Psychological Medicine (2014), **44**, 685–695. © Cambridge University Press 2013
doi:10.1017/S0033291713000457

REVIEW ARTICLE

Comparison of psychotherapies for adult depression to pill placebo control groups: a meta-analysis

P. Cuijpers[1,2]*, E. H. Turner[3,4], D. C. Mohr[5], S. G. Hofmann[6], G. Andersson[7,8], M. Berking[9] and J. Coyne[10]

[1] *Department of Clinical Psychology, VU University Amsterdam, The Netherlands*
[2] *EMGO Institute for Health and Care Research, The Netherlands*
[3] *Behavioral Health and Neurosciences Division, Portland Veterans Affairs Medical Center, Portland, OR, USA*
[4] *Departments of Psychiatry and Pharmacology, Oregon Health and Science University, Portland, OR, USA*
[5] *Department of Preventive Medicine, Northwestern University, Chicago, IL, USA*
[6] *Department of Psychology, Boston University, Boston, MA, USA*
[7] *Department of Behavioral Sciences and Learning, Swedish Institute for Disability Research, Linköping University, Sweden*
[8] *Department of Clinical Neuroscience, Psychiatry Section, Karolinska Institutet, Stockholm, Sweden*
[9] *Department of Clinical Psychology, Philipps-University of Marburg, Germany*
[10] *Health Psychology Section, Department of Health Sciences, University Medical Center, Groningen, University of Groningen, The Netherlands*

Background. The effects of antidepressants for treating depressive disorders have been overestimated because of selective publication of positive trials. Reanalyses that include unpublished trials have yielded reduced effect sizes. This in turn has led to claims that antidepressants have clinically insignificant advantages over placebo and that psychotherapy is therefore a better alternative. To test this, we conducted a meta-analysis of studies comparing psychotherapy with pill placebo.

Method. Ten 10 studies comparing psychotherapies with pill placebo were identified. In total, 1240 patients were included in these studies. For each study, Hedges' g was calculated. Characteristics of the studies were extracted for subgroup and meta-regression analyses.

Results. The effect of psychotherapy compared to pill placebo at post-test was g=0.25 [95% confidence interval (CI) 0.14–0.36, I^2=0%, 95% CI 0–58]. This effect size corresponds to a number needed to treat (NNT) of 7.14 (95% CI 5.00–12.82). The psychotherapy conditions scored 2.66 points lower on the Hamilton Depression Rating Scale (HAMD) than the placebo conditions, and 3.20 points lower on the Beck Depression Inventory (BDI). Some indications for publication bias were found (two missing studies). We found no significant differences between subgroups of the studies and in meta-regression analyses we found no significant association between baseline severity and effect size.

Conclusions. Although there are differences between the role of placebo in psychotherapy and pharmacotherapy research, psychotherapy has an effect size that is comparable to that of antidepressant medications. Whether these effects should be deemed clinically relevant remains open to debate.

Received 11 July 2012; Revised 1 February 2013; Accepted 5 February 2013; First published online 3 April 2013

Key words: Depression, meta-analysis, placebo, psychotherapy.

Introduction

Comparisons of psychotherapy for depression *versus* antidepressants have direct relevance to practice guidelines and to policy issues concerning deployment of clinical resources. Provision of medication and psychotherapy require different clinician training and skills and certification and licensure. However, previous estimates of the efficacy of antidepressants relative to pill placebo conditions based on published trials have been shown to be exaggerated because of selective publication. Meta-analyses incorporating data from both published and unpublished trials obtained from the US Food and Drug Administration (FDA) have yielded markedly lower estimates than those based on published data alone (Melander *et al.* 2003; Turner *et al.* 2008). Although these meta-analyses did not evaluate psychotherapy for depression, some have drawn inferences about the relative efficacy of antidepressants *versus* psychotherapy. The claim is that antidepressants have clinically insignificant advantages over pill placebo, and therefore alternative treatments such as psychotherapy should be exhausted before turning to medication for depression (Kirsch *et al.* 2008).

* Address for correspondence: P. Cuijpers, Ph.D., Professor of Clinical Psychology, Department of Clinical Psychology, VU University Amsterdam, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands.
(Email: p.cuijpers@vu.nl)

https://doi.org/10.1017/S0033291713000457 Published online by Cambridge University Press

Turner *et al.* (2008) extracted data for 12 antidepressants approved by the FDA between 1987 and 2004 and compared the FDA's regulatory decisions to what was reported in the literature. According to the FDA analyses, only half of the trials were positive whereas, according to the published literature, almost all of the trials were positive. Studies that the FDA deemed positive were 12 times more likely to be published than those deemed negative or questionable. The overall effect size (mean standardized difference) for antidepressants relative to placebo was reduced from 0.41 to 0.31. Nonetheless, antidepressant/pill placebo differences were statistically significant for all medications. Kirsch *et al.* (2008) subsequently examined FDA reviews for four of the 12 drugs examined by Turner *et al.* (2008) and found essentially the same effect size (0.32). However, Kirsch *et al.* (2008) applied two criteria for clinical significance proposed by the National Institute for Clinical Excellence (NICE): an effect size of 0.50 and an improvement on the Hamilton Depression Rating Scale (HAMD; Hamilton, 1960) of 3 points. In so doing, they concluded that the effects of antidepressants were clinically significant (though small) for severe depression but not clinically significant for mild to moderate levels of depression (NICE, 2004). Although NICE no longer uses these cutoffs as indicators of clinical relevance in its current guideline, this judgment became the basis of Kirsch *et al.*'s (2008) widely quoted claims to the general public that patients who took a placebo would obtain an effect almost as large as that obtained by taking an antidepressant. These results have been interpreted elsewhere to indicate that psychotherapy for depression is preferable to antidepressants (Kirsch, 2009).

