Short-term psychotherapy for depression

Driessen, E.

2013

Part II

Reviews and meta-analyses

2

Does pretreatment severity moderate the efficacy of psychological treatment of adult outpatient depression? A meta-analysis

This chapter is published as:
Driessen, E., Cuijpers, P., Hollon, S. D., & Dekker, J. J. M. (2010).
Does pretreatment severity moderate the efficacy of psychological treatment of adult outpatient depression? A meta-analysis.
Journal of Consulting and Clinical Psychology, 78, 668-680.

Abstract

Objective: It is widely believed that psychological treatment has little effect on more severely depressed patients. This study assessed whether pretreatment severity moderates psychological treatment outcome relative to controls by means of meta-analyses.

Method: We included 132 studies (10,134 participants) from a database of studies (www.evidencebasedpsychotherapies.org) in which the effects of psychological treatment on adult outpatients with a depressive disorder or an elevated level of depressive symptoms were compared with a control condition in a randomized controlled trial. Two raters independently extracted outcome data and rated study characteristics. We conducted metaregression analyses assessing whether mean pretreatment depression scores predicted psychological treatment versus control condition posttreatment effect size and subgroup analyses summarizing the results of studies reporting within-study analyses of depression severity and psychological treatment outcome.

Results: Psychological treatment was found to be consistently superior to control conditions (*d*=0.40–0.88). We found no indication that pretreatment mean depression scores predicted psychological treatment versus control condition posttreatment effect size, even after adjusting for relevant study characteristics. However, among the smaller subset of studies that reported within-study severity analyses, posttreatment effect sizes were higher for high-severity patients (*d*=0.63) than for low-severity patients (*d*=0.22) when psychological treatment was efficacious relative to a more stringent control.

Conclusion: Contrary to conventional wisdom, our findings suggest that when compared with control conditions, psychological treatment might be more efficacious for high-severity than for low-severity patients. Because the number of studies reporting within-study severity analyses is small, we recommend that future studies routinely report tests for Severity x Treatment interactions.

Introduction

It is widely believed that psychological treatment has little effect on more severely depressed patients. For example, treatment guidelines published by the American Psychiatric Association (2000) suggest that although psychotherapy may be sufficient for patients with less severe depressions, antidepressant medications or electroconvulsive therapy (ECT) is necessary for more severely depressed patients. This belief is based, in part, on the findings of the Treatment of Depression Collaborative Research Program, in which cognitive behavior therapy (CBT) did not separate from placebo in more severely depressed patients but antidepressant medications did (Elkin et al., 1989, 1995). However, what is often overlooked in that study is that interpersonal psychotherapy did separate from placebo among patients with more severe depression, suggesting that psychological treatment can be efficacious in this population.

There are two distinct questions that are often confounded in the literature. The first is whether psychological treatment works, and the second is how it compares with other alternative interventions. Whether psychological treatment works is answered by comparisons with minimal treatment controls (absolute efficacy) or nonspecific controls (specificity), whereas its relative efficacy is determined by comparisons with alternative interventions. In this article, we address the question of whether psychological treatment works, and more specifically, whether the magnitude of this effect (if any) is affected by depression severity. This is a question of moderation that asks whether psychological treatment is more or less likely to differentiate from control conditions as a function of pretreatment severity (Kraemer, Wilson, Fairburn, & Agras, 2002).

This study assessed whether pretreatment depression severity is related to psychological treatment outcome compared with control conditions by means of two different meta-analytic techniques: (a) metaregression analyses to assess whether mean pretreatment depression scores predict psychological treatment versus control condition posttreatment effect sizes and (b) analyses of studies reporting results of within-study analyses of depression severity and psychological treatment outcome.

Methods

Search strategy

We retrieved studies assessing the efficacy of psychological treatment for depression compared with control conditions from a database of 1,036 studies on the psychological treatment of depression. This database has been described in detail elsewhere (Cuijpers, van Straten, Warmerdam, & Andersson, 2008) and is publically available for all interested investigators (www .evidencebasedpsychotherapies.org). So far, 24 meta-analyses have been published from this database by our research group, 16 of which used at least a subset of the 132 studies that were included in this article. Pretreatment depression severity was examined as a possible predictor of treatment

outcome in five previous meta-analyses (Cuijpers, Dekker, Hollon, & Andersson, 2009; Cuijpers, van Straten, Smit, & Andersson, 2009; Cuijpers, van Straten, van Oppen, & Andersson, 2008; Cuijpers, van Straten, Warmerdam, & Smits, 2008; Driessen et al., 2010), but in none of these studies was it addressed explicitly as a primary focus, nor was it addressed in depth by using multiple outcome measures and different meta-analytic strategies.

