, Ito M, Lelliott M, Cipriani A, Tomlinson A, Baethge C, Furukawa TA (2022) Assessment of blinding in randomized controlled trials of antidepressants for depressive disorders 2000–2020: A systematic review and meta-analysis. EClinicalMedicine 50.

McGuinness LA, Higgins JPT (2020) Risk-of-bias VISualization (robvis): An R package and Shiny web app for visualizing risk-of-bias assessments. Research Synthesis Methods n/a.

Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG (2010) CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 340.

Moustgaard H, Clayton GL, Jones HE, Boutron I, Jørgensen L, Laursen DR, Olsen MF, Paludan-Müller A, Ravaud P, Savović J (2020) Impact of blinding on estimated treatment effects in randomised clinical trials: meta-epidemiological study. BMJ 368.

Olesen J, Gustavsson A, Svensson M, Wittchen HU, Jönsson B (2012) The economic cost of brain disorders in Europe. European journal of neurology 19:155-162.

Oosterbaan DB, van Balkom A, Spinhoven P, van Oppen P, van Dyck R (2001) Cognitive therapy versus moclobemide in social phobia: A controlled study. Clinical Psychology & Psychotherapy 8:263-273.

Page MJ et al. (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Systematic Reviews 10.

Rutherford BR, Wall MM, Brown PJ, Choo T-H, Wager TD, Peterson BS, Chung S, Kirsch I, Roose SP (2017) Patient Expectancy as a Mediator of Placebo Effects in Antidepressant Clinical Trials. Am J Psychiatry 174:135-142.

Sackett DL (2004) Turning a blind eye: why we don't test for blindness at the end of our trials. BMJ: British Medical Journal 328:1136.

Schulz KF, Grimes DA (2002) Blinding in randomised trials: hiding who got what. The Lancet 359:696-700.

Schulz KF, Altman DG, Moher D (2010a) CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. Journal of Pharmacology and pharmacotherapeutics 1:100-107.

Schulz KF, Altman DG, Moher D, Fergusson D (2010b) CONSORT 2010 changes and testing blindness in RCTs. The Lancet 375:1144-1146.

Scott AJ, Sharpe L, Colagiuri B (2022) A systematic review and meta-analysis of the success of blinding in antidepressant RCTs. Psychiatry Res 307:114297.

Sterne JAC et al. (2019) RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 366:l4898.

Szigeti B, Heifets B (2024) Expectancy effects in psychedelic trials. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging.

Szigeti B, Nutt D, Carhart-Harris R, Erritzoe D (2023) The difference between ‘placebo group’and ‘placebo control’: a case study in psychedelic microdosing. Sci Rep 13:12107.

van der Ende NA, Roozenbeek B, Broderick JP, Khatri P, Lingsma HF, Dippel DW (2023) Blinding of outcome assessors and its association with outcome in a randomized open-label stroke trial. Int J Stroke 18:562-568.

Van Vliet IM, den Boer JA, Westenberg HG (1992) Psychopharmacological treatment of social phobia: Clinical and biochemical effects of brofaromine, a selective MAO-A inhibitor. European Neuropsychopharmacology 2:21-29.

van Vliet IM, Westenberg HG, Den Boer JA (1993) MAO inhibitors in panic disorder: clinical effects of treatment with brofaromine. A double blind placebo controlled study. Psychopharmacology (Berl) 112:483-489.

Viechtbauer W (2010) Conducting meta-analyses in R with the metafor package. Journal of statistical software 36:1-48.

Webster RK, Bishop F, Collins GS, Evers AW, Hoffmann T, Knottnerus JA, Lamb SE, Macdonald H, Madigan C, Napadow V (2021) Measuring the success of blinding in placebo-controlled trials: Should we be so quick to dismiss it? J Clin Epidemiol 135:176-181.

Wood L, Egger M, Gluud LL, Schulz KF, Jüni P, Altman DG, Gluud C, Martin RM, Wood AJ, Sterne JA (2008) Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ 336:601-605.

World Health Organization (2017) Depression and other common mental disorders: global health estimates. In: World Health Organization.

12. Appendix

12.1. Search Strategy

Search String	Database Source
((panic disorder OR agoraphobia) OR (generalized anxiety disorder) OR (social anxiety disorder) OR (social phobics) OR (social phobia)) AND ((Medication) OR (Selective serotonin reuptake inhibitors) OR (SSRIs) OR (serotonin and norepinephrine reuptake inhibitors) OR (SNRIs) OR (benzodiazepines) OR (gabapentinoids)) AND ((Randomised) OR (placebo))	PubMed, Cochrane Central Register of Controlled Trials, PsycINFO, Embase + Embase classic, OVID MEDLINE, CINAHL, Web of Science, Google Scholar (first 100 pages), clinicaltrials.gov

12.2. Data Extracted

a. Article ID
b. Authors
c. Date
d. Type of publication
e. Diagnosis
f. Diagnostic criteria
g. Age
h. Sex
i. Intervention medications given
j. Control: interventions serving as control
k. Method
l. Study design
m. Duration of study
n. Method of randomisation
o. Nature of blinding
p. Method of blinding
q. Patient and clinician number of correct/ incorrect guesses for treatment and placebo
r. Other blinding data
s. Primary outcome measures of study
t. Secondary outcome measures of study
u. Results
v. Sample size for primary outcome measure
w. Mean and standard deviations for treatment and placebo group primary outcome
x. If no mean/SD for primary outcome, other data if available
y. Statistical tests used for comparison
z. Change in score of anxiety or affective symptoms in treatment group (if not primary outcome)
aa. Author conclusions

