

## Review article

# Efficacy of cognitive–behavioural therapy and other psychological treatments for adult depression: meta-analytic study of publication bias

Pim Cuijpers, Filip Smit, Ernst Bohlmeijer, Steven D. Hollon and Gerhard Andersson

**Background**

It is not clear whether the effects of cognitive–behavioural therapy and other psychotherapies have been overestimated because of publication bias.

**Aims**

To examine indicators of publication bias in randomised controlled trials of psychotherapy for adult depression.

**Method**

We examined effect sizes of 117 trials with 175 comparisons between psychotherapy and control conditions. As indicators of publication bias we examined funnel plots, calculated adjusted effect sizes after publication had been taken into account using Duval & Tweedie's procedure, and tested the

symmetry of the funnel plots using the Begg & Mazumdar rank correlation test and Egger's test.

**Results**

The mean effect size was 0.67, which was reduced after adjustment for publication bias to 0.42 (51 imputed studies). Both Begg & Mazumdar's test and Egger's test were highly significant ( $P < 0.001$ ).

**Conclusions**

The effects of psychotherapy for adult depression seem to be overestimated considerably because of publication bias.

**Declaration of interest**

None.

Most meta-analytic studies of cognitive–behavioural and other psychotherapies for adult depression have found that these therapies have moderate to large effects on depression compared with control conditions,<sup>1–5</sup> effects that are comparable to those of antidepressive medication.<sup>6</sup> It may be possible, however, that these effects are overestimated because of publication bias – the tendency for increased publication rates among studies that show a statistically significant effect of treatment.<sup>7,8</sup> Publication bias can be considered as one of the major drawbacks of meta-analytic studies and a threat to their validity. Evidence of such bias has been found in many intervention fields, including that of depression treatment. In two recent meta-analytic reviews of studies examining the effects of antidepressive medication,<sup>9,10</sup> it was found that considerably more positive than negative trials were reported, which resulted in a considerable overestimation of the effect sizes of drugs. Although many meta-analytic studies have examined the effects of cognitive–behavioural and other psychotherapies for depression,<sup>11</sup> publication bias has not been examined thoroughly in these studies. Most comprehensive meta-analyses examining the effects of psychotherapies have not examined publication bias at all.<sup>12,13</sup> One meta-analysis did examine publication bias,<sup>2</sup> but did not test whether the effects of published and unpublished studies differed significantly from each other, nor did it use advanced techniques to formally test whether significant publication bias was present. That study also did not calculate effect sizes that were adjusted for publication bias by imputing missing studies, and assessed the mean effect sizes of psychotherapies after adjustment for publication bias; furthermore, the meta-analysis included only a limited number of the published studies. However, the study found some indications that publication bias might be present in research on cognitive–behavioural and other psychotherapies for adult depression. We decided, therefore, to conduct a new meta-analysis of studies examining the effects of psychotherapies for adult depression and focus on the question whether publication bias may have resulted in an overestimation of the mean effect size.

**Method****Identification and selection of studies**

We used a database of 1036 papers on the psychological treatment of depression, which includes studies on combined treatments and comparisons with pharmacotherapies. This database has been described in detail elsewhere,<sup>14</sup> and has been used in a series of earlier meta-analyses ([www.evidencebasedpsychotherapies.org](http://www.evidencebasedpsychotherapies.org)). It was developed through a comprehensive literature search (from 1966 to January 2009) in which we examined 9011 abstracts in PubMed (1629 abstracts), PsycINFO ( $n = 2439$ ), Embase ( $n = 2606$ ) and the Cochrane Central Register of Controlled Trials ( $n = 2337$ ). These abstracts were identified by combining terms indicative of psychological treatment and depression (both MeSH terms and text words). For this database we also collected the primary studies from 42 meta-analyses of psychological treatment for depression ([www.evidencebasedpsychotherapies.org](http://www.evidencebasedpsychotherapies.org)). For the study reported here we examined the full texts of these 1036 papers. The abstracts from the electronic bibliographic databases were screened by the first author (P.C.) and the 1036 retrieved reports were examined independently by two reviewers for possible inclusion. When the two reviewers disagreed, they discussed the differences with the third reviewer until agreement was reached.

We included published studies in which the effects of a psychological treatment on adults with a diagnosed depressive disorder or scoring above a cut-off point on a depression measurement instrument were compared with a control condition in a randomised controlled trial. Depressive disorders could be diagnosed with one of the versions of the DSM, the ICD, Research Diagnostic Criteria or another diagnostic system. 'Psychological treatments' were defined as interventions in which verbal communication between a therapist and a client was the core element, or in which a psychological treatment was written down in book format (bibliotherapy) which the client worked through more or less independently, but with some kind of personal support from a therapist (by telephone, email or otherwise).