The Kirsch *et al.* (2008) study did not involve comparing psychotherapy for depression to pill placebo, and comparisons between psychotherapy and pill placebo control groups are controversial. The ascendancy of the view that such comparisons are inappropriate for the evaluation of psychotherapy is reflected in, and reinforced by, the small number of studies allowing such a comparison. However, these psychotherapy studies were designed to enable the evaluation of an antidepressant condition *versus* pill placebo. In the context of a clinical trial, it seems clear enough why pill placebo should be contrasted with another pill (the active medication being evaluated), but it is less clear why pill placebo should be contrasted with something that is not in a pill form, namely psychotherapy. The difference between a pill and psychotherapy is obvious to both clinicians and patients, and this unblinding allows potentially strong clinician and patient preferences and expectations to come into play. Several more common options for control conditions exist, ranging from no treatment and waitlist controls, which may also lead to unblinding, to comparison conditions that control for the effects of nonspecific and specific treatment factors, which may limit unblinding of patients (Posternak & Zimmerman, 2007; Mohr *et al.* 2009).

However, there are arguments in favor of examining the effect sizes of psychotherapy *versus* pill placebo conditions, even while acknowledging these issues. First, a pill placebo condition is more than the simple administration of a pill, it involves active clinical management instilling positive expectations, and considerable support and attention. These aspects of a pill placebo condition potentially control for the effects of similar ingredients in the provision of psychotherapy in a way that is not achieved with waitlist or no-treatment conditions. Because it is clear to patients assigned to such treatment conditions that they are not receiving treatment, the associated effects would be expected to be fairly small. Such small effects will tend to inflate the difference between these conditions and psychotherapy treatment conditions, thereby inflating the apparent effects of psychotherapy. If such comparisons are to be used, it seems warranted to include pill placebo at least an auxiliary comparison.

However, the most compelling argument for a pill placebo/psychotherapy comparison is pragmatic: the efficacy of psychotherapy *versus* antidepressants treatment for depression has important clinical, policy and economic implications. To compare the effects of two treatment modalities, we should not compare the effects of one to apples and the other to oranges. Rather, the effects of the two modalities should be compared to a common benchmark. One approach would be to reassess the efficacy of antidepressants by contrasting them to waitlist or no-treatment control conditions; however, because the design of such trials would be open-label rather than double-blind, the resulting effect sizes would probably be spuriously inflated. The alternative approach is to compare both treatment modalities to pill placebo. Although there are ample efficacy data on antidepressants compared to pill placebo, there are relatively few data on psychotherapy compared to pill placebo.

Another alternative would be to examine direct comparisons of the two treatments. Several meta-analyses have examined these direct comparisons of psychotherapy and pharmacotherapy, and they typically find no significant differences (Cuijpers *et al.* 2008*b*; Imel *et al.* 2008). However, there is a persuasive argument that head-to-head studies, involving two active treatments, can be uninterpretable (Temple & Ellenberg, 2000). The reason is that, in the absence of placebo, there is no way of knowing whether the two treatments are equally effective or equally ineffective. One example is a well-known trial comparing

https://doi.org/10.1017/S0033291713000457 Published online by Cambridge University Press

St John's Wort to sertraline to placebo for depression (Hypericum Depression Trial Study Group, 2002). Ignoring the placebo, it could be concluded that St John's Wort performs about as well as sertraline, hence it must be effective. Full response occurred more often in patients treated with placebo than in those treated with either active treatment. Hence, even though comparative studies are informative, comparisons with placebo remain important.

We therefore focus in this study on the efficacy of psychotherapy relative to pill placebo. Because pill placebo more plausibly controls for positive expectations, support and attention, it is conceivable that this approach will lead to effect sizes that are smaller than those historically found using waitlist and no-treatment control conditions. This would have important implications for clinical and policy decisions concerning provision of psychotherapy *versus* antidepressants for depression. Because earlier research based on head-to-head comparisons of psychotherapy and pharmacotherapy have found that their effects on depression are comparable (Cuijpers *et al.* 2008), our hypothesis is that the effect size of psychotherapy is comparable to that of pharmacotherapy, that is $g=0.3$ (Turner *et al.* 2008).

Method

Identification and selection of studies

We used a database of 1344 papers on the psychological treatment of depression that has been described in detail elsewhere (Cuijpers *et al.* 2008), and that has been used in a series of earlier published meta-analyses (www.evidencebasedpsychotherapies.org). This database is continuously updated through comprehensive literature searches (from 1966 to January 2012). In these searches we examined 13407 abstracts in PubMed (3320 abstracts), PsycInfo (2710), EMBASE (4389) and the Cochrane Central Register of Controlled Trials (2988). These abstracts were identified by combining terms indicative of psychological treatment and depression (both MeSH terms and text words). For this database, we also checked the primary studies from 42 meta-analyses of psychological treatment for depression to ensure that no published studies were missed (www.evidencebasedpsychotherapies.org). From the 13407 abstracts (9860 after removal of duplicates), 1344 full-text papers were retrieved for possible inclusion in the database.

We included trials that (*a*) were randomized, (*b*) examined the effects of a psychological treatment and (*c*) used pill placebo as a comparator, in patients who were (*d*) adults and (*e*) diagnosed with a depressive disorder. Studies that included participants with co-morbid general medical or psychiatric disorders were not used as an exclusion criterion. No language restrictions were applied.

Quality assessment and data extraction

We assessed the validity of included studies using four criteria of the Risk of Bias assessment tool, developed by the Cochrane Collaboration (Higgins & Green, 2008). This tool assesses possible sources of bias in randomized trials, including the adequate generation of allocation sequence; the concealment of allocation to conditions; the prevention of knowledge of the allocated intervention (masking of assessors); and dealing with incomplete outcome data (this was assessed as positive when intention-to-treat analyses were conducted, meaning that all randomized patients were included in the analyses). Two other criteria of the Risk of Bias assessment tool were not used in this study because we found no clear indication in any of the studies that these had influenced the validity of the study (suggestions of selective outcome reporting; and other problems that could put it at a high risk of bias).

In addition to indicators of study quality, we coded several aspects of the included studies, including the following participant characteristics: recruitment method (community, from clinical samples, or other), definition of depression (assessment with a diagnostic interview or not), and target group (adults in general, or more specific target groups such as older adults). We also assessed the following intervention characteristics: format (individual or group), number of sessions, and the type of psychotherapy [cognitive behavior therapy (CBT), interpersonal psychotherapy (ITP), or other].