The database was developed through a comprehensive literature search (from 1966 to January 2009) in which 9,011 abstracts were examined: 1,629 from PubMed, 2,439 from PsycINFO, 2,606 from Embase, and 2,337 from the Cochrane Central Register of Controlled Trials. These abstracts were identified by combining terms indicative of psychological treatment (*psychotherapy, psychological treatment, cognitive therapy, behavior therapy, interpersonal therapy, counseling, family therapy, marital therapy, problem-solving therapy, psychoanalysis, psychodynamic therapy, relaxation, reminiscence, life review*) and depression (both medical subject heading terms and text words). Studies from 42 meta-analyses regarding the psychological treatment of depression were also collected. In addition, the references of included studies were checked.

Selection of studies

From this database, we included studies in which the effects of a psychological treatment on adults with a depressive disorder or an elevated level of depressive symptoms were compared with the effects of a control condition in a randomized controlled trial.

Psychological treatments were defined as interventions in which verbal communication between a therapist and a client was the core element or as interventions in which a psychological treatment was written down in book format (bibliotherapy), which the client worked through with some kind of personal support from a therapist (by telephone, e-mail, or otherwise). We excluded studies on children and adolescents below 18 years of age. Studies in which the psychological intervention could not be discerned from other elements of the intervention were also excluded (e.g., managed care interventions and disease management programs), as were studies of inpatients and studies in which a standardized effect size could not be calculated (mostly because no test was performed in which the difference between the psychological treatment and control condition was examined). In some studies, a combination of psychological treatment and placebo was compared with placebo. These studies were excluded because a placebo may have an effect on depression or may alter the effects of psychological treatment (Wampold, Minami, Tierney, Baskin, & Bhati, 2005). We also excluded studies aimed at maintenance treatments and relapse prevention and studies that included participants who were anxious but not currently depressed. Comorbid general medical or psychiatric disorders were not used as an exclusion criterion. No language restrictions were applied.

Meta-analyses

This study assessed whether pretreatment depression severity is related to psychological treatment outcome when compared with a control condition. First, we conducted metaregression analyses to assess whether the mean pretreatment depression scores of the study's population predicted psychological treatment versus control condition posttreatment effect size. Although the mean pretreatment depression score is a straightforward measure of study population depression severity, it has limitations when used in these meta-analyses. Because it is a mean score, it ignores the range of depression severity in the study population. If the pretreatment mean depression score of the study population falls within the moderate category, for instance, it remains uncertain whether every participant was in this moderate range of scores or whether some were actually in the severe range and others were in the mild range, leading to a moderate mean score. To ensure that reliance on pretreatment study means did not obscure within-study variation in depression severity, we selected studies that reported results of moderation analyses within the context of a given study. These studies all examined whether the magnitude of the psychological treatment effect relative to control was a function of pretreatment severity. We then conducted a meta-analysis contrasting high- and low-severity groups of participants within these studies. These analyses ensured that participants were sorted in to high- or low-severity groups in a way that cannot be guaranteed when relying on pretreatment means. This constituted the second meta-analytic approach to our question.

Psychological treatment versus control condition posttreatment effect size

For each study, the comparative effect size of psychological treatment with control conditions at posttreatment was calculated by subtracting the average score of the control condition from the average score of the psychological treatment condition and dividing the result by the pooled standard deviations of both conditions. Effect sizes of 0 through 0.32 are assumed to be small, whereas effect sizes of 0.33 through 0.55 are considered moderate, and effect sizes of 0.56 through 1.20 are large (Lipsey & Wilson, 1993).

We used depressive symptoms as the sole outcome measure for this meta-analysis. Three outcome measures were used in the calculation of effect sizes: the Beck Depression Inventory (BDI-I; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), the revised Beck Depression Inventory (2nd ed.; BDI-II; Beck, Steer, & Brown, 1996), and the Hamilton Depression Rating Scale (HDRS; Hamilton, 1960, 1967). However, different HDRS variants exist. Because the number of items affects the mean total HDRS score, we included studies referring to the original Hamilton (1960, 1967) publications or explicitly mentioning the use of the 17-item HDRS. We excluded studies that used HDRS variants with other than 17 items. When multiple outcome measures were used in a given study, posttreatment effect sizes were calculated for all instruments reported.

When means and standard deviations were not reported, we used other statistics (e.g., *t* values, *p* values) to compute the effect sizes. When means and standard

deviations were not present and no statistical test between the relevant scores was presented, the effect size could not be calculated, and the study was excluded from the meta-analysis. If the treatment conditions included different subgroups (e.g., typical and atypical depressed participants) a single mean effect size was computed.

To calculate the pooled mean effect sizes, we used the computer program Comprehensive Meta-Analysis (Version 2.2.021; Biostat, Englewood, NJ). Because considerable heterogeneity was expected among the included studies, we used the random effects model to compute the pooled mean effect sizes. In the random effects model, the included studies are seen as a sample drawn from a population of studies rather than as replications of each other, so that not only the random error within the studies but also the true variations of effect sizes from one study to the next are taken into account. Consequently, the random effects model results in broader 95% confidence intervals and more conservative results.