In earlier meta-analyses of psychotherapies for adult depression, few indications were found that different types of psychotherapy have differential effects on depression. Few indications for significant differences are found in multivariate meta-regression analyses,<sup>1</sup> as well as in meta-analytic research in which direct comparisons of different types of psychotherapy are examined.<sup>15,16</sup> However, we decided to examine the effects of the full sample of psychotherapies, as well as the sample of studies examining cognitive-behavioural therapies. We defined cognitive-behavioural therapy as a psychological treatment in which the therapist focuses on the impact a patient's present dysfunctional thoughts have on current behaviour and future functioning.<sup>16</sup> The therapy is aimed at evaluating, challenging and modifying the patient's dysfunctional beliefs (cognitive restructuring). The definitions of other psychotherapies we distinguished have been given elsewhere.<sup>16</sup>

We excluded studies on children and adolescents (below 18 years of age). Studies in which the psychological intervention could not be distinguished from other elements of the intervention were also excluded (managed care interventions and disease management programmes), as were studies in which a standardised effect size could not be calculated (mostly because no test was performed in which the difference between experimental and control groups was examined), studies in which patients were not randomly assigned to conditions, and studies of in-patients. We also excluded unpublished dissertations, studies aimed at maintenance treatments and relapse prevention, and studies that also included participants with anxiety disorders. Comorbid general medical or psychiatric disorder was not used as an exclusion criterion. No language restriction was applied.

### Data extraction

Studies were coded according to patient characteristics, intervention characteristics and general characteristics of the study. We coded the following characteristics: type of recruitment (through the community, from clinical samples, other recruitment method); diagnosis of depression (diagnosed mood disorder, other definition of depression – usually a high score on a self-rating instrument); target group (adults in general, older adults, student populations, women with postpartum depression, general medical patients with depression, other target groups); type of psychotherapy (cognitive-behavioural therapy according to the manual by Beck *et al*,<sup>17</sup> other cognitive-behavioural therapy in which cognitive restructuring is the core element, interpersonal psychotherapy, problem-solving therapy, non-directive supportive therapy, behavioural activation treatment, other psychotherapy; full definitions of these treatments are given elsewhere);<sup>16</sup> treatment format (individual, group, guided self-help); type of control group (waiting list, care as usual, pill placebo, other control group); and data analyses (intention to treat, completers only).

All studies were coded by two independent assessors. When they disagreed, they discussed the differences with a third reviewer until agreement was reached. A study of the quality of the included studies has been reported elsewhere.<sup>18</sup>

### Statistical analysis

We first calculated effect sizes ( $d$ ) for each study by subtracting (at post-test) the average score of the control group from the average score of the experimental group and dividing the result by the pooled standard deviations of the experimental and control group. An effect size of 0.5 thus indicates that the mean of the experimental group is half a standard deviation larger than the mean of the control group. Effect sizes of 0.80 can be assumed

to be large, effect sizes of 0.50 are moderate and effect sizes of 0.20 are small.<sup>19</sup> In the calculations of effect sizes, only standardised instruments were used that explicitly measured depression (online Table DS1). If more than one depression measure was used, the mean of the effect sizes was calculated, such that each study (or contrast group) contributed only one effect size to the meta-analysis. When means and standard deviations were not reported we used other statistics ( $t$ ,  $P$ ) to calculate effect sizes.

To calculate the pooled mean effect size we used the computer program Comprehensive Meta-analysis version 2.2.021 for Windows, developed for support in meta-analysis ([www.meta-analysis.com](http://www.meta-analysis.com)). We conducted all analyses using the random effects model.<sup>20</sup> Subgroup analyses were conducted according to the Comprehensive Meta-analysis procedures. In the subgroup analyses we used mixed effects analyses that pooled studies within subgroups with the random effects model but tested for significant differences between subgroups with the fixed effects model.

In order to assess the heterogeneity of effect sizes we calculated the  $I^2$  statistic, which is an indicator of heterogeneity in percentages. A value of 0% indicates no observed heterogeneity; larger values show increasing heterogeneity, with 25% as low, 50% as moderate and 75% as high heterogeneity.<sup>21</sup> We also calculated the  $Q$  statistic, but only report whether this was significant or not.