12.3. Characteristics of included RCT's and whether blinding integrity information was obtainable

First Named Author	Publication Year	Anxiety Definition Used	Overall Efficacy Outcome	Was blinding assessed?	Was blinding integrity reported?	Email Response?
G. Klerman,	1992	PD: DSM-III	Alprazolam and imipramine superior to placebo (p<0.001)	No	No	One named author, passed away
H. Katschnig	1997	SAD: DSM-IV, ICD-10	moclobemide was superior to placebo (p=0.0017)	No	No	Blinding was not assessed
J. Hartford	2007	GAD: DSM-IV, HADS ≥ 10, CAS ≥ 9	Duloxetine (p= 0.007) and venlafaxine XR (p< 0.001) groups superior to placebo group.	No	No	Email not found through search
J. Adams	1995	GAD: DSM-III-R, HAM-A ≥ 20, AMIS ≥ 2	No difference in mean change between treatment CI-988 and placebo (p= 0.0426) , but there was highly variable placebo response rate between centres.	No	No	No response
K. Alaka	2014	GAD: DSM-IV- TR, CGI-S ≥ 4, CAS ≥ 9, HADS≥ 10	Duloxetine superior to placebo (p<0.001)	No	No	No response
P. Alexander	1993	PD: DSM-III-R	Alprazolam XR superior to placebo group - 4mg (p= 0.042) 6mg (p=0.022)	No	No	No response

N. Aliyev	2008	GAD: DSM-IV, HAM-A ≥ 25	Depakine-chrono superior to the placebo group (p < 0.001).	No	No	No response
C. Allgulander	2004	GAD: DSM-IV, HAM-A ≥ 18	Sertraline superior to placebo (p=0.002)	No	No	Blinding was not assessed
C. Allgulander	2001	GAD: DSM-IV, HRSA ≥ 20	Venlafaxine ER is superior to placebo (p <0.001)	No	No	Blinding was not assessed
C. Allgulander	2007	GAD: DSM-IV, CGI-S ≥ 4, CAS ≥ 9, HADS≥ 10	Duloxetine superior to placebo (p≤0.001)	No	No	Blinding was not assessed
C. Allgulander	2004	SAD: DSM-IV	Venlafaxine ER superior to placebo (p<0.05)	No	No	Blinding was not assessed
C. Allgulander	2008	GAD: DSM-IV TR, HADS≥ 9, RDS >3, CGI-S≥ 4	Duloxetine and venlafaxine superior to placebo (p≤0.001)	No	No	Blinding was not assessed
S. Andersch	1991	PD: DSM-III	Alprazolam, imipramine superior to placebo (p=0.001)	No	No	No response

M. Ansseau	1985	GAD: RDC, HAM-A ≥ 20	Methylclonazepam, lorazepam both superior to placebo (p<0.00001)	No	No	Email not found through search
M. Ansseau	1991	GAD: DSM III-R, HAM-A ≥ 20, CAS ≥ 9	Suriclone, diazepam both superior to placebo (p=0.0001)	No	No	Email not found through search
S. Asakura	2016	SAD: DSM-IV-TR, LSAS-J ≥ 60, CGI-S ≥ 4	Escitalopram efficacious over placebo (p<0.001)	No	No	No response
S. Asakura	2007	SAD: DSM-IV	Fluvoxamine efficacious over placebo (p<0.0197)	No	No	No response
G. Asnis	2001	PD: DSM-III-R	Fluvoxamine efficacious over placebo (p=0.002)	No	No	Data not available
D. Bakish	1996	PD: DSM-III-R	Fluvoxamine and imipramine (equally as effective) both more efficacious over placebo	No	No	No response
A. Bakker	1999	PD: DSM III-R	Antidepressants were superior to placebo (p<0.05)	Yes	Yes	Email not found through search
D. Baldwin	1999	SAD: DSM-IV	Paroxetine effective over placebo (p≤0.001)	No	No	Blinding was not assessed
D. Baldwin	2006	GAD: DSM-IV, HAM-A ≥ 20	Escitalopram more efficacious than paroxetine (p<0.05). Both superior to placebo (p<0.05)	No	No	Blinding was not assessed

J. Ballenger	1988	PD: DSM-III	Alprazolam superior to placebo (p=0.001)	No	No	No response
J. Ballenger	1991	GAD: DSM III-R, HAM-A ≥ 18	Benzodiazepines agonist, abecarnil 3-9mg/day group (rather than higher doses), superior to the placebo	No	No	No response
J. Ballenger	1998	PD: DSM-III-R	Paroxetine effective over placebo (p≤0.001)	No	No	No response
B. Bandelow	2010	GAD: DSM-IV, HAM-A ≥ 20, CGI-S ≥4	Quetiapine XR (p<0.01) and paroxetine (p<0.05) superior to placebo	No	No	No response
D. Barlow	2000	PD: not specified	Imipramine was superior to placebo (p<0.05)	No	No	Blinding was not assessed. 2 studies provided to review for systematic review.
S. Barnett	2002	SAD: DSM-IV, BSPS≥ 20	Olanzapine superior to placebo on BSPS(p=0.02) and SPIN (p=0.01)	No	No	No response
M. Basoglu	1997	PD: Unclear which definition was used	Alprazolam was superior to placebo (p<0.001)	Yes	Yes	N/A
L. Beauclair	1994	PD: DSM-III	Clonazepam superior to placebo (p<0.001)	No	No	No response
J. Benjamin	1995	PD: DSM-III-R	Inositol superior to placebo (p<0.05)	No	No	No success of blinding evaluations