Because all studies reported the baseline score on the HAMD, we also examined the effects of baseline severity by comparing randomized controlled trials (RCTs) in which patients had a mean baseline HAMD<17, reflecting mild depression, with RCTs in which patients had a mean baseline of HAMD>18, reflecting moderate (HAMD 18–24) or severe (HAMD>25) depression (Katz *et al.* 1995). We also conducted a meta-regression analysis with effect size as the dependent variable and baseline severity (as the continuous variable) as the predictor, to be consistent with previous research on the question of the effect of antidepressant medication as a function of initial severity (Kirsch *et al.* 2008; Fournier *et al.* 2010). Data extraction was conducted by two independent researchers.

Meta-analyses

For each comparison between a psychotherapy and a pill placebo control group, the effect size indicating the difference between the two groups at post-test

https://doi.org/10.1017/S0033291713000457 Published online by Cambridge University Press

was calculated (Hedges' g or the standardized mean difference). Effect sizes were calculated by subtracting (at post-test) the average score of the psychotherapy group from the average score of the placebo group, and dividing the result by the pooled standard deviation. Because several studies had relatively small sample sizes, we corrected the effect size for small sample bias according to the procedures suggested by Hedges & Olkin (1985).

In the calculations of effect sizes, we only used those instruments that explicitly measured symptoms of depression, such as the Beck Depression Inventory (BDI; Beck *et al.* 1961) or the HAMD (Hamilton, 1960). If more than one depression measure was used, the mean of the effect sizes was calculated so that each comparison yielded only one effect. If dichotomous outcomes were reported without means and standard deviations, we used the procedures of the Comprehensive Meta-Analysis software to calculate the standardized mean difference. To calculate pooled mean effect sizes, we used the Comprehensive Meta-Analysis version 2.2.021. As we expected considerable heterogeneity among the studies, we used a random effects pooling model.

Because the standardized mean difference (Hedges' g) is not easy to interpret from a clinical perspective, we transformed these values into the NNT, using the formulae provided by Kraemer & Kupfer (2006). The NNT indicates the number of patients that have to be treated to generate one additional positive outcome (Laupacis *et al.* 1988).

As a test of homogeneity 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 and larger values indicate increasing heterogeneity, with 25% as low, 50% as moderate and 75% as high heterogeneity (Higgins *et al.* 2003). We calculated 95% confidence intervals (CIs) around I^2 (Ioannidis *et al.* 2007), using the non-central χ^2-based approach within the heterogi module for Stata (Orsini *et al.* 2005). We also calculated the Q statistic, but only report whether this was significant.

Subgroup analyses were conducted according to the mixed effects model, in which studies within subgroups are pooled with the random effects model whereas tests for significant differences between subgroups are conducted with the fixed effects model. For continuous variables, we used meta-regression analyses to test whether there was a significant relationship between the continuous variable and effect size, as indicated by a Z value and an associated p value.

Publication bias was tested by inspecting the funnel plot on primary outcome measures and by the trim-and-fill procedure of Duval & Tweedie (2000), which yields an estimate of the effect size after the publication bias has been taken into account (as implemented in Comprehensive Meta-Analysis version 2.2.021). We also conducted Egger's test of the intercept to quantify the bias captured by the funnel plot and to test whether it was significant. Furthermore, we calculated the fail-safe N, which indicates the number of studies with an effect size of zero that cause the resulting effect size to be non-significant.

Power calculation

Because we expected only a limited number of studies, we conducted a power calculation to examine how many studies would have to be included to have sufficient statistical power to identify relevant effects. We conducted a power calculation according to the procedures described by Borenstein *et al.* (2009).

First, we calculated how many studies are needed to identify an effect size of 0.5, which has been proposed as the threshold for a clinically relevant effect. These calculations indicated that we would need to include at least five studies with a mean sample size of 60 (30 participants for each condition) to be able to detect an effect size of $g=0.50$ (conservatively assuming a high level of between-study variance, τ^2, a statistical power of 0.80, and a significance level, α, of 0.05). Alternatively, we would need three studies with 80 participants each to detect an effect size of $g=0.50$. Second, we calculated how many studies are needed to identify an effect size of $g=0.31$, the effect size found for pharmacotherapy *versus* placebo (Turner *et al.* 2008). We found that we would need 11 studies with a sample size of 60 (30 for each condition) or nine studies with a sample size of 80 (40 for each condition) to identify an effect size of $g=0.31$.

Results

Selection and inclusion of studies

After examining a total of 13407 abstracts (9860 after removal of duplicates), we retrieved 1344 full-text papers for further consideration. We excluded 1334 of the retrieved papers. The flowchart describing the inclusion process, including the reasons for exclusion, is presented in Fig. 1. Ten of the 1344 retrieved full-text papers included a comparison between a psychotherapy and a pill placebo control group and were included in this meta-analysis.

Characteristics of included studies

Selected characteristics of the included studies are presented in Table 1. A total of 1240 patients were

https://doi.org/10.1017/S0033291713000457 Published online by Cambridge University Press