### Sensitivity analyses

In our analyses we included studies in which two or more psychological treatments were compared with a control group. This means that multiple comparisons from these studies were included in the same analysis. These multiple comparisons, however, are not independent of each other, which may have resulted in an artificial reduction of heterogeneity. This may influence the overall effect size and our estimates of publication bias. Therefore, we conducted additional meta-analyses in which we included only one comparison per study. First, we included only the comparison with the largest effect size (i.e. the largest difference between the psychotherapy and control group), followed by another analysis in which only the comparison with the smallest effect size was included.

Because many different effect measures were used in the studies, we also conducted separate analyses in which the effect sizes were based on the two measures that were used in most studies: the Beck Depression Inventory (BDI) and the Hamilton Rating Scale for Depression (HRSD). Furthermore, we conducted analyses in which possible outliers (defined as studies with an effect size of 1.5 or larger) were removed, because these might distort the overall results.

### Indications of publication bias

In order to examine the possibility of publication bias, we conducted several tests.

**Funnel plot.** Perhaps the most common method that has been proposed to detect the existence of publication bias in a meta-analysis is the funnel plot.<sup>22</sup> This plots a measure of study size (the standard error) on the vertical axis as a function of effect size on the horizontal axis. Large studies appear at the top of the graph and tend to cluster near the mean effect size. Smaller studies appear towards the bottom of the graph. As there is more sampling variation in effect size estimates in the smaller studies, they will be dispersed across a range of values.<sup>23</sup> Visual inspection of a funnel plot can give an indication of publication bias. The

studies can be expected to be distributed symmetrically about the pooled effect size when publication bias is absent. In the presence of bias, it can be expected that the lower part of the plot will show a higher concentration of studies on one side of the mean than on the other. This is caused by the fact that smaller studies (appearing towards the bottom of the funnel plot) are more likely to be published if they have larger than average effects, which makes them more likely to meet the criterion for statistical significance.<sup>23</sup> Although funnel plots can be constructed on the basis of the standard error or on trial size,<sup>24</sup> we decided to use the standard error on the vertical axis because this is the most widely used method, and because it is readily available in the Comprehensive Meta-analysis software.

**Duval & Tweedie's trim and fill procedure.** If a meta-analysis has included all relevant studies, the funnel plot can be expected to be symmetric and dispersed equally on either side of the mean effect.<sup>23</sup> If, on the other hand, the funnel plot is asymmetric (with more studies on the right of the mean effect size than on the left) this could indicate publication bias. Duval & Tweedie developed a method of imputing missing studies, based on the assumption that the studies should be equally distributed on both sides of the mean effect size.<sup>25</sup> This procedure, usually called the Duval & Tweedie trim and fill procedure, yields an estimate of the effect size after the publication bias has been taken into account (adjusted effect size) and also indicates how many studies were imputed to correct for publication bias.

**Begg & Mazumdar rank correlation test.** The Begg & Mazumdar rank correlation test is based on the assumption that studies with larger sample sizes are published more often and that studies with an equal sample size are published less often when the effect size is smaller.<sup>26</sup> Therefore, it can be expected that in the case of publication bias there is a negative correlation between the standardised effect size and the standard errors of these effects. This correlation is tested with Kendall's tau: a significant value indicates possible publication bias. Because publication bias is expected to reduce the mean effect size, the significance test is one-tailed.

**Egger's test of the intercept.** Egger's linear regression method,<sup>27</sup> like the Begg & Mazumdar rank correlation test, is intended to quantify the bias captured by the funnel plot. In the Egger test the standard normal deviation is regressed on precision, defined as the inverse of the standard error.<sup>23</sup> The intercept in this regression corresponds to the slope in a weighted regression of the effect size on the standard error. In our results we report the intercept, the 95% confidence interval and the significance (one-tailed). Power for this test is generally higher than power for the rank correlation method, but is still low unless there is severe bias or a substantial number of studies.<sup>28</sup>

## Results

The selection and inclusion of studies are summarised in Fig. 1.<sup>29</sup> All inclusion criteria were met by 117 studies, in which 175 psychological treatment conditions were compared with a control group. These studies included a total of 9537 participants (5481 in the psychotherapy conditions and 4056 in the control conditions). An overview of selected characteristics of the included studies is presented in online Table DS1 together with a full list of references.