V. Bergnik	2005	PD: DSM-IV	paroxetine was not superior to placebo (p>0.05)	No	No	No success of blinding evaluations
I. Bertin	1989	GAD: DSM III-R	Chlordesmethyldiazepam and lorazepam superior to placebo (p<0.05)	No	No	Email not found through search
L.Bidzan	2012	GAD: HAM-A ≥ 20, MADRS <15	Vortioxetine is superior to placebo (p<0.001)	No	No	No response
H. Bjerrum	1992	GAD: DSM III-R, HAM-A ≥ 16	Flupenthixol was superior to placebo P≤0.01	No	No	Email not found through search
D. black	1996	PD: DSM-III-R	Fluvoxamine was superior to placebo (p<0.05)	No	No	Dataset no longer exists
C. Blanco	2010	SAD: DSM-IV	Phenelzine was superior to placebo (p=0.001)	No	No	No response
S. Blomhoff	2001	SAD: DSM-IV, CGI-SP≥ 4	Sertraline superior to placebo (p<0.001)	No	No	Email not found through search
A. Bose	2008	GAD: DSM-IV, HAM-A ≥ 20	Escitalopram (p=0.09) and venlafaxine (p=0.01) were superior to placebo	No	No	Email not found through search
W. Boyer	1993	GAD: HAM-A	ipsapirone and diazepam were superior to placebo (p<0.05)	No	No	No response
J. Bradwejn	2005	PD: DSM-IV	Venlafaxine ER is superior to placebo (P <0.01)	No	No	No response

O. Brawman-Mintzer	2005	GAD: DSM-IV, HAM-A ≥ 18	Risperidone was superior to placebo (p=0.034)	No	No	No response
O. Brawman-Mintzer	2005	GAD: DSM-IV, HAM-A ≥ 20	Sertraline superior to placebo (p=0.032)	No	No	No response
N. Bresolin	1988	GAD: HAM-A	Ketazolam was superior to placebo (p<0.01)	No	No	Email not found through search
A. Broocks	1998	PD: DSM-III-R, ICD-10	Clomipramine is superior to placebo (p<0.0008)	No	No	No response
J. Careri	1999	GAD: DSM-IV-TR, LSAS≥70, GCI-S ≥4	Vilazodone was superior to placebo group (p=0.04)	No	No	No response
C. Carter	1995	PD: DSM-III-R	Benzodiazepine not significant against placebo (p<0.37), adinazolam-SR is better No than placebo (p<0.04)	No	No	No response
M. Casacchia	1990	GAD: DSM III-R	Etizolam was superior to placebo (p<0.001)	No	No	No response
A. Castillo	1987	GAD: DSM III, HAM-A>17	Alprazolam and clobazam are superior to placebo (p<0.01)	No	No	Email not found through search
D. Ceulemans	1985	GAD: DSM III	Riranserin was superior to palcebo (p<0.05)	No	No	Email not found through search
D. Charney	1989	PD: DSM-II	benzodiazepines were superior to placebo (p<0.05)	No	No	No response

G. Chouinard	1982	GAD and PD: RDC	Alprazolam superior to placebo (P<0.05)	No	No	No response
V. Coric	2010	GAD: DSM-IV TR, HAM-A≥ 18, CGI-S≥ 4	Escitalopram was superior to the placebo (p<0.02). Pexacerfont was not superior to the placebo (P=0.82)	No	No	No response
N. Cutler	1994	GAD: DSM III, HAM-A≥18	Ipsapirone superior to placebo (p<0.05)	No	No	No response
N. Cutler	1993	GAD: DSM III	Ipsapirone and lorazepam were superior to placebo (p<0.05)	No	No	No response
P. Czobor	2010	GAD: DSM-IV TR, HAM-A≥ 20	Ocinaplon was superior to the placebo (p=0.023)	No	No	No response
S. Dager	1992	PD: DSM-III	Alprazolam superior to placebo (p<0.05)	No	No	Email not found through search
A. Dahl	2005	GAD: DSM-IV TR, HAM-A≥ 18	Sertraline was superior to placebo (p<0.001)	No	No	No response
T. Darcis	1995	GAD: DSM III-R	Hydroxyzine was superior to placebo (p<0.001)	No	No	Email not found through search
J. Davidson	2004	SAD: DSM-IV, LSAS ≥60	Fluvoxamine was superior to placebo (p<0.017)	No	No	No response
J. Davidson	1994	PD: DSM-III-R	Adinazolam was superior to placebo (p<0.001)	No	No	No response

J. Davidson	2004	GAD: DSM-IV TR, HAM-A≥ 18, CAS<17	Escitalopram was superior to the placebo (p<0.001)	No	No	No response
J. Davidson	1999	GAD: DSM-IV	Venlafaxine XR superior to placebo (P<0.05)	No	No	No response
J. Davidson	2004	SAD: DSM-IV	Fluoxetine was superior to placebo (p<0.05)	No	No	No response
J. Davidson	1993	SAD: DSM-III-R	Clonazepam superior to placebo (p<0.01)	No	No	No response
F. de Jonghe	1989	GAD: DSM III, HAM-A≥ 18	Sriclone and lorazepam are superior to the placebo (p≤0.05)	No	No	Email not found through search
J. de la Barquera	2008	SAD: DSM-III-R	Clonazepam was superior to placebo (p<0.001)	No	No	No response
J. Deltito	1991	PD: DSM-III, HAM-D<10	Alprazolam and imipramine superior to placebo (P<0.01)	No	No	No response
J. Den Boer	1990	PD: DSM-III-R	Fluvoxamine and ritanserin is superior to placebo (p<0.001)	No	No	No response
J. Den Boer	1992	GAD and SAD: DSM III-R	No significant changes between MSG/ACTH analogs compared to placebo	No	No	No response
E. Dunayevich	2008	GAD: DSM-IV, CAS ≥ 9, HADS≥ 10, RDS<8	LY544344 is superior to placebo (p=0.008)	No	No	Email not found through search