As an indicator of homogeneity, we calculated the *Q* statistic. A significant *Q* value rejects the null hypothesis of homogeneity. We also calculated the I^2 statistic, which is an indicator of heterogeneity in percentages. A value of 0% indicates no observed heterogeneity, and larger values show increasing heterogeneity, with 25% indicating low, 50% indicating moderate, and 75% indicating high heterogeneity (Higgins, Thompson, Deeks, & Altman, 2003).

Simple and multiple metaregression analyses

First, we conducted simple metaregression analyses to assess whether pretreatment mean depression scores alone predicted posttreatment effect size. We then conducted multiple metaregression analyses controlling for the following commonly reported variables to see if they affected psychological treatment versus control condition effect sizes:

- Recruitment method: community (recruiting participants from the general community through local media announcements or flyers, with participants taking the initiative to participate in the study), clinical (recruiting participants from general practice populations or outpatient samples who first actively sought help for depression and were then asked to participate in the study), or other (e.g., systematic screening, recruiting participants from hospital populations, a combination of methods, or no recruitment method reported);
- Depression diagnosis: major depressive disorder (MDD according to criteria from the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM–IV*; American Psychiatric Association, 1994), mood disorder (MDD according to non-*DSM–IV* criteria or mood disorder according to specified diagnostic criteria), or other (typically a high score on a standardized depression measure);
- Target group: adults, older adults, student population, women with postpartum depression, people with general medical disorder, or other;
- Mean age of the participants in the psychological treatment condition;
- Percentage of women in the psychological treatment condition;
- Psychological treatment format: individual, group, or bibliotherapy;

- Number of sessions in the psychological treatment condition;
- Type of control condition: waiting list, care as usual, or other (e.g., nonspecific or pill placebo);
- Use of a treatment manual: yes or no (no manual used or no manual reported);
- Therapist training: yes (therapists already trained for the specific intervention or receiving training for the study) or no (therapists not trained or no training reported);
- Treatment integrity check: yes (integrity check by means of supervision of the therapists during treatment or the recording of treatment sessions) or no (no integrity check used or no check reported);
- Treatment condition unknown to the outcome assessor: yes or no (treatment condition known to assessors or not reported);
- Independent randomization: yes (randomization was conducted by an independent party) or no (randomization was not conducted by an independent party or randomization method was not reported);
- Outcome analyses: intention-to-treat analyses or completers-only analyses;
- Type of psychological treatment: CBT according to Beck's manual (Beck, Rush, Shaw, & Emery, 1979), CBT according to another manual, interpersonal psychotherapy, problem-solving therapy, nondirective supportive therapy, behavioral activation therapy, or other. More specific definitions of these interventions are described elsewhere (Cuijpers, van Straten, Andersson, & van Oppen, 2008).

Two raters independently rated the study characteristics and disagreements were resolved in consensus.

We used Stata/SE 8.2 to conduct metaregression analyses and used the random-effects model for these analyses. Because scores on the BDI-II trend slightly higher than scores on the BDI-I (Beck et al., 1996), we kept the two measures separate in the primary analyses but also conducted secondary analyses in which the two were combined both as raw and as standardized scores. To avoid collinearity among the predictors in the regression model, we calculated the correlations between all variables previously described, using SPSS Version 16. When the correlation between two variables was found to be greater than .55, one of the variables was excluded from the analyses.

We defined a reference group within each category of variables (Table 1; e.g., among the recruitment methods, we defined community recruitment as the reference group; among the three psychological treatment formats, the individual format was defined as the reference group). When a variable is found to significantly predict posttreatment effect size in a multiple metaregression analysis, that means it predicts posttreatment effect size after adjusting for all other variables entered in the model. For continuous variables, such as pretreatment BDI score, a one-unit increase in BDI score is associated with an increase of the base rate by one unit of regression coefficient *B*. For categorical variables, the base rate plus *B* of a given category constitutes the effect size difference of that category with the reference group.

Studies reporting outcomes for low and high baseline severity.
Second, to ensure that reliance on pretreatment study means did not obscure within-study variation in depression severity, we screened the studies for those reporting results of moderation analyses of depression severity and psychological treatment outcome within the same sample. We selected studies reporting on different subgroups, including participants with high-severity and low-severity baseline depression only. We then conducted subgroup analyses. In subgroup analyses, studies are divided into two or more subgroups. For each subgroup, the pooled mean effect size is calculated, and a test is conducted to examine whether the subgroups' effect sizes differ significantly from one another. We calculated the pooled mean effect size comparing psychological treatment with control conditions at posttreatment in the subgroups including high-severity participants only and in the subgroups including low-severity participants only and assessed whether the effect sizes differed significantly from each other.