## Indicators of publication bias in the full sample

The overall effect size of the 175 comparisons between psychotherapy and a control condition was 0.67 (95% CI 0.60–0.75), with high heterogeneity ( $I^2 = 70.27$ ). Adjustment for publication bias according to Duval & Tweedie's trim and fill procedure resulted in a mean effect size of 0.42 (95% CI 0.33–0.51) with 51 studies missing. Both Begg & Mazumdar's test and Egger's test resulted in highly significant indicators of publication bias ( $P < 0.001$ ). The results of these analyses are summarised in Table 1. The funnel plot of the effect sizes of the studies (with the standard error on the vertical axis and the effect size at the horizontal axis), first without and then with the imputed studies, clearly shows that smaller studies with lower effect sizes are missing (Fig. 2) and this is confirmed after the imputation of missing studies.

Because the overall results of our analyses may have been influenced by possible outliers, we conducted a new meta-analysis in which all effect sizes from 1.5 or larger were removed. As can be seen in Table 1, the mean effect size dropped somewhat ( $d = 0.51$ ), but all indicators of publication bias remained highly significant ( $P < 0.001$ ), and Duval & Tweedie's trim and fill procedure estimated the adjusted effect size to be considerably smaller ( $d = 0.39$ ), with a total of 38 missed studies.

In order to examine the influence of multiple comparisons from one study, we conducted another meta-analysis in which we included only one comparison per study (online Table DS2). From the 46 studies with multiple comparisons we included only the comparison with the largest effect size (i.e. the largest difference between the psychotherapy and control groups). We then conducted another analysis in which only the comparison with the smallest effect size was included. As can be seen in Table 1, the resulting mean effect sizes differed somewhat from the overall analyses, and heterogeneity dropped a little, but all indicators of publication bias remained highly significant ( $P < 0.001$ ).

When we limited the analyses to the effect sizes found for the BDI, somewhat larger effect sizes were found, but again all indicators for publication bias remained highly significant ( $P < 0.001$ ). The same was true when we limited the analyses to the effect sizes found for the HRSD.

## Publication bias in subgroups of studies

It is possible that the publication bias we found differs for subtypes of studies. We therefore conducted a series of analyses in which we first selected a subgroup of studies based on a specific characteristic, and then examined the indicators of publication bias within this subgroup (online Table DS2). Significant indicators of publication bias were found for most subgroups of studies. No indication of publication bias was found for studies examining interpersonal psychotherapy, studies examining psychotherapy for women with postpartum depression, and studies in which patients were not recruited through the community or from clinical samples (usually from general medical patient groups or systematic screening).

## Publication bias in studies of cognitive-behavioural therapy

Because more than half of the comparisons examined cognitive-behavioural therapy, we decided to investigate publication bias in these studies in more detail. The results of these analyses are presented in Fig. 2 and in online Table DS3. The main analyses pointed at a considerable and significant risk of publication bias

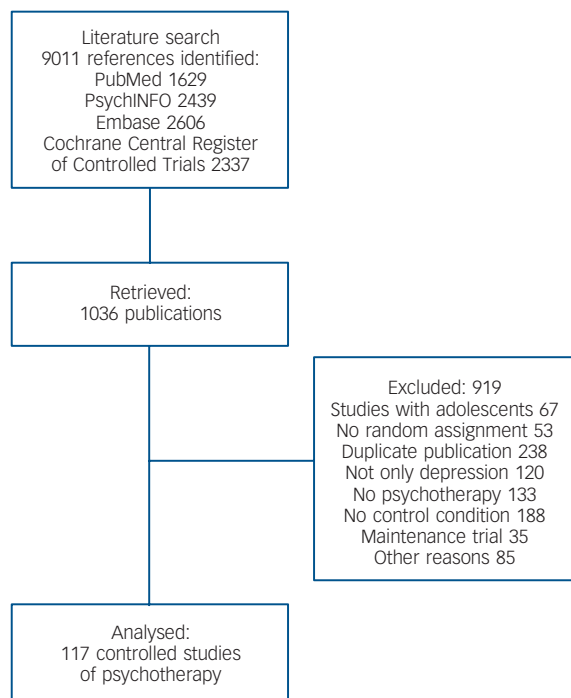


Fig. 1 Selection of studies.

among studies examining cognitive-behavioural therapy, and this remained high in all sensitivity analyses. The overall effect size of the 89 comparisons between cognitive-behavioural therapy and a control condition was 0.69, and after adjustment for publication bias this was reduced to 0.49 with 26 studies missing. Both Begg & Mazumbar's test and Egger's test were highly significant ( $P < 0.001$ ).