D. Dunner	1986	PD: DSM-III	Alprazolam and diazepam are superior to placebo ($p<0.05$)	No	No	No response
S. Durgam	2016	GAD: DSM-IV TR	Vilazodone was superior to placebo group ($p=0.0236$)	No	No	Email not found through search
R. Enkelmann	1991	GAD: DSM III	Alprazolam and buspirone are superior to placebo ($p<0.05$)	No	No	Email not found through search
D. Feltner	2007	GAD: DSM-IV TR, HAM-A≥ 20	Pregabalin significantly superior to placebo ($p<0.0002$)	No	No	No response
E. Euctr	2006	GAD: DSM-IV TR, HAM-A≥ 20, CGI-S≥ 4	QTP XR was superior to placebo ($p<0.001$)	No	No	Email not found through search
A. Mahableshwarkar	2014	GAD: HAM-A ≥ 20	Vortiozetine was not superior to placebo ($p=0.279$)	No	No	Email not found through search
L. Evans	1986	PD: DSM-III	Zimeldine and imipramine were not superior to placebo $p>0.05$	No	No	Email not found through search
T. Fahlén	1995	SAD: DSM-III-R	Brofaromine was superior to placebo ($p<0.001$)	No	No	No response
T. Fahy	1992	PD: DSM-III-R	Clomipramine was superior to placebo ($p<0.05$)	No	No	Email not found through search

A. Guastella	2008	SAD: DSM-IV, ADIS-IV	Oxytocin (p=0.014) was superior to the placebo (p=0.883)	Yes	No	Email not found through search
D. Feltner	2003	GAD: DSM-IV	Pregabalin superior to placebo (p=0.0013)	No	No	Email not found through search
D. Feltner	2011	SAD: DSM-IV, LSAS ≥50	Pregabalin superior to placebo (p=0.0099)	No	No	Email not found through search
T. Darcis	1996	GAD: DSM-III, HAM-A≥ 20	Hydroxyzine was superior to placebo (p=0.001)	No	No	Email not found through search
M. Ferreri	1994	GAD: DSM III-R	Hydroxyzine was superior to placebo (p=0.001)	No	No	Email not found through search
R. Fontaine	1983	GAD: DSM III-R, HAM-A ≥ 20	Bromazepam (p<0.001) and diazepam (p<0.05) was superior to placebo	No	No	No response
R. Fontaine	1987	GAD: DSM III	Diazepam (P<0.05) is superior to placebo and buspirone (p<0.01)	No	No	No response
R. Fontaine	1986	GAD: DSM III	Bromazepam and lorazepam (p<0.05) is superior to the placebo	No	No	No response
A. Fresquet	2003	GAD: DSM-IV, HAM-A ≥ 18	Lesopitron and lorazepam are superior to the placebo(p=0.044)	No	No	Email not found through search

T. Furmark	2005	SAD: DSM-IV	NKI antagonist and citalopram are superior to placebo (p<0.05)	No	No	No response
A. Gelenberg	1997	GAD: DSM-IV TR	Venlafaxine ER was superior to placebo (p<0.001)	No	No	No response
C. Gommoll	2016	GAD: DSM-IV TR, HAM-A≥ 20	Vilazodone was superior to placebo group (p=0.312)	No	No	Email not found through search
R. Hoehn-saric	1993	PD: DSM-III-R	Fluvoxamine was superior to placebo (p<0.02)	No	No	Email not found through search
A. Hoffart	1993	PD: DSM III-R	Clomipramine was superior to placebo (p<0.05)	Yes - kappa values for patient's guesses	No	No response
D. Ionescu	2013	PD: LSAS	escitalopram was superior to placebo (p<0.05)	No	No	No response
D. Johnston	1988	PD: DSM III-R	Clomipramine was superior to placebo (p<0.002)	No	No	Email not found through search
D. Johnston	1995	PD: DSM-III	Clomipramine was superior to placebo (p<0.05)	No	No	Email not found through search
R. Kahn	1986	PD: DSM-II, HAM-A, CAS	Imipramine is superior to chlordiazepoxide and placebo (p<0.05)	No	No	Email not found through search