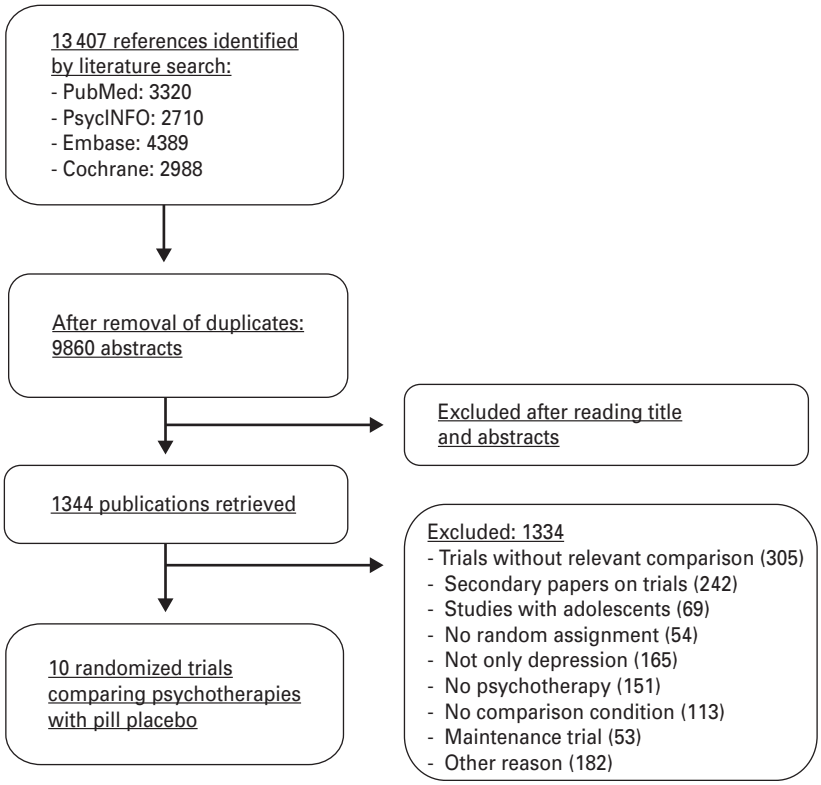

Fig. 1. Flowchart of inclusion of studies.

included in the 10 studies (668 in the psychotherapy conditions and 572 in the placebo conditions).

Five studies recruited patients from clinical samples, four studies recruited patients from the community and one study used another recruitment strategy. Seven studies were aimed at patients with major depressive disorder (MDD), two studies were aimed at patients with dysthymia or minor depression, and one study was aimed at patients with major depression, dysthymia or minor depression. The mean baseline severity was mild (HAMD<17) in three studies, moderate (HAMD 18–24) in six studies and severe (HAMD>25) in one study (Katz *et al.* 1995).

In the 10 studies, 12 psychotherapies were compared with a placebo condition (in each of two studies, two different psychotherapies were compared with the placebo condition). Four of the 12 psychotherapies were cognitive behavior therapy, three were problem-solving therapy, two ITP, another two were supportive therapy, and one was behavioral activation therapy. Eleven of the 12 psychotherapies used an individual format, and one used a group format. The number of treatment sessions ranged from six to 20. Patients in all studies were randomized to a one of three conditions (psychotherapy, medication or placebo), apart from one study where there were more than three conditions (Hegerl *et al.* 2010: CBT, medication, placebo, patient preference arm, guided self-help control).

The quality of the included studies varied somewhat, but was generally high. All but one study reported an adequate sequence generation. Six of the 10 studies reported allocation to conditions by an independent (third) party. All 10 studies reported blinding of outcome assessors and in nine of the 10 studies intention-to-treat analyses were conducted. Six of the 10 studies met all four quality criteria, three met three of four criteria; one study had a lower quality (it met only one of the four criteria).

Effects of psychotherapy compared to pill placebo control groups

The effect of psychotherapy compared to pill placebo at post-test was $g=0.25$ (95% CI 0.14–0.36, $I^2=0\%$, 95% CI 0–58%). This effect size corresponds with an NNT of 7.14 (95% CI 5.00–12.82). The effect sizes and 95% CIs of each study are presented in Fig. 2 and separately for each type of psychotherapy in Supplementary Table S1. A *post-hoc* power calculation showed that the statistical power was 0.99.

We included two studies, each of which compared two psychological treatments with a pill placebo group (Elkin *et al.* 1989; Dimidjian *et al.* 2006). Thus,

https://doi.org/10.1017/S0033291713000457 Published online by Cambridge University Press

Table 1. *Selected characteristics of studies comparing psychotherapy for adult depression with pill placebo control groups*

Study	Recruitment	Target group	Diagnosis	Conditions	n	Format	$n_{sessions}$	Medication[a]	Contact with pharmacotherapist	Country	Quality			
											SG	CA	BA	ITT
Barber, 2011	Community	Adults	MDD (DSM-IV, SCID), HAMD>14	1. Supportive 2. Placebo	51 50	Individual	20	Sertraline	Weekly/biweekly meetings with psychopharmacologist	USA	+	–	+	+
Barrett, 2001	Clinic	Adults	Dysthymia or minor depression (DSM-III-R)+HAMD>10	1. PST 2. Placebo	80 81	Individual	6	Paroxetine	Six treatment sessions of 10–15 min over 11 weeks	USA	+	+	+	+
DeRubeis, 2005	Community	Adults	MDD (SCID, DSM-IV)	1. CBT 2. Placebo	60 60	Individual	14	Paroxetine	Weekly/biweekly meetings with pharmacotherapist (20–45 min)	USA	+	–	+	+
Dimidjian, 2006	Community	Adults	MDD (DSM-IV)+BDI-II>14+HAMD>17	1. CBT 2. Behavioral activation 3. Placebo	38 37 41	Individual	16	Paroxetine	Weekly/biweekly meetings with pharmacotherapist (30–45 min)	USA	+	–	+	+
Elkin, 1989	Clinic	Adults	MDD (RDC)+HAMD>14	1. CBT 2. IPT 3. Placebo	59 61 59	Individual	16	Imipramine	Weekly meetings with pharmacotherapist (20–60 min)	USA	+	+	+	+
Hegerl, 2010	Clinic	Adults	Dysthymia, minor depression or MDD+HAMD score 8–22	1. CBT 2. Placebo	59 61	Group	10	Sertraline	Biweekly meetings with psychiatrist	Germany	+	+	+	+
Jarrett, 1999	Clinic	Adults	MDD with atypical features (SCID, DSM-III-R)	1. CBT 2. Placebo	36 36	Individual	20	Sertraline	Weekly meetings with pharmacotherapist	USA	+	+	+	+
Mynors-Wallis, 1995	Clinic	Adults	MDD (RDC)+HAMD>13	1. PST 2. Placebo	30 30	Individual	6	Amitriptyline	Six sessions of the same duration as psychotherapy	UK	+	+	+	+
Sloane, 1985	Other	Older adults	MDD	1. IPT 2. Placebo	19 14	Individual	6	Nortriptyline	N.R.	USA	–	–	+	–
Williams, 2000	Community	Older adults	Dysthymia or minor depression (DSM-III-R)+HAMD>10	1. PST 2. Placebo	138 140	Individual	6	Paroxetine	Six treatment sessions of 15 min over 11 weeks	USA	+	+	+	+