Subgroup analyses were conducted according to the procedures implemented in Comprehensive Meta-Analysis Version 2.2.021. We used the mixed effects method of subgroup analyses, which pools studies within subgroups with the random effects model but tests for significant differences between the subgroups with the fixed effects model. We considered groups of study participants as high or low severity if all participants in the subgroups scored above or below a cutoff score on a standardized depression measure or met diagnostic criteria that differed with respect to severity level (e.g., subclinical depression vs. MDD). We classified the high- or low-severity categories according to the cutoff criteria used in the original studies.

Results

Description of the included studies

A total of 132 studies (198 comparisons) were included, totaling 10,134 participants (5,858 in the psychological treatment conditions and 4,276 in the control conditions). The majority of the studies examined CBT (101 comparisons), whereas smaller numbers of studies examined interpersonal psychotherapy, problem-solving therapy, nondirective supportive therapy, or behavioral activation (14, 15, 154, and 12 comparisons, respectively). In the remaining 41 comparisons, a variety of other psychological interventions were examined. In 105 comparisons, the study used specified diagnostic criteria to define depression (58 MDD according to *DSM–IV* criteria and 47 MDD according to non-*DSM–IV* criteria or other diagnosed mood disorders), whereas in 93 comparisons the study defined depression in other ways, usually by an elevated score on a self-report measure.

The effect sizes comparing psychological treatment with control conditions at posttreatment are summarized in Table 1. In 74 studies (121 comparisons), the BDI-I was used as an outcome measure. The pooled mean effect size in this group of studies was 0.80 (95% CI [0.69, 0.91]). In 12 studies (14 comparisons), the BDI-II was used as an outcome measure. The posttreatment effect size relative to controls was smaller in

this subgroup of studies (*d*=0.40; 95% CI [0.16, 0.65]). However, BDI type was not a significant predictor of posttreatment effect size in the metaregression analyses, suggesting that the smaller mean pooled effect size in studies using the BDI-II as an outcome measure was more a consequence of differences in study characteristics than of differences in the scales. In 48 studies (70 comparisons), the HDRS was used as an outcome measure. The pooled mean effect size in this group of studies was 0.88 (95% CI [0.74, 1.01]). Heterogeneity was moderate with all three outcome measures (*Q*=35.88–372.13, *p*<.01; I^2=63.77–67.75%). A similar pattern of effect sizes resulted from these analyses when outliers (*d*>2.0) were excluded or when only the highest effect sizes or only the lowest effect sizes in studies with multiple comparisons were used (Table 1). Thus, psychological treatment was found to be consistently superior to control conditions.

Table 1: Psychological treatment versus control condition posttreatment effect sizes

Posttreatment effect size	*N*	*d*	95% CI	*Z*	*Q*	I^2
BDI-I						
All comparisons	121	0.80	0.69 - 0.91	14.65**	372.13**	67.75
Highest ES per study only	74	0.86	0.72 - 0.99	12.21**	249.26**	70.71
Lowest ES per study only	74	0.66	0.54 - 0.79	10.73**	199.29**	63.37
Outliers (*d*>2.0) excluded[a]	112	0.71	0.61 - 0.81	14.21**	279.56**	60.29
BDI-II						
All comparisons	14	0.40	0.16 - 0.65	3.23**	35.88**	63.77
Highest ES per study only	12	0.38	0.11 - 0.64	2.80*	30.11**	63.47
Lowest ES per study only	12	0.36	0.10 - 0.62	2.70*	29.52**	62.74
HDRS						
All comparisons	70	0.88	0.74 - 1.01	13.11**	191.85**	64.03
Highest ES per study only	48	0.92	0.76 - 1.08	11.44**	147.80**	68.20
Lowest ES per study only	48	0.75	0.61 - 0.90	10.51**	118.92**	60.48
Outliers (*d*>2.0) excluded[b]	64	0.78	0.66 - 0.90	12.61**	144.95**	56.54

Note. BDI-I = Beck Depression Inventory; BDI-II = Beck Depression Inventory (2nd ed.); ES = effect size; HDRS = Hamilton Depression Rating Scale.

[a] Ayen and Hautzinger (2004); Carrington (1979); Nezu and Perri (1989); Pecheur (1980); Taylor and Marshall (1977); Wilson, Goldin, and Charbonneau- Powis (1983).

[b] Arean et al. (1993); Cullen (2002); Floyd, Scogin, McKendree-Smith, Floyd, and Rokke (2004); Nezu and Perri (1989); Pecheur (1980); Selmi, Klein, Greist, Sorrell, and Erdman (1990).

* p<.05. ** p<.01. Italic numbers indicate a nonsignificant trend (p<.10).