In the subgroup analyses we merged several subgroups because of the small number of studies per group (the target group had two categories: adults and 'more specific target group'; and placebo and other control groups were merged into 'other control group'). We also conducted a subgroup analysis in which we examined cognitive-behavioural therapy according to the manual by Beck *et al.*<sup>17</sup> and other types of cognitive-behavioural therapy: in the majority of subgroups, all indicators pointed at significant publication bias (online Table DS3).

## Discussion

We used a large sample of controlled studies of psychotherapy for adult depression to examine the possibility of publication bias. All tests for publication bias gave strong and highly significant indications of publication bias. The overall mean effect size of psychotherapy was 0.67, which corresponds with a number needed to treat (NNT) of 2.75.<sup>30</sup> After adjustment for publication bias the effect size was reduced to 0.42, which corresponds to an NNT of 4.27. When we examined the subsample of studies examining cognitive-behavioural therapy, the results were comparable. The overall effect size of cognitive-behavioural therapy was 0.69, and after adjustment for publication bias this was reduced to 0.49, with 26 studies missing. There were some subgroups of studies for which we found no indication of significant publication bias, including research on interpersonal psychotherapy and research on psychotherapy for women with postpartum depression. These results have to be considered with caution, however, because these subsamples were relatively small and may have been the result of random error.

## Limitations

This meta-analytic review has several limitations. The most important is that our tests for publication bias do not provide direct evidence of such bias. These procedures only test whether the funnel plot is symmetrical and whether small negative studies are missing. This cannot be considered as direct evidence, and there may in principle be other reasons why these studies are missing. It may be possible, for example, that early pilot studies of new treatments result in large effect sizes because the developer of such a new treatment is the best expert, realising larger effect sizes than later studies by other groups who test the effects of this treatment in routine care. Alternatively, it may be possible that such early pilot projects attract a specific type of patient, who is willing to undergo a new treatment. If these patients were more receptive to change or to treatment in general, that would result in larger effect sizes of smaller pilot studies. However, funnel plot asymmetry cannot be considered to be proof of bias in a meta-analysis,<sup>31</sup> and no statistical imputation method can recover the hidden and missing truth. With the highly significant indicators of funnel plot asymmetry, it is very likely that meta-analyses of these studies overestimate the true effect size of psychotherapy for adult depression.

We also used a rather narrow definition of publication bias in this study. A broader definition of publication bias might indicate

Table 1 Mean effect sizes ( $d$ ) for psychological treatments of adult depression from published studies: tests for publication bias in the overall sample

	$N_c$	Unadjusted values		Adjusted values <sup>a</sup>			Egger's test
		$d$ (95% CI)	$I^2$	$d$ (95% CI)	$N_{imp}$ <sup>a</sup>	Tau <sup>b</sup>	$I^c$ (95% CI)
All comparisons	175	0.67 (0.60–0.75)	70.27***	0.42 (0.33–0.51)	51	0.37***	1.58 (1.06–2.11) ***
Outliers excluded <sup>d</sup>	153	0.51 (0.45–0.57)	43.98***	0.39 (0.32–0.45)	38	0.26***	1.01 (0.58–1.44) ***
Multiple comparisons excluded (lowest retained)	117	0.57 (0.49–0.66)	67.98***	0.39 (0.29–0.49)	28		
Multiple comparisons excluded (highest retained)	117	0.70 (0.60–0.80)	74.87***	0.45 (0.33–0.56)	33	0.37***	1.63 (0.92–2.34) ***
BDI only	113	0.76 (0.65–0.87)	66.81***	0.45 (0.33–0.56)	38	0.42***	2.39 (1.78–2.99) ***
HRSD only	67	0.83 (0.70–0.96)	65.61***	0.54 (0.40–0.69)	21	0.47***	1.68 (0.99–2.36) ***

BDI, Beck Depression Inventory; HRSD, Hamilton Rating Scale for Depression;  $N_c$ , number of comparisons;  $N_{imp}$ , number of imputed studies.