BA, Blinding of assessors; BDI, Beck Depression Inventory; CA, concealment of allocation to conditions; CBT, cognitive behavior therapy; HAMD, Hamilton Depression Rating Scale; IPT, interpersonal psychotherapy; ITT, intention-to-treat analyses; MDD, major depressive disorder; PST, problem-solving therapy; RDC, research diagnostic criteria; SG, adequate generation of allocation sequence; N.R., not reported.

[a] In all studies patients were randomized to one of three conditions: (1) psychotherapy, (2) medication or (3) placebo condition (only in the study from Hegerl *et al.* 2010 were patients randomized to one of five conditions: CBT, medication, placebo, patient preference arm, guided self-help control).

https://doi.org/10.1017/S0033291713000457 Published online by Cambridge University Press

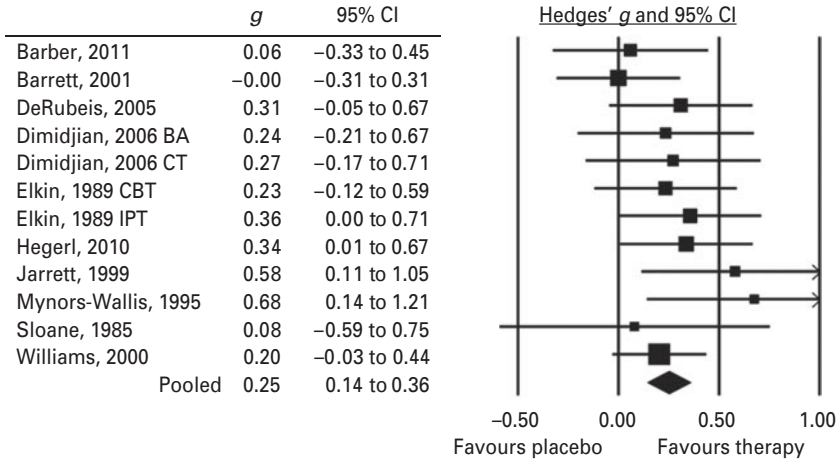

Fig. 2. Standardized effect sizes of psychotherapy for adult depression compared with control conditions: Hedges' *g*.

multiple comparisons from these studies were included in the same analysis, while these comparisons are not independent of each other. This may have resulted in an artificial reduction in heterogeneity and may have affected the pooled effect size. We examined the possible effects of this by conducting an analysis in which we included only one effect size per study. First, we included only the comparisons with the largest effect size from these studies and then we conducted another analysis in which we included only the smallest effect sizes. As shown in Table 2, the resulting effect sizes were almost the same as in the overall analyses.

We also calculated the effect sizes based on the HAMD (while excluding effect sizes based on other measurement instruments) and found comparable results (g=0.34, 95% CI 0.21–0.46, I^2=0, NNT=5.26, 95% CI 3.91–8.47). The psychotherapy conditions scored 2.66 points lower on the HAMD than the placebo conditions (95% CI 1.62–3.71). The effect size based exclusively on the BDI was also comparable with the overall effect size (g=0.30, 95% CI 0.13–0.46, I^2=0, NNT=5.95, 95% CI 3.91–13.51). The psychotherapy conditions scored 3.20 points lower on the BDI than the placebo conditions (95% CI 1.35–5.04).

Inspection of the funnel plot and Duval & Tweedie's trim-and-fill procedure indicated the presence of some publication bias. After adjustment for missing studies, the effect size dropped from g=0.25 to g=0.21 (95% CI 0.10–0.32, number of trimmed studies=2) and Egger's test did not indicate an asymmetric funnel plot (intercept: 1.25, 95% CI 1.00–3.50, df 10, p=0.24). The fail-safe N was 58, indicating that 58 studies with an effect size of zero would have to be found to make the result non-significant.

Long-term follow-up effects were examined in only three of the 10 studies (Elkin *et al.* 1989; Jarrett *et al.* 1999; Dimidjian *et al.* 2006), and because of this small number of studies we decided not to examine these further.

Subgroup and meta-regression analyses

We conducted a series of analyses to examine associations between characteristics of the studies and the effect sizes.

One study delivered psychotherapy in group format (Hegerl *et al.* 2010) whereas the other studies used an individual format. In the same study, no formal diagnostic interview was used to establish the presence of a depressive disorder. We examined whether removal of this study resulted in a different mean effect size. We found no indication that this study had an effect on the mean effect size (after removal of this study g=0.24, 95% CI 0.13–0.35, I^2=0, NNT=7.46; Table 2).

In one study, the randomization procedure was unclear, and no intention-to-treat analyses were reported, so the overall quality score for this study was low. Removal of this study had little impact on the overall outcomes (g=0.25, 95% CI 0.15–0.36, I^2=0, 95% CI 0–60, NNT=7.14, 95% CI 5.00–11.90).

We conducted a series of subgroup analyses to examine the association between study characteristics and the effect size. We found no indication that the effects size was significantly associated with type of psychotherapy, recruitment method (clinical samples *versus* other), target group (adults in general *versus* more specific target group), or whether the study was aimed exclusively at patients with MDD or at patients who might also have dysthymia, minor depression or dysthymia. The results are summarized in Table 2.