Simple metaregression analyses

We conducted simple metaregression analyses to assess whether mean pretreatment depression scores in the psychological treatment condition predicted psychological treatment versus control condition posttreatment effect size. Mean pretreatment BDI-I scores (range=14.11–35.85) and mean pretreatment BDI-II scores (range=17.89 – 36.50) did not predict posttreatment effect size significantly (BDI-I: *B*=-0.016; 95% CI [-0.047, 0.015]; *p*=.30; BDI-II: *B*=-0.001; 95% CI [-0.052, 0.054]; *p*=.97). Mean pretreatment HDRS scores also did not predict posttreatment effect size significantly (range=8.34 –31.00; *B*=0.016; 95% CI [-0.023, 0.054]; *p*=.43). The ranges of the mean pretreatment depression scores represented a broad distribution from mild to severe depressions and should support efforts to look for correlations.

To address the extent to which these nonsignificant findings might be the result of missed studies, we examined how many extra studies would have been needed to find pretreatment depression severity significantly predicting posttreatment effect size. We simulated larger samples of studies by copying the existing data sets, thereby creating extra studies that were in all ways similar to the studies already included in the meta-analyses. We then reran the simple metaregression analyses to examine whether pretreatment depression severity predicted posttreatment effect size significantly (*p*<.05) in these larger samples. We had to multiply the data set of studies using the BDI-I as outcome measure four times (thus adding 322 studies) to find pretreatment BDI-I scores predicting posttreatment effect size significantly (*B*=-0.016, *p*=.04). Similarly, we had to multiply the data set of studies using the HDRS as outcome measure five times (thus adding 230 studies) to find pretreatment mean HDRS scores predicting posttreatment effect size significantly (*B*=0.015, *p*=.05). BDI-II scores were not found to be significantly predictive even after multiplying the data set of studies using this outcome measure 1,000 times (thus adding 12,000 studies; *B*=-0.0005, *p*=.47). Given the comprehensive search strategy, we consider it unlikely that more than 230 studies would have been missed.

Multiple metaregression analyses

In addition to the simple metaregression analyses treating pretreatment severity as the only predictor, we also conducted multiple metaregression analyses to control for additional variables that might influence posttreatment effect size. With regard to collinearity, a number of variables were found to have correlations higher than .55. In the subgroup of studies using the BDI-I as the outcome measure, mean age in the psychological treatment condition correlated with both older adults (*r*=.58, *p*<.01) and students (*r*=-.57, *p*<.01) as target populations. In addition, training of the therapists correlated with the use of an integrity check in the study (*r*=.62, *p*<.01). We decided to exclude mean age as a predictor in the model and to do separate analyses excluding either training or integrity. In the subgroup of studies using the HDRS as outcome measure, we again found that mean age correlated with older adults as the target population (*r*=.83, *p*<.01). In addition, the percentage of women correlated with other populations as target group (*r*=.57, *p*<.01). We therefore excluded both mean age and percentage of women as predictors from the model.

The results of the multiple metaregression analyses are presented in Table 2. As can be seen, neither the mean pretreatment BDI-I nor the mean pretreatment HDRS scores predicted posttreatment effect size when controlling for other variables, regardless of the model tested. The results of the metaregression analyses suggest that other factors than pretreatment depression severity did predict posttreatment BDI and HDRS effect sizes. For instance, studies using other recruitment methods yielded lower effect sizes than studies using community recruitment, and effect sizes generally were higher in studies with student populations when compared with adult populations. Thus, although other factors were found to be significant predictors, mean pretreatment depression score was not a significant predictor of BDI and HDRS posttreatment effect sizes when controlling for relevant variables. Because of the small number of studies using the BDI-II, we could not conduct a separate metaregression analysis with this outcome measure. However, secondary analyses combining BDI-I and BDI-II scores (both raw and standardized within measure) resulted in a highly similar pattern of findings in which pretreatment severity did not predict posttreatment effect size, but other variables did. In addition, we replicated the BDI-I and HDRS continuous analyses in a categorical fashion and found a similar pattern of results.

One might argue that pooling different subgroups of studies could obscure possible differential moderation relationships in specific subgroups. To address this possibility, we also conducted BDI-I and HDRS simple and multiple metaregression analyses in subgroups of studies. We assessed whether pretreatment depression severity predicted posttreatment effect size in subgroups of studies with different types of psychological treatments (CBT and other treatments), control groups (waiting list and other controls), treatment formats (individual and group), depression diagnosis (diagnostic criteria and other diagnosis), and target groups (adults and other target groups). We adjusted the number of predictors to account for the smaller samples of studies examined in these subgroups and used a simplified metaregression model including pretreatment mean depression scores and the variables defining the subgroups that were examined. The results of these analyses are presented in Table 3. In none of the subgroups of studies, metaregression analyses indicated that mean pretreatment BDI-I scores predicted posttreatment effect size significantly. Simple metaregression analyses did indicate that pretreatment HDRS scores predicted posttreatment effect size significantly in the subgroups of individual format, other diagnosis, and other target group, but this result was not replicated in the multiple metaregression analyses when adjusting for possible confounds in these three subgroups. In the subgroup of studies that used a group format, both the simple and multiple regression analyses indicated that pretreatment HDRS score predicted posttreatment effect size significantly (simple: B=0.0762, p<.01; multiple: B=0.0930, p<.01), such that higher pretreatment depression scores were associated with larger posttreatment effect sizes. However, this subgroup included only 10 studies, and the results were not replicated in the larger set of studies that used the BDI-I as the outcome measure.