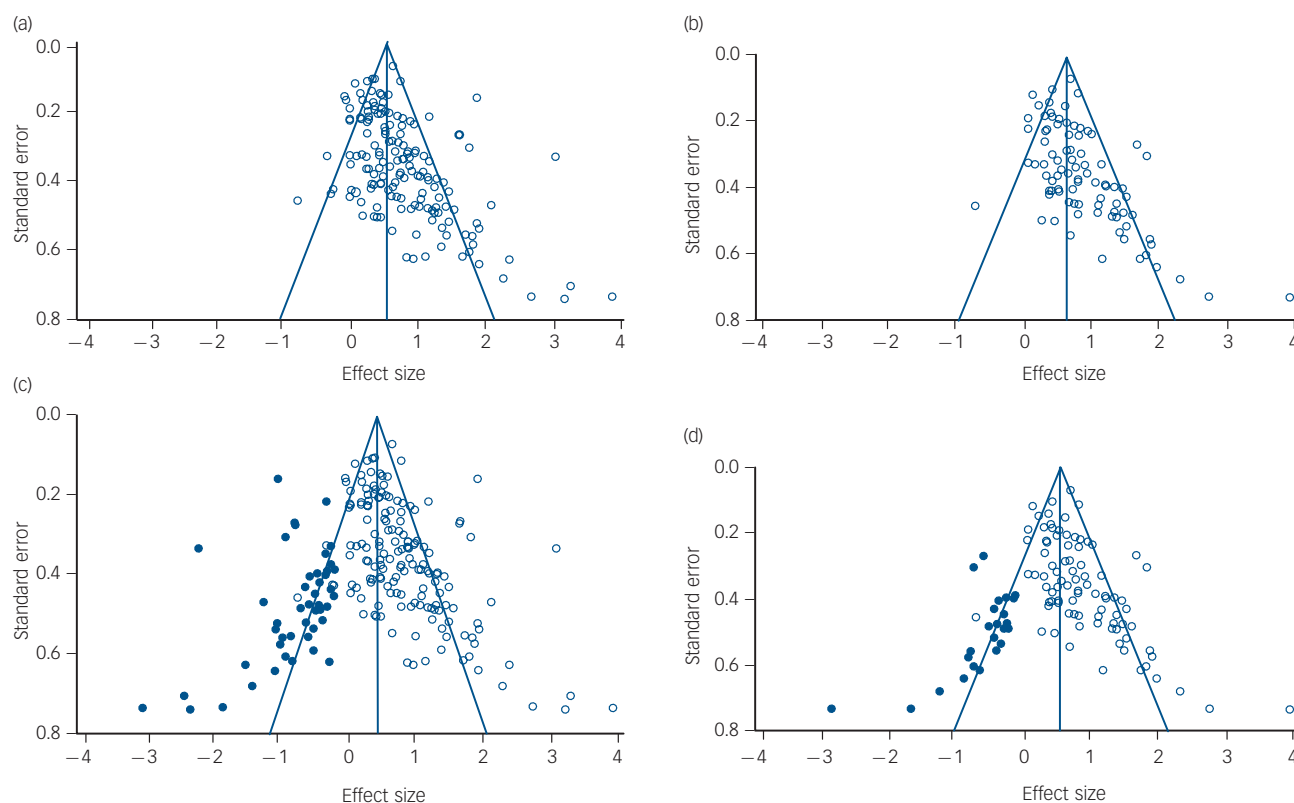
a. Based on Duval & Tweedie's trim & fill procedure.

b. Kendall's tau based on Begg & Mazumbar rank correlation.

c. Intercept based on Egger's test.

d. In these analyses, mean effect sizes of 1.5 and higher were excluded.

\* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ .



**Fig. 2** Funnel plots. (a) All psychotherapy studies, without imputed studies; (b) studies of cognitive-behavioural therapy (CBT) only, without imputed studies; (c) all psychotherapy studies, with imputed studies (black circles); (d) CBT studies only, with imputed studies. Imputation according to Duval & Tweedie trim and fill procedure.

not only selective publication of studies but also selective reporting of outcomes.<sup>32,33</sup> Such selective reporting of outcomes might also cause asymmetry of the funnel plot. A further limitation is that many of the included studies were small and although they were all randomised controlled trials they may not meet standard quality criteria for clinical trials. On the other hand the number of included studies was high, enabling us to control for several basic characteristics of the populations, interventions and study designs.

The statistical tests we used to assess the asymmetry of the funnel plot have several weaknesses.<sup>20,34</sup> A problem with Duval & Tweedie's trim and fill procedure is that it depends strongly on the assumptions of the model for why studies are missing, and the algorithm for detecting asymmetry can be influenced by one or two aberrant studies.<sup>34</sup> Begg & Mazumdar's test and Egger's test may also yield a different picture depending on the index used in the analyses,<sup>24</sup> and they tend to have low power. Furthermore, they only make sense if there is a reasonable amount of dispersion in the sample sizes and a reasonable number of studies. In our study, however, we included a large number of studies and we found strong indications of publication bias. In this situation, these weaknesses do not suggest that the main conclusion – that there is considerable risk of publication bias – should be doubted.