Finally, a meta-regression analysis with Hedges' g as the dependent variable did not indicate a significant

https://doi.org/10.1017/S0033291713000457 Published online by Cambridge University Press

Table 2. *Effects of psychotherapies compared with pill placebo control groups: Hedges' g*[a]

	n	*g* (95% CI)	*Z* value	I^{2b} (95% CI)	p^c	NNT (95% CI)
All studiess	12	0.25 (0.14–0.36)	4.59**	0 (0–58)		7.14 (5.00–12.82)
HAMD only	10	0.34 (0.21–0.46)	5.09**	0 (0–62)		5.26 (3.91–8.47)
BDI only	7	0.30 (0.13–0.46)	3.50**	0 (0–71)		5.95 (3.91–13.51)
One effect size per study (highest)	10	0.25 (0.14–0.37)	4.28**	0 (0–62)		7.14 (4.85–12.82)
One effect size per study (lowest)	10	0.24 (0.12–0.35)	4.02**	0 (0–62)		7.46 (5.10–14.71)
One study removed						
Study without diagnosis removed[c]	11	0.24 (0.13–0.35)	4.17**	0 (0–60)		7.46 (5.10–13.51)
Low quality study removed[d]	11	0.25 (0.15–0.36)	4.61**	0 (0–60)		7.14 (5.00–11.90)
Subgroup analyses						
Type of therapy						
CBT	5	0.33 (0.16–0.50)	3.81**	0 (0–79)	0.23	5.43 (3.62–11.11)
Other therapy	7	0.20 (0.06–0.33)	2.84*	1 (0–71)		8.93 (5.26–29.41)
Recruitment						
Clinical samples	6	0.31 (0.13–0.50)	3.30*	30.49 (0–72)	0.38	5.75 (3.62–13.51)
Other	5	0.21 (0.06–0.36)	2.71*	0 (0–75)		8.47 (5.00–29.41)
Target group						
Adults in general	10	0.27 (0.15–0.39)	4.32**	0 (0–62)	0.55	6.58 (4.59–11.90)
Specific group	2	0.19 (–0.03 to 0.41)	1.67 N.S.	0		9.43[e]

HAMD, Hamilton Depression Rating Scale; BDI, Beck Depression Inventory; CBT, cognitive behavior therapy; CI, confidence interval; NNT, numbers needed to treat; N.S., not significant ($p>0.05$).

[a] According to the random effects model.

[b] The *p* values indicate whether the difference between the effect sizes in the subgroups is significant.

[c] This was also the only study in which psychotherapy was delivered in group format.

[d] This was also the only study in which the randomization procedure was unclear.

[e] The 95% CI was not calculated for this NNT because the lower limit was below zero.

* $p<0.01$, ** $p<0.001$.

association between effect size and the number of sessions in the psychotherapies (slope: 0.015, 95% CI –0.007 to 0.038, $p>0.1$), nor did we find an association between effect size and baseline depression severity according to the HAMD (slope: 0.016, 95% CI –0.014 to 0.045, $p>0.1$). We also conducted a meta-regression analysis with the number of sessions with the pharmacotherapist in the placebo conditions as predictor, but did not find a significant association (slope: 0.005, 95% CI –0.019 to 0.029, $p>0.1$).

Discussion

A systematic search of the literature yielded only 10 studies comparing 12 psychotherapies for depression to pill placebo. However, the finding of an effect size equal to 0.25 was robust across specific forms of therapies, and removal of any individual trial. Focusing only on the HAMD, we obtained an effect size of $g=0.34$, which, as hypothesized, was in the range that Turner *et al.* (2008) and Kirsch *et al.* (2008) obtained for antidepressants (effect size=0.32 and $g=0.31$ respectively).

Overall, when compared to pill placebo, psychotherapy for depression was observed to have an effect size well below the medium range ($g=0.5$), which some researchers have recently used as a cut-off value for clinical significance, that is a virtual litmus test. It has also been suggested that 3 points on the HAMD should be regarded as a cut-off for clinical significance. Compared to pill placebo, psychotherapy reduced depression by 2.66 points, which falls below this 3-point cut-off. Thus, if either of these cut-offs were applied to the effect of psychotherapy, consistent with the way others have applied them to the effect of antidepressant medications (Kirsch *et al.* 2008; Fournier *et al.* 2010), the conclusion would be that the effect of psychotherapy is also clinically insignificant. We would argue, however, that such a conclusion is not valid because this reasoning begins with a flawed premise; these cut-offs were chosen arbitrarily and without empirical evidence.

We also calculated the NNT, and found that one in seven patients has to be treated with psychotherapy to have one more successful outcomes than treatment with placebo. For a potentially debilitating disorder such as depression, some might consider this to be a

https://doi.org/10.1017/S0033291713000457 Published online by Cambridge University Press

clinically relevant outcome. However, what NNT value should qualify as clinically relevant seems open to debate.

There are several important differences between the studies of psychotherapy for depression included in our meta-analysis and the studies of antidepressants included in the meta-analyses of Turner *et al*. (2008) and Kirsch *et al*. (2008). The present meta-analysis relied on published trials and did not allow for the inclusion of both published and unpublished trials. In the absence of an FDA-like repository of data, how many trials are unpublished or which published trials have been subjected to outcome reporting bias cannot be known with certainty. Nevertheless, for psychotherapy trials, there is ample statistical evidence of publication bias (Cuijpers *et al.* 2010*a*) and strong investigator allegiance effects (Luborsky *et al.* 1999). The studies included in our meta-analysis were of high quality, relative to much of the psychotherapy literature (Cuijpers *et al.* 2010*b*), resulting in a better estimate of the true effect size.

It is impossible to blind comparisons of psychotherapy to pill placebo. In such a situation, patient preferences may affect outcomes more than in comparisons between antidepressants and pill placebo, although the direction of this is not clear. Because most patients prefer psychotherapy (Dwight-Johnson *et al.* 2000; van Schaik *et al.* 2004), those who are randomized to their preferred treatment may have better outcomes, which may result in higher effect sizes. Moreover, not only patients but also providers are not blinded as to whether patients are assigned to psychotherapy or pill placebo, and provider preferences may also affect the outcomes in unknown ways. Whether these preferences indeed affect outcomes should be examined in future research. It is also important to remember that patients and providers were blinded for the psychotherapy *versus* placebo conditions, but because the studies also included a pharmacotherapy condition they were blinded for the pharmacotherapy *versus* placebo conditions. This may have had an influence on the outcomes, although it is not clear which direction such an influence may have had on our outcomes.