Table 2: Predictors of BDI-I and HDRS psychological treatment versus control condition posttreatment effect sizes: multiple metaregression analyses

Predictor	BDI-I excl. Integrity			BDI-I excl. Training			HDRS		
	B	*SE*		*B*	*SE*		*B*	*SE*	
Base rate	0.848	0.755		0.591	0.699		3.557	0.756	**
Mean pretreatment BDI-I/HDRS score	-0.029	0.018		-0.023	0.017		-0.031	0.022	
Recruitment									
Community[a]									
Clinical	-0.020	0.180		-0.049	0.168		0.347	0.172	*
Other	-0.687	0.264	**	-0.618	0.246	*	-1.457	0.519	**
Diagnosis									
MDD *DSM-IV*[a]									
Mood disorder	0.037	0.187		0.067	0.171		0.351	0.358	
Other	-0.127	0.165		-0.045	0.155		0.301	0.362	
Target group									
Adults[a]									
Older adults	-0.515	0.357		-0.516	0.336		0.231	0.331	
Student population	0.626	0.251	*	0.573	0.239	*	-	-	
Women with postpartum depression	0.250	0.319		0.217	0.298		0.611	0.598	
People with a general medical disorder	0.565	0.393		0.435	0.362		-0.725	0.659	
Other	0.675	0.266	*	0.692	0.246	**	-0.957	0.491	
Mean age	-	-		-	-		-	-	
Percentage women	0.011	0.006		0.012	0.006	*	-	-	
Format									
Individual[a]									
Group	-0.130	0.146		-0.173	0.136		0.526	0.259	*
Bibliotherapy	0.114	0.293		-0.065	0.270		-0.413	0.371	
Number of sessions	0.006	0.017		0.017	0.016		-0.035	0.020	

Control group								
Waiting list[a]								
Care as usual	-0.051	0.215	-0.072	0.198		-0.502	0.265	
Other	-0.320	0.224	-0.348	0.208		-0.428	0.227	
Manual	0.219	0.160	0.294	0.150	*	-1.391	0.451	**
Training	0.054	0.185	-	-		-0.135	0.349	
Integrity	-	-	-0.390	0.141	**	0.395	0.338	
Treatment condition unknown to outcome assessors	-0.225	0.217	-0.174	0.207		-1.297	0.457	**
Independent randomization	-0.092	0.241	-0.083	0.219		-0.210	0.245	
Outcome analyses	-0.249	0.190	-0.201	0.177		0.739	0.270	**
Psychological treatment type								
Cognitive behavioral therapy (Beck) [a]								
Cognitive behavioral therapy (other)	-0.039	0.207	0.068	0.190		0.450	0.360	
Interpersonal psychotherapy	-0.009	0.306	0.090	0.282		-0.062	0.226	
Problem-solving therapy	0.265	0.280	0.387	0.258		1.350	0.454	**
Nondirective supportive therapy	-0.220	0.257	-0.161	0.240		-0.348	0.645	
Behavioral activation therapy	0.329	0.362	0.305	0.342		0.081	0.226	
Other	-0.200	0.241	0.006	0.220		0.114	0.391	

Note. Dashes indicate that B and SE could not be calculated for this predictor, either because the predictor was excluded from the model due to collinearity or because not enough studies were included in that category. BDI-I = Beck Depression Inventory; excl. = excluding the following predictor from the model; HDRS = Hamilton Depression Rating Scale; MDD DSM–IV = major depressive disorder according to criteria from the Diagnostic and Statistical Manual of Mental Disorders (4th ed.).

[a] No Bs and SEs were calculated for this predictor, because it constituted the reference group.

* p<.05. ** p<.01.

Table 3: Mean pretreatment BDI-I/HDRS score as a predictor of posttreatment effect size within different subgroups of studies: simple and multiple metaregression analyses

Subgroup	BDI-I simple			BDI-I multiple			HDRS simple			HDRS multiple		
	N	*B*	*p*	*N*	*B*	*p*	*N*	*B*	*p*	*N*	*B*	*p*
Psychological treatment type												
CBT	61	-0.0231	.26	51	-0.0101	.62	24	-0.0073	.77	24	-0.0020	.94
Other therapies	43	-0.0059	.81	42	-0.0121	.67	22	0.0140	.64	22	0.0095	.77
Control group												
Waiting list	74	-0.0254	.16	64	-0.0074	.76	25	0.0255	.33	25	0.0145	.60
Other control groups	30	0.0063	.79	29	0.0199	.42	21	-0.0079	.61	20	-0.0331	.10
Format												
Individual	49	-0.1820	.44	49	0.0293	.22	31	-0.0482	.02	31	-0.0307	.09
Group	44	-0.0166	.52	44	-0.0222	.37	10	0.0762	<.01	10	0.0930	<.01
Depression diagnosis												
Diagnostic criteria	62	-0.0383	.14	56	-0.2645	.37	33	-0.0018	.95	28	-0.0232	.16
Other (elevated depression score)	42	0.0129	.53	37	0.0153	.42	13	0.4860	.04	13	0.0064	.84
Target group												
Adults	68	-0.0213	.29	58	-0.0198	.35	33	-0.4486	.09	29	-0.0313	.14
Other (specific target groups)	36	0.0171	.51	35	0.0052	.84	13	0.0628	<.01	12	0.0478	.14