### Implications

Research on psychotherapy for adult depression does not seem to be any freer from publication bias than research on medication treatment, although it is likely that the sources of this bias differ between the two types of studies. There have long been concerns that the large pharmaceutical companies that fund much of the

drug research have an economic incentive to make the most positive case possible for the medications that they sell and that this incentive may influence the investigators whose research they support. Although psychological treatments are not supported by large pharmaceutical companies with strong economic interests in larger effect sizes, there appears to be a considerable risk of publication bias in studies of psychotherapy as well. Researchers of psychological treatments do have personal interests in publication of (larger) effects, as these are more likely to lead to tenure and lucrative workshop fees. Pharmaceutical companies have clear financial reasons to inflate research findings, and psychological investigators have both personal and professional reasons for doing the same. Results of a psychotherapy study might not be published because of authors' failure to submit the results of negative studies, journal editors preferring large significant outcomes over small non-significant effects, and negative reports by reviewers.<sup>9</sup> A recent study examining psychotherapy for depression among children and adolescents also found strong indications of publication bias, suggesting that this problem is not restricted to psychotherapy for adult depression alone.<sup>35</sup> Furthermore, there is a growing concern not only that research findings in the field of depression treatment are biased, but that most current published research findings are false and that claimed research findings may be simply accurate measures of prevailing bias.<sup>36</sup>

We strongly encourage psychotherapy researchers to register their studies in trial registries, as this will facilitate later investigations into publication bias. We also believe that the investigation of publication bias is a concern not only for psychotherapy research but for psychology in general. The 'file drawer' effect is probably present in all areas of psychological

research,<sup>37</sup> and may have implications for how research findings are interpreted. Meta-analysts are therefore encouraged to quantify systematically the number of studies likely to have been missed in their review, and to conduct sensitivity analyses as exemplified in this study to obtain an impression of the impact of publication bias on the overall estimate.

Not only are the effects of antidepressant medication overestimated because of publication bias, but the same seems to be true for psychotherapy for adult depression. The two most important first-line treatments of adult depression appear not to be as effective as is often assumed and it may be that the presumably active ingredients of treatment account for a smaller proportion of the outcomes observed than is widely believed. It may be time to search for new and more effective treatments of depression or to isolate the active ingredients of those that already exist, and to examine in more depth which treatments are most effective for which patients.

**Pim Cuijpers**, PhD, Department of Clinical Psychology, Vrije Universiteit (VU), and EMGO Institute, VU University Medical Centre, Amsterdam; **Filip Smit**, PhD, Department of Clinical Psychology, VU University, EMGO Institute, VU University Medical Center, Amsterdam, and Trimbos Institute, Utrecht; **Ernst Bohlmeijer**, PhD, Technical University Twente, Deventer, The Netherlands; **Steven D. Hollon**, PhD, Vanderbilt University, Nashville, Tennessee, USA; **Gerhard Andersson**, PhD, Department of Behavioural Sciences and Learning, Swedish Institute for Disability Research, Linköping University, and Department of Clinical Neuroscience, Psychiatry Section, Karolinska Institute, Stockholm, Sweden

**Correspondence:** Professor Pim Cuijpers, Department of Clinical Psychology, VU University Amsterdam, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands. Email: p.cuijpers@psy.vu.nl