Pill placebo groups are not inert conditions, and it is not surprising therefore that the difference between psychotherapy and pill placebo reported here is substantially smaller than the difference between psychotherapy *versus* waitlist control or no treatment (Cuijpers *et al.* 2008). Patients in pill placebo conditions are provided with positive expectations and considerable encouragement and support that may be sufficient to produce improvement (Rief *et al.* 2009). In our meta-analysis we found that there was frequent contact between patients and pharmacotherapists in the pill placebo conditions, and this may have made an important contribution to the improvement of patients.

This study has several additional limitations. There may have been a sampling bias because the included studies compared psychotherapy with pharmacotherapy, and patients with a strong preference for antidepressants or psychotherapy may have declined to enroll in these trials altogether. This bias could limit generalizability of these findings. Another important limitation was the relatively small number of included studies. Although we had sufficient statistical power to detect small effects and the quality of included studies was relatively high, the power was limited for some of the analyses, such as tests for subgroup differences and for publication bias. We also limited the study to the effects of acute treatments. There are indications, however, that the effects of psychological treatments last longer than those of pharmacological treatments (Hollon *et al.* 2005; David *et al.* 2008; Dobson *et al.* 2008). A final limitation is that we considered psychotherapy for depression as a monolithic treatment whereas in fact several different treatments were used in the included studies, such as CBT and IPT, among others. We did find a significant effect for CBT compared to placebo ($p<0.01$), but did not have enough comparisons for other types of psychotherapy. When other psychotherapies were compared directly with one another, very little evidence for a differential effect of different types of psychotherapies was found (Cuijpers *et al.* 2008).

The results of this meta-analysis suggest that psychotherapy for adult depression has an effect size that is comparable to that of antidepressant medications when compared with placebo. Both fall below the proposed cut-offs for clinical significance. Although some might question the clinical significance of both treatment modalities, we would instead question the validity of the proposed cut-offs. Whether cut-offs for clinical significance should exist and, if they do, what they should be are questions that, in our view, remain unresolved.

Supplementary material

For supplementary material accompanying this paper visit http://dx.doi.org/10.1017/S0033291713000457.

Declaration of Interest

None.

References

Barber J, Barrett MS, Gallop R, Rynn MA, Rickels K (2012). Short-term dynamic psychotherapy versus

https://doi.org/10.1017/S0033291713000457 Published online by Cambridge University Press

pharmacotherapy for major depressive disorder: a randomized, placebo-controlled trial. *Journal of Clinical Psychiatry* **73**, 66–73.

Barrett JE, Williams JW, Oxman TE, Frank E, Katon W, Sullivan M, Hegel MT, Cornell JE, Sengupta AS (2001). Treatment of dysthymia and minor depression in primary care: a randomized trial in patients aged 18 to 59. *Journal of Family Practice* **50**, 405–412.

Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J (1961). An inventory for measuring depression. *Archives of General Psychiatry* **4**, 561–571.

Borenstein M, Hedges LV, Higgins JPT, Rothstein HR (2009). *Introduction to Meta-Analysis*. Wiley: Chichester, UK.

Cuijpers P, Smit F, Bohlmeijer ET, Hollon SD, Andersson G (2010*a*). Is the efficacy of cognitive behaviour therapy and other psychological treatments for adult depression overestimated? A meta-analytic study of publication bias. *British Journal of Psychiatry* **196**, 173–178.

Cuijpers P, van Straten A, Andersson G, van Oppen P (2008*a*). Psychotherapy for depression in adults: a meta-analysis of comparative outcome studies. *Journal of Consulting and Clinical Psychology* **76**, 909–922.

Cuijpers P, van Straten A, Bohlmeijer E, Hollon SD, Andersson G (2010*b*). The effects of psychotherapy for adult depression are overestimated: a meta-analysis of study quality and effect size. *Psychological Medicine* **40**, 211–223.

Cuijpers P, van Straten A, van Oppen P, Andersson G (2008*b*). Are psychological and pharmacological interventions equally effective in the treatment of adult depressive disorders? A meta-analysis of comparative studies. *Journal of Clinical Psychiatry* **69**, 1675–1685.

Cuijpers P, van Straten A, Warmerdam L, Andersson G (2008*c*). Psychological treatment of depression: a meta-analytic database of randomized studies. *BMC Psychiatry* **8**, 36.

Cuijpers P, van Straten A, Warmerdam L, Smits N (2008*d*). Characteristics of effective psychological treatments of depression: a meta-regression analysis. *Psychotherapy Research* **18**, 225–236.

David D, Szentagotai A, Lupu V, Cosman D (2008). Rational emotive behavior therapy, cognitive therapy, and medication in the treatment of major depressive disorder: a randomized clinical trial, posttreatment outcomes, and six month follow-up. *Journal of Clinical Psychology* **64**, 728–746.

DeRubeis RJ, Hollon SD, Amsterdam JD, Shelton RC, Young PR, Salomon RM (2005). Cognitive therapy vs medications in the treatment of moderate to severe depression. *Archives of General Psychiatry* **62**, 409–416.

Dimidjian S, Hollon SD, Dobson KS, Schmaling KB, Kohlenberg RJ, Addis ME (2006). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology* **74**, 658–670.

Dobson KS, Hollon SD, Dimidjian S, Schmaling KB, Kohlenberg RJ, Gallop RJ, Rizvi SL, Gollan JK (2008). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the prevention of relapse and recurrence in major depression. *Journal of Consulting and Clinical Psychology* **76**, 468–477.

Duval S, Tweedie R (2000). Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* **56**, 455–463.

Dwight-Johnson M, Sherbourne CD, Liao D, Wells KB (2000). Treatment preferences among depressed primary care patients. *Journal of General Internal Medicine* **15**, 527–534.

Elkin I, Shea MT, Watkins JT, Imber SD, Sotsky SM, Collins JF (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program: general effectiveness of treatments. *Archives of General Psychiatry* **46**, 971–982.