Note. BDI-I = Beck Depression Inventory; HDRS = Hamilton Depression Rating Scale; CBT = cognitive behavior therapy.

Studies reporting outcomes for within-study severity analyses

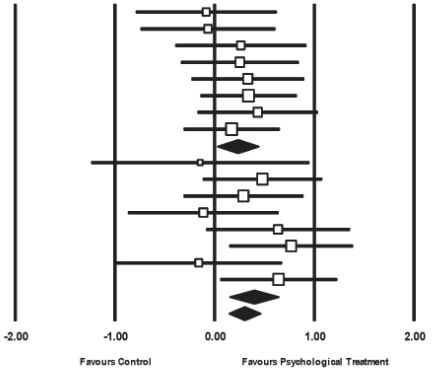
Of the 132 studies we included, 16 studies looked at the relation between pretreatment depression severity and subsequent outcome. Five of those studies provided prognostic information only; that is, information that predicted outcome irrespective of the treatment (Brown & Lewinsohn, 1984; McLean & Hakstian, 1979; Rohen, 2002; Teri, Logsdon, Uomoto, & McCurry, 1997; Wright et al., 2005). Such information can be used to predict future status but does not speak to the issue of differential treatment effects (Fournier et al., 2009). On the other hand, prescriptive information (in the methodological literature this is often referred to as a *moderator*; Kraemer et al., 2002) can detect a different pattern of outcomes between different treatment conditions. Another three studies reported insufficient information to determine whether these analyses were prognostic or prescriptive (Dunn et al., 2007; Holden, Sagovsky, & Cox, 1989; McKendree-Smith, 1998), and two prescriptive studies did not provide the requisite data for reanalysis (Murphy, Carney, Knesevich, Wetsel, & Whitworth, 1995; Usaf & Kavanagh, 1990).

Six studies reported the information necessary to calculate posttreatment effect sizes comparing psychological treatment with control conditions in subgroups including high-severity participants only and in subgroups including low-severity participants only (Barlow, 1986; Dimidjian et al., 2006; Elkin et al., 1989; Haringsma, Engels, Cuijpers, & Spinhoven, 2006; Simpson, Corney, Fitzgerald, & Beecham, 2003; Van Schaik et al., 2006). We used these studies to quantitatively summarize the results of these within-study severity analyses (Figure 1). Combining the psychological treatment outcome data of all low-severity within-study subgroups resulted in a pooled mean effect size of 0.23 (95% CI [0.02, 0.43]; Z=2.18; p=.03). Combining the outcome data of all high-severity within-study subgroups resulted in a posttreatment effect size of 0.39 (95% CI [0.15, 0.64]; Z=3.16; p=.00). Although the psychological treatment condition differed significantly from the control condition in both subgroups, the two subgroups did not differ significantly from each other (p=.31). Heterogeneity was low in both subgroups (low severity: Q=2.32, *ns*; I^2=0.00; high severity: Q=7.04, *ns*; I^2=0.57).

However, prescriptive effects can be found only in studies that actually found a difference between psychological treatment and control conditions. Pretreatment severity cannot moderate treatment outcome if the treatment itself is not efficacious for at least some subset of the sample. Therefore, only those studies that found a treatment effect for one or more of the subgroups are germane to the question of moderation. Closer examination of the six studies revealed that some found psychological treatment more efficacious than the control condition (the behavioral activation condition in Dimidjian et al., 2006; the interpersonal therapy condition in Elkin et al., 1989; Haringsma et al., 2006; Van Schaik et al., 2006), whereas others did not (Barlow, 1986; the cognitive therapy condition in Dimidjian et al., 2006; the CBT condition in Elkin et al., 1989; Simpson et al., 2003). Comparing high-severity subgroups (d=0.04; 95% CI [-0.35, 0.42]; Z=0.19; p=.85) with low-severity subgroups (d=0.24; 95% CI [-0.08, 0.56]; Z=1.46; p=.14) in the psychological treatment conditions that were not found to be efficacious relative to controls resulted in no significant