M. Kampman	2002	PD: DSM-IV	paroxetine was superior to palcebo (p<0.05)	No	No	Provided patient guesses
S. Kasper	2005	GAD: DSM-IV	Venlafaxine XR was not superior to placebo (P=0.968)	No	No	No response
S. Kasper	2005	SAD: DSM-IV, LSAS ≥70	Escitalopram was superior to placebo (p=0.005)	No	No	No response
R. Katz	1993	GAD: DSM-III-R, HAM-A ≥ 18	Serazepine was superior to placebo (p=0.012)	No	No	Email not found through search
D. Katzelnick	1995	SAD: DSM-III-R	Sertraline was superior to placebo (p=0.001)	No	No	No response
H. Koponen	2007	GAD: DSM-IV	Duloxetine and venlafaxine superior to placebo (p≤0.001)	No	No	No response
D. Koszycki	2011	PD: DSM-IV	Sertraline was superior to placebo (p=0.0180)	No	No	Blinding was not assessed
P. Kragh-Sørensen	1990	GAD: DSM-III	Bromazepam was superior to placebo (p<0.05)	No	No	Email not found through search
M. Kramer	1995	PD: DSM III-R	L-365,260 was not superior to placebo (p>0.05)	No	No	Email not found through search
G. Kronenberg	2005	PD: DSM-IV, PAS ≥ 18, CGI≥ 4	CCK-4 was not superior to placebo (p>0.05)	No	No	No response

M. Krüger	1999	PD: DSM-III-R	Moclobemide and Clomipramine were superior to placebo ($p<0.05$)	No	No	No response
G. Laakmann	1997	GAD: DSM-III	Buspirone and lorazepam were both superior to the placebo ($p\leq0.05$)	No	No	Email not found through search
M. Lader	1998	GAD: DSM-IV	Hydroxyzine was superior to placebo ($p<0.05$). Buspirone was not superior to placebo.	No	No	Email not found through search
M. Lader	2004	SAD: DSM-IV, LSAS ≥70, SDS≥5	Escitalopram was superior to placebo ($p<0.05$)	No	No	Email not found through search
Y. Lapierre	1997	PD: DSM III-R	Paroxetine and clomipramine were superior to placebo ($p<0.05$)	No	No	Email not found through search
Y. Lecrubier	1993	GAD: DSM III, CAS ≥ 6	Tropisetron was superior to placebo ($p<0.05$)	No	No	Email not found through search
E. Leinonen	2000	PD: DSM-III-R	Citalopram was superior to placebo ($p<0.05$)	No	No	No response
A. Lenox-Smith	2013	GAD: DSM-IV, HAM-A≥ 20	Venlafaxine XL was superior to placebo ($p=0.05$)	No	No	Email not found through search
E. Lenze	2009	GAD: DSM-IV, HAM-A≥ 18	Escitalopram was superior to placebo ($p=0.03$)	Yes	Patient guesses only	Email not found through search

C. León	1990	PD: DSM-III	Alprazolam and imipramine were superior to the placebo ($p<0.05$)	No	No	Email not found through search
U. Lepola	2004	SAD: DSM-IV, HAM-A≥ 15	Paroxetine was superior to placebo ($p<0.01$)	No	No	No response
U. Lepola	1998	PD: DSM III-R, MADRS≥ 22	Citalopram was superior to placebo ($p<0.04$)	No	No	Email not found through search
M. Liebowitz	2009	PD: DSM-IV, CGI≥ 4	Venlafaxine ER superior to placebo ($p=0.006$)	No	No	No response
M. Liebowitz	2003	SAD: DSM-IV, LSAS ≥68	Sertraline superior to placebo ($p<0.001$)	No	No	No response
M. Liebowitz	2005	SAD: DSM-IV	Venlafaxine and paroxetine were superior to placebo ($p<0.001$)	No	No	No response
M. Liebowitz	2016	SAD, GAD: DSM-IV, LSAS≥60, CGI-S≥ 4	PH94B was superior to placebo ($p=0.2$)	No	No	No response
M. Liebowitz	1992	SAD: DSM-IV, LSAS≥50, CGI-S≥ 4	Venlafaxine ER was superior to placebo ($p<0.01$)	No	No	No response
M. Liebowitz	2002	SAD: DSM-III	Phenelzine was superior to placebo ($p=0.02$)	No	No	No response
M. Liebowitz	2002	SAD: DSM-IV	Paroxetine was superior to placebo ($p<0.01$)	No	No	No response

M. Linden	1997	GAD: DSM III-R, HAM-A ≥ 18	DN-2327 was superior to placebo (p<0.05)	No	No	No response
P. Llorca	2002	GAD: DSM- IV, HAM-A≥ 20	Hydroxyzine was superior to placebo (p=0.019)	No	No	No response
P. Londborg	1998	PD: DSM-III	Sertraline was superior to placebo (p<0.05)	No	No	Email not found through search
H. Lôo	1991	GAD: HAM-A	Lorazepam was superior to placebo (p=0.008)	No	No	No response
M. Lott	1997	SAD: DSM-III	Brofaromine was superior to the placebo (p<0.001)	No	No	Email not found through search
B. Lydiard	1997	GAD: DSM-III	Abecarnil and alprazolam were superior to placebo (p<0.05)	No	No	No response
B. Lydiard	1992	PD: DSM-III	Alprazolam was superior to placebo (p<0.05)	No	No	No response
B. Lydiard	1993	PD: DSM III-R	Desipramine was not superior to placebo overall (p<0.09)	No	No	No response
A. Mahableshwarkar	2014	GAD: DSM-IV, HAM-A ≥ 20	Duloxetine and vortioxetine were superior to placebo (p=0.036, p<0.05)	No	No	Email not found through search
A. Mahableshwarkar	2014	GAD: DSM-IV, HAM-A ≥ 20	Vortioxetine was not superior to placebo (p=0.279)	No	No	Email not found through search