Fournier JC, DeRubeis RJ, Hollon SD, Dimidjian S, Amsterdam JD, Shelton RC (2010). Antidepressant drug effects and depression severity: a patient-level meta-analysis. *Journal of the American Medical Association* **303**, 47–53.

Hamilton M (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry* **23**, 56–62.

Hedges LV, Olkin I (1985). *Statistical Methods for Meta-Analysis*. Academic Press: San Diego, CA.

Hegerl U, Hautzinger M, Mergl R, Kohnen R, Schütze M, Scheunemann W, Allgaier AK, Coyne J, Henkel V (2010). Effects of pharmacotherapy and psychotherapy in depressed primary-care patients: a randomized, controlled trial including a patients' choice arm. *International Journal of Neuropsychopharmacology* **13**, 31–44.

Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003). Measuring inconsistency in meta-analyses. *British Medical Journal* **327**, 557–560.

Higgins JPT, Green S (eds) (2008). *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.0.1 [updated September 2008]. Wiley: Chichester, UK.

Hollon SD, DeRubeis RJ, Shelton RC, Amsterdam JD, Salomon RM, O'Reardon JP, Lovett ML, Young PR, Haman KL, Freeman BB, Gallop R (2005). Prevention of relapse following cognitive therapy versus medication in moderate to severe depression. *Archives of General Psychiatry* **62**, 417–422.

Hypericum Depression Trial Study Group (2002). Effect of *Hypericum perforatum* (St John's wort) in major depressive disorder: a randomized controlled trial. *Journal of the American Medical Association* **287**, 1807–1814.

Imel ZE, Malterer MB, McKay KM, Wampold BE (2008). A meta-analysis of psychotherapy and medication in unipolar depression and dysthymia. *Journal of Affective Disorders* **110**, 197–206.

Ioannidis JPA, Patsopoulos NA, Evangelou E (2007). Uncertainty in heterogeneity estimates in meta-analyses. *British Medical Journal* **335**, 914–916.

Jarrett RB, Schaffer M, McIntire D, Witt-Browder A, Kraft D, Risser RC (1999). Treatment of atypical depression with cognitive therapy or phenelzine: a double-blind, placebo-controlled trial. *Archives of General Psychiatry* **56**, 431–437.

https://doi.org/10.1017/S0033291713000457 Published online by Cambridge University Press

Katz R, Shaw BF, Vallis TM, Kaiser AS (1995). The assessment of severity and symptom patterns in depression. In *Handbook of Depression* (ed. E. E. Beckham and W. R. Leber), 2nd edn, pp. 61–85. Guilford Press: New York.

Kirsch I (2009). *The Emperor's New Drugs: Exploding the Antidepressant Myth*. The Bodley Head: London.

Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT (2008). Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Medicine* **5**, 250–268.

Kraemer HC, Kupfer DJ (2006). Size of treatment effects and their importance to clinical research and practice. *Biological Psychiatry* **59**, 990–996.

Laupacis A, Sackett DL, Roberts RS (1988). An assessment of clinically useful measures of the consequences of treatment. *New England Journal of Medicine* **318**, 1728–1733.

Luborsky L, Diguer L, Seligman DA, Rosenthal R, Krause ED, Johnson S, Halperin G, Bishop M, Berman JS, Schweizer E (1999). The researcher's own therapy allegiances: a 'wild card' in comparisons of treatment efficacy. *Clinical Psychology: Science and Practice* **6**, 95–106.

Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B (2003). Evidence b(i)ased medicine selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *British Medical Journal* **326**, 1171–1173.

Mohr DC, Spring B, Freedland KE, Beckner V, Arean P, Hollon SD, Ockene J, Kaplan R (2009). The selection and design of control conditions for randomized controlled trials of psychological interventions. *Psychotherapy and Psychosomatics* **78**, 275–284.

Mynors-Wallis LM, Gath DH, Lloyd-Thomas AR, Tomlinson D (1995). Randomised controlled trial comparing problem solving treatment with amitriptyline and placebo for major depression in primary care. *British Medical Journal* **310**, 441–335.

NICE (2004). Depression: management of depression in primary and secondary care. Clinical Practice Guideline CG23. National Institute for Clinical Excellence: London.

Orsini N, Higgins J, Bottai M, Buchan I (2005). Heterogi: Stata module to quantify heterogeneity in a meta-analysis (http://EconPapers.repec.org/RePEc:boc:bocode:s449201). Accessed 27 February 2013.

Posternak MA, Zimmerman M (2007). Therapeutic effect of follow-up assessments on antidepressant and placebo response rates in antidepressant efficacy trials: meta-analysis. *British Journal of Psychiatry* **190**, 287–292.

Rief W, Nestoriuc Y, Weiss S, Welzel E, Barsky AJ, Hofmann SG (2009). Meta-analysis of the placebo response in antidepressant trials. *Journal of Affective Disorders* **118**, 1–9.

Sloane RB, Staples FR, Schneider LS (1985). Interpersonal therapy versus nortriptyline for depression in the elderly. In *Clinical and Pharmacological Studies in Psychiatric Disorders* (ed. G. Burrows, T. R. Norman and L. Dermerstein), pp. 344–346. John Libbey: London.

Temple R, Ellenberg SS (2000). Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: Ethical and scientific issues. *Annals of Internal Medicine* **133**, 455–463.

Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. *New England Journal of Medicine* **358**, 252–260.

van Schaik DJF, Klijn AFJ, van Hout HPJ, van Marwijk HWJ, Beekman ATF, de Haan M, van Dyck R (2004). Patients' preferences in the treatment of depressive disorder in primary care. *General Hospital Psychiatry* **26**, 184–189.

Williams JW Jr., Barrett J, Oxman T, Frank E, Katon W, Sullivan M, Cornell J, Sengupta A (2000). Treatment of dysthymia and minor depression in primary care: a randomized controlled trial in older adults. *Journal of the American Medical Association* **284**, 1519–1526.

https://doi.org/10.1017/S0033291713000457 Published online by Cambridge University Press