First received 10 Mar 2009, final revision 20 Sep 2009, accepted 28 Oct 2009

## References

- Cuijpers P, van Straten A, Warmerdam L, Smits N. Characteristics of effective psychological treatments of depression: a meta-regression analysis. *Psychother Res* 2000; **18**: 225–36.
- Churchill R, Hunot V, Corney R, Knapp M, McGuire H, Tylee A, et al. A systematic review of controlled trials of the effectiveness and cost-effectiveness of brief psychological treatments for depression. *Health Technol Assess* 2001; **5**: 35.
- Gaffan EA, Tsaousis I, Kemp-Wheeler SM. Researcher allegiance and meta-analysis: the case of cognitive therapy for depression. *J Consult Clin Psychol* 1995; **63**: 966–80.
- Gloaguen V, Cottraux J, Cucherata M, Blackburn IM. A meta-analysis of the effects of cognitive therapy in depressed patients. *J Affect Disord* 1998; **49**: 59–72.
- McDermut W, Miller IW, Brown RA. The efficacy of group psychotherapy for depression: a meta-analysis and review of the empirical research. *Clin Psychol Sci Pract* 2001; **8**: 98–116.
- Cuijpers P, van Straten A, van Oppen P, Andersson G. Are psychological and pharmacological interventions equally effective in the treatment of adult depressive disorders? A meta-analysis of comparative studies. *J Clin Psychiatry* 2008; **69**: 1675–85.
- Rothstein HR, Sutton AJ, Borenstein M, eds. Publication bias. In *Publication Bias in Meta-analysis: Prevention, Assessment and Adjustment*: 277–302. Wiley & Sons, 2005.
- Sterne JAC, Egger M. Regression methods to detect publication and other bias in meta-analysis. In *Publication Bias in Meta-analysis: Prevention, Assessment and Adjustment* (eds HR Rothstein, AJ Sutton, M Borenstein): 88–111. Wiley, 2005.
- Turner EH, Matthews AA, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008; **358**: 252–60.
- Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the food and drug administration. *PLoS Med* 2008; **5**: e45.
- Cuijpers P, Dekker J. Psychologische behandeling van depressie: een systematisch overzicht van meta-analyses [Psychological treatment of depression: a systematic overview of meta-analyses]. *Ned Tijdschr Geneesk* 2005; **149**: 1892–7.
- Westen D, Morrison K. A multidimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: an empirical examination of the status of empirically supported therapies. *J Consult Clin Psychol* 2001; **69**: 875–99.
- Robinson LA, Berman JS, Neimeyer RA. Psychotherapy for the treatment of depression: a comprehensive review of controlled outcome research. *Psychol Bull* 1990; **108**: 30–49.
- Cuijpers P, van Straten A, Warmerdam L, Andersson G. Psychological treatment of depression: a meta-analytic database of randomized studies. *BMC Psychiatry* 2008; **8**: 36.
- Wampold BE, Mondin GW, Moody M, Stich F, Benson K, Hyun-nie A. A meta-analysis of outcome studies comparing bona fide psychotherapies: empirically, 'All must have prizes'. *Psychol Bull* 1997; **122**: 203–15.
- Cuijpers P, van Straten A, Andersson G, van Oppen P. Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. *J Consult Clin Psychol* 2008; **76**: 909–22.
- Beck AT, Rush AJ, Shaw BF, Emery G. *Cognitive Therapy of Depression*. Guilford, 1979.
- Cuijpers P, van Straten A, Bohlmeijer E, Hollon SD, Andersson G. The effects of psychotherapy for adult depression are overestimated: a meta-analysis of study quality and effect size. *Psychol Med* 2009; epub ahead of print: doi: 10.1017/S0033291709006114.
- Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. Academic Press, 1977.
- Higgins JPT, Green S, eds. *Cochrane Handbook for Systematic Reviews of Interventions* 4.2.6. Wiley, 2006.
- Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; **327**: 557–60.
- Light R, Pillemer D. *Summing Up: The Science of Reviewing Research*. Harvard University Press, 1984.
- Borenstein M. Software for publication bias. In *Publication Bias in Meta-analysis: Prevention, Assessment and Adjustments* (eds H Rothstein, AJ Sutton, M Borenstein): 391–403. Wiley, 2005.
- Tang JL, Liu JL. Misleading funnel plot for detection of bias in meta-analysis. *J Clin Epidemiol* 2000; **53**: 477–84.
- Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 2000; **56**: 455–63.
- Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994; **50**: 1088–101.
- Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997; **315**: 629–34.
- Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol* 2000; **53**: 1119–29.
- Moher D, Cook DJ, Eastwood S, Olkin I, Rennie D, Stroup DF. Improving the quality of reports of meta-analyses of randomised controlled trials: the QUOROM statement. *Lancet* 1999; **354**: 1896–900.
- Kraemer HC, Kupfer DJ. Size of treatment effects and their importance to clinical research and practice. *Biol Psychiatry* 2006; **59**: 990–6.
- Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. *J Clin Epidemiol* 2000; **53**: 1119–29.
- Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004; **291**: 2457–65.
- Furukawa TA, Watanabe N, Omori IM, Montori VM, Guyatt GH. Association between unreported outcomes and effect size estimates in Cochrane meta-analyses. *JAMA* 2007; **297**: 468–70.
- Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to Meta-Analysis*. Wiley, 2009.
- Watanabe N, Hunot V, Omori IM, Churchill R, Furukawa TA. Psychotherapy for depression among children and adolescents: a systematic review. *Acta Psychiatr Scand* 2007; **116**: 84–95.
- Ioannidis JP. Why most published research findings are false. *PLoS Med* 2005; **2**: e124.
- Rosenthal R. The 'file drawer problem' and tolerance for null results. *Psychol Bull* 1979; **86**: 638–41.