M. Mavissakalian	1985	PD: DSM-III	Imipramine was superior to placebo (p<0.05)	No	No	No response
M. Mavissakalian	1995	PD: DSM-III	Imipramine was superior to placebo (p<0.05)	No	No	No response
D. McLeod	1992	GAD: DSM III-R	Alprazolam and imipramine were superior to placebo (p<0.001)	No	No	Email not found through search
J. Mendels	1986	GAD: DSM III, HAM-A≥ 20	Trifluoperazine was superior to placebo (p<0.001)	No	No	Email not found through search
C. Merideth	2012	GAD: DSM-IV-TR, HAM-A≥ 20	Quetiapine XR (p<0.001) , escitalopram (p<0.05) were superior to placebo	No	No	Email not found through search
I. Mezhebovsky	2013	GAD: DSM-IV	Quetiapine XR was superior to placebo (p<0.001)	No	No	No response
D. Michelson	2001	PD: DSM-IV TR, CGI-S ≥ 12	Fluoxetine was superior to placebo (p<0.05)	No	No	Email not found through search
D. Michelson	2012	GAD: DSM IV	L-758274 was not superior to placebo (p=0.359)	No	No	Email not found through search
D. Michelson	1998	PD: DSM-III	Fluoxetine was superior to placebo (p=0.006)	No	No	Email not found through search
K. Modigh	1992	PD: DSM III	Clomipramine was superior to imipramine (p<0.001) and placebo	No	No	Did not assess blinding

H. Möller	2001	GAD: ICD-10, HAM-A≥ 17	Alprazolam was superior to placebo (p<0.05)	No		No access to data
S. Montgomery	2006	GAD: DSM-IV, HAM-A≥ 20, CAS<8	Pregabalin (p=0.008) and velafaxine (p=0.03) were superior to placebo	No	No	No response
G. Moroz	1999	PD: DSM III-R	Clonazepam was superior to placebo (p<0.05)	No	No	No response
M. Muehlbacher	2005	SAD: DSM-IV	Mirtazapine was superior toplacebo (p<0.001)	No	No	No access to data
Murphy, S. M.	1989	PD: DSM III	Alprazolam was superior to placebo (p<0.05), but not propranolol	No	No	No response
N. Nair	1996	PD: DSM III-R	Imipramine was superior to placebo (p<0.05)	No	No	No response
H. Naukkarinen	2005	GAD: DSM-IV, HAM-A ≥ 18, CGI-S≥4	Deramicilane was superior to placebo (p=0.024)	No	No	Did not assess blinding
H. Nicolini	2008	GAD: DSM-IV , CAS≥ 9, CGI-S≥ 4	Duloxetine and venlafaxine superior to placebo (p≤0.01)	No	No	No response
I. Nimatoudis	2004	GAD: DSM-IV, HAM-A ≥ 18	Venlafaxine ER was superior to placebo (p=0.0006)	No	No	No response
H. Nordahl	2016	SAD: DSM-IV	paroxetine was not superior to placebo (p>0.05)	No	No	No response

R. Noyes	1996	PD: DSM III	Diazepam and alprazolam were superior to placebo (P<0.05)	No	No	Email not found through search
R. Noyes	1988	PD: DSM III	Diazepam and alprazolam were superior to placebo (P<0.05)	No	No	Email not found through search
R. Noyes	1997	SAD: DSM-III-R	Moclobemide was superior to placebo (p<0.05)	No	No	Email not found through search
S. Oehrberg	1995	PD: DSM III-R	Paroxetine was superior to placebo (p=0.001)	No	No	Email not found through search
D. Olajide	1987	GAD: DSM-III, HAM-A≥ 18	Diazepam was superior to buspirone and placebo (p<0.01)	No	No	No response
D. Oosterbaan	2001	PD: DSM III-R	Moclobemide was not superior to placebo (p>0.05)	Yes	No	No response
A. Pande	2003	GAD: DSM IV	Pregabalin superior to placebo (p<0.05)	No	No	No response
A. Pande	2004	SAD: DSM-IV	Gabapentin was superior to placebo (p=0.05)	No	No	No response
A. Pande	2004	SAD: DSM-IV	Pregabalin superior to placebo (p=0.024)	No	No	No response
A. Pande	1999	PD: DSM III-R	CI-988 was not superior to placebo	No	No	No response

A. Pande	2000	PD: DSM-IV	Gabapentin was not superior to placebo (p=0.606)	No	No	No response
L. Pangalila-Ratu	1988	GAD: DSM-III	Ritanserin was superior to placebo (p<0.05)	No	No	No response
J. Pecknold	1994	PD: DSM III-R	Alprazolam and Alprazolam XR were superior to placebo (p<0.01)	No	No	No response
J. Pecknold	1988	GAD: DSM III, HAM-A≥ 18	Buspirone and lorezepam were both superior to the placebo (p≤0.05)	No	No	Blinding was not assessed
J. Pecknold	1986	PD: DSM-III	Alprazolam was superior to placebo (p<0.05)	No	No	Email not found through search
J. Pecknold	1988	PD: DSM III	Alprazolam was superior to placebo (p<0.02)	No	No	No response
R. Pohl	1989	PD: DSM III	Buspirone was superior to the placebo (p<0.02)	No	No	Email not found through search
R. Pohl	2005	GAD: HAM-A ≥ 20	Pregabalin superior to placebo (p<0.05)	No	No	No response
R. Pohl	1998	PD: DSM-III-R	Sertraline was superior to placebo (p=0.03)	No	No	Email not found through search
M. Pollack	2007	PD: DSM-IV	Venlafaxine ER was superior to placebo (p<0.01)	No	No	No response

M. Pollack	2007	PD: DSM-IV	Venlafaxine ER is superior to placebo (p <0.01)	No	No	No response
M. Pollack	1998	PD: DSM III-R	Sertraline is superior to placebo (p<0.01)	No	No	No response
M. Pollack	2005	GAD: DSM-IV	Tiagabine was superior to placebo (p<0.05)	No	No	No response
M. Pollack	1997	GAD: DSM III-R, HAM-A ≥ 20	Abecarnil effective over placebo (p<0.05)	No	No	No response
M. Pollack	1996	PD: DSM III-R	Venlafaxine XR superior to placebo (p<0.05)	No	No	No response
M. Pollack	2001	GAD: DSM-IV, HAM-A≥ 20	paroxetine was superior to placebo (p<0.05)	No	No	No response
G. Post	1991	GAD: DSM III-R	Escitalopram was superior to placebo (p<0.01)	No	No	Email not found through search
T. Pourmotabbed	1996	PD: DSM III-R	Diazepam was superior to placebo (P<0.05)	No	No	No response
K. Power	1990	GAD: DSM III	Diazepam was superior to placebo (P<0.05)	No	No	Email not found through search
J. Prasko	2006	SAD: ICD-10	Moclobemide was superior to placebo (p<0.05)	No	No	No response

K. Rickels	1970	SAD: DSM-IV	Atomoxetine was not superior to placebo (p=0.91)	No	No	Email not found through search
K. Rickels	1998	GAD: DSM III, HAM-A≥ 20	Diazepam was superior to placebo (P<0.01), abecarnil was not	No	No	Email not found through search
K. Rickels	1993	GAD: DSM III, HAM-A≥ 18	Imipramine and diazepam superior to placebo (p<0.01), trazodone only better at trend level (p<0.1).	No	No	Email not found through search
K. Rickels	1985	GAD: DSM-III, HAM-A≥ 18	diazepam was superior to placebo (p<0.05)	No	No	Email not found through search
K. Rickels	2004	SAD: DSM- IV, CGI-S≥ 4	Venlafaxine ER superior to placebo (p<0.05)	No	No	Email not found through search
K. Rickels	2008	GAD: DSM-IV, HAM-A≥ 20	PRX-00023 is superior to placebo (p=0.0094)	No	No	Email not found through search
K. Rickels	2005	GAD: DSM-IV, HAM-A ≥ 20	Pregabalin was superior to placebo (p<0.02)	No	No	Email not found through search
K. Rickels	2000	GAD: DSM-IV	venlafaxine was superior to placebo (p<0.05)	No	No	Email not found through search
K. Rickels	1997	GAD: DSM III	Gepirone and diazepam were superior to placebo (p<0.05)	No	No	Email not found through search
K. Rickels	2003	GAD: DSM-IV, HAM-A≥ 20	Paroxetine was superior to placebo (p<0.05)	No	No	Email not found through search

P. Rolland	2000	GAD: DSM IV	Venlafaxine XR was superior to placebo ($p<0.05$)	No	No	No access to data
J. Rosenbaum	1991	PD: DSM III-R	Clonazepam was superior to placebo ($p<0.05$)	No	No	Email not found through search
C. Ross	1987	GAD: DSM-III	Buspirone and diazepam were both not superior to the placebo	No	No	No response
A. Rothschild	2012	GAD: HAM-A ≥ 20	vortioxetine was not superior to placebo ($p=0.518$)	No	No	No response
F. Savoldi	1990	PD: DSM-III	Etizolam was superior to placebo ($p<0.05$)	No	No	Email not found through search
E. Scarpini	1988	GAD: HAM-A ≥ 8	Ketazolam was superior to placebo ($p<0.001$)	No	No	Email not found through search
M. Schmidt	2021	SAD: DSM-V, LSAS≥ 70	JNJ-42165279 was superior to placebo ($p=0.04$)	No	No	Blinding was not assessed
S. Schutters	2010	SAD: DSM-IV	Mirtazapine was not superior to placebo	No	No	Email not found through search
E. Schweizer	1992	PD: DSM-III-R	Midazolam was superior to placebo ($p<0.05$)	No	No	No response
E. Schweizer	1988	PD: DSM III	Alprazolam was superior to placebo ($p<0.05$)	No	No	Email not found through search

E. Schweizer	1988	PD: HARS	Buspirone and Clorazepate were superior to placebo ($p<0.02$)	No	No	No access to data
E. Schweizer	1990	GAD: DSM III, HAM-A≥ 18	Enciprazine was superior to placebo ($p<0.05$)	No	No	No response
D. Sharp	1993	PD: DSM III	Alprazolam and imipramine superior to placebo ($p<0.05$)	No	No	Email not found through search
D. Sharp	1996	PD: DSM III	Fluoxetine was superior to placebo ($p<0.05$)	No	No	Email not found through search
D. Sheehan	1993	PD: DSM-III	Alprazolam was superior to placebo ($p<0.05$) , buspirone was not	No	No	No response
D. Sheehan	2013	GAD: DSM-IV, HAM-A≥ 20	Quetiapine was superior to placebo ($p<0.001$)	No	No	No response
A. Simen	2017	GAD: DSM-IV-TR	PF-06372865 was not superior to placebo	No	No	No access to data
S. Stahl	2003	PD: DSM-IV	Escitalopram was superior to placebo ($p=0.04$)	No	No	Blinding was not assessed
D. Stein	2017	GAD: DSM-IV-TR	Agomelatine was superior to placebo ($p<0.0001$)	No	No	Email request referred to other staff member, and no response received
D. Stein	2014	GAD: DSM-IV-TR	Agomelatine was superior to placebo ($p<0.0001$)	No	No	Email request referred to other staff member, and

						no response received
D. Stein	2008	GAD: DSM-IV-TR	Agomelatine was superior to placebo (p=0.04)	No	No	Email request referred to other staff member, and no response received
D. Stein	1999	SAD: DSM-IV	Paroxetine was not superior to placebo	No	No	Email request referred to other staff member, and no response received
D. Stein	2002	SAD: DSM-IV	Moclobemide was superior to placebo (p<0.05)	No	No	Email request referred to other staff member, and no response received
D. Stein	2002	SAD: DSM-IV	Paroxetine was superior to placebo (p<0.01)	No	No	Email request referred to other staff member, and no response received
D. Stein	2002	SAD: DSM-IV	Fluvoxamine was superior to placebo (p=0.028)	No	No	Email request referred to other staff member, and no response received

M. Stein	2014	GAD: DSM-IV TR	Alprazolam was superior to placebo ($p<0.05$)	No	No	Did not assess blinding
M. Stein	1999	SAD: DSM-IV	Fluvoxamine was superior to placebo ($p<0.05$)	No	No	Did not assess blinding
M. Stein	1998	SAD: DSM-IV	Paroxetine was superior to placebo ($p<0.05$)	No	No	Did not assess blinding
M. Stein	2010	SAD: DSM-IV	Venlafaxine ER was superior to placebo ($p<0.001$)	No	No	Did not assess blinding
M. Stein	2008	SAD: DSM-IV	Levetiracetam was not superior to placebo	No	No	Did not assess blinding
J. Tauscher	2010	SAD: DSM-IV	Paroxetine was superior to placebo ($p<0.05$)	No	No	No response
C. Taylor	1990	SAD: DSM-IV-TR, CGI-S ≥ 4	Alprazolam and imipramine were superior to placebo ($p<0.05$)	No	No	No response
G. Tesar	1991	PD: DSM-III	Clonazepam and Alprazolam were superior to placebo ($p<0.05$)	No	No	Email not found through search
T. Uhde	1989	PD: DSM-III	Clonidine is superior to placebo ($p<0.02$)	No	No	No response
E. Uhlenhuth	1989	PD: DSM III	Alprazolam and imipramine superior to placebo ($p<0.05$)	No	No	Email not found through search
S. Vaishnavi	2007	SAD: DSM-IV	Quetiapine was not any superior to placebo	No	No	No response

A. Valença	2000	PD: DSM-IV	Clonazepam superior to placebo (p=0.079)	No	No	No response
M. Van Ameringen	2007	SAD: DSM-IV	Nefazodone was superior to placebo (p<0.05)	No	No	Did not assess blinding
A. van Balkom	1996	SAD: DSM-IV	Sertraline is superior to placebo (p<0.001)	No	No	Email not found through search
I. Van Vliet	1992	SAD: DSM-III-R	Brofaromine was superior to placebo (p<0.05)	No	No	Provided more blinding information: 100% of clinicians guessed 100% of treatments correctly
I. Van Vliet	1997	SAD: DSM-III-R	Buspirone is not superior to placebo - statistical difference from baseline but not compared to each other (p<0.001)	No	No	Blinding was not assessed
I. Van Vliet	1994	SAD: DSM-III-R	Fluvoxamine was superior to placebo (p<0.001)	No	No	Blinding was not assessed
I. Van Vliet	1993	PD: DSM-III-R	Bromazepam was superior to placebo (p<0.01)	No	No	Provided more blinding information: 100% of clinicians guessed 100% of treatments correctly
M. Versiani	2002	PD: DSM-III-R	Reboxetine was superior to the placebo (p<0.05)	No	No	Email not found through search

M. Versiani	2001	SAD: DSM-III-R, CGI≤4	Bromazepam was superior to placebo (p<0.05)	No	No	Blinding was not assessed
M. Versiani	1992	SAD: LSAS	Moclobemide and phenelzine were superior to placebo (p<0.0001)	No	No	Email not found through search
J. Walker	2000	PD: DSM-III-R	Citalpram and clomipramine were superior to placebo (p<0.05)	No	No	Email not found through search
H. Westernberg	1994	SAD: DSM-IV, LSAS≥60	Fluvoxamine was superior to placebo (p<0.02)	No	No	Email not found through search
C. Wilcox	1991	GAD: DSM III-R , HAM-A≥ 20	Adinazolam was superior to placebo (p<0.0178)	No	No	Authors no longer have access to the information
W. Wu-yuan	2011	GAD: DSM-IV	Duloxetine was superior to placebo (p=0.022)	No	No	No response
W. Zhang	2005	SAD: DSM-IV, BSPS≥ 20	Levetiracetam was superior to placebo (p<0.05)	No	No	Email not found through search

Abbreviations: GAD, generalised anxiety disorder; SAD, social anxiety disorder; PD, panic disorde; DSM diagnostic and statistical manual of mental disorders; LSAS, Liebowitz social anxiety scale; CGI, clinical global impression scale; HAM-A/HARS, Hamilton anxiety rating scale; HADS, hospital anxiety and depression scale; RDS, rejection sensitive dysphoria; BSPS, brief social phobia scale; MADRS, Montgomery-asberg depression rating scale; ICD, international classification of disease; AIMS, Abnormal Involuntary Movement Scale; XR/ ER, extended release