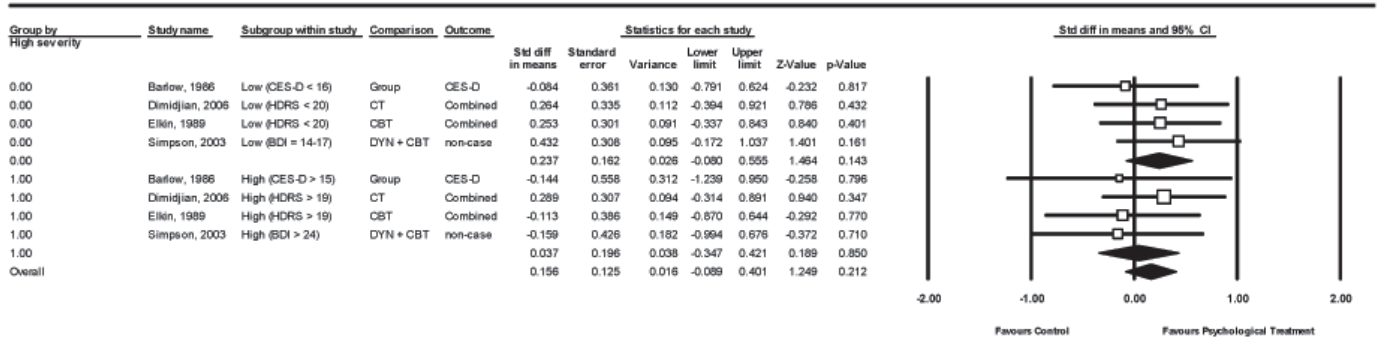
Figure 1: Effect sizes of low- and high-severity subgroups in studies reporting within-study moderation analyses of depression severity.

Group by High severity	Study name	Subgroup within study	Comparison	Outcome	Statistics for each study: Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
0.00	Barlow, 1986	Low (CES-D < 16)	Group	CES-D	-0.084	0.361	0.130	-0.791	0.624	-0.232	0.817
0.00	Dimidjian, 2006	Low (HDRS < 20)	BA	Combined	-0.066	0.345	0.119	-0.743	0.611	-0.190	0.849
0.00	Dimidjian, 2006	Low (HDRS < 20)	CT	Combined	0.264	0.335	0.112	-0.394	0.921	0.786	0.432
0.00	Elkin, 1989	Low (HDRS < 20)	CBT	Combined	0.253	0.301	0.091	-0.337	0.843	0.840	0.401
0.00	Elkin, 1989	Low (HDRS < 20)	IPT	Combined	0.333	0.289	0.084	-0.234	0.900	1.153	0.249
0.00	Haringsma, 2005	Low (Subclinical depr)	CWD	Combined	0.340	0.247	0.061	-0.144	0.823	1.377	0.169
0.00	Simpson, 2003	Low (BDI = 14-17)	DYN + CBT	non-case	0.432	0.308	0.095	-0.172	1.037	1.401	0.161
0.00	Van Schaik, 2006	Low (MADRS < 21)	IPT	non-case	0.171	0.246	0.061	-0.311	0.654	0.696	0.486
0.00					0.229	0.105	0.011	0.023	0.434	2.184	0.029
1.00	Barlow, 1986	High (CES-D > 15)	Group	CES-D	-0.144	0.558	0.312	-1.239	0.950	-0.258	0.796
1.00	Dimidjian, 2006	High (HDRS > 19)	BA	Combined	0.481	0.306	0.094	-0.119	1.080	1.571	0.116
1.00	Dimidjian, 2006	High (HDRS > 19)	CT	Combined	0.289	0.307	0.094	-0.314	0.891	0.940	0.347
1.00	Elkin, 1989	High (HDRS > 19)	CBT	Combined	-0.113	0.386	0.149	-0.870	0.644	-0.292	0.770
1.00	Elkin, 1989	High (HDRS > 19)	IPT	Combined	0.636	0.369	0.136	-0.087	1.359	1.724	0.085
1.00	Haringsma, 2005	High (Major depression)	CWD	Combined	0.768	0.317	0.100	0.147	1.388	2.424	0.015
1.00	Simpson, 2003	High (BDI > 24)	DYN + CBT	non-case	-0.159	0.426	0.182	-0.994	0.676	-0.372	0.710
1.00	Van Schaik, 2006	High (MADRS >20)	IPT	non-case	0.642	0.299	0.090	0.055	1.229	2.145	0.032
1.00					0.392	0.124	0.015	0.149	0.636	3.157	0.002
Overall					0.296	0.080	0.006	0.140	0.453	3.704	0.000



CES-D = Center for Epidemiological Studies Depression Scale; HDRS = Hamilton Depression Rating Scale; BDI-I = Beck Depression Inventory; MADRS = Montgomery Åsberg Depression Rating Scale; depr = depression; BA = behavioral activation; CT = cognitive therapy; CBT = cognitive behavior therapy; IPT = interpersonal psychotherapy; CWD = Coping With Depression course; DYN = psychodynamic therapy; Std diff = standard difference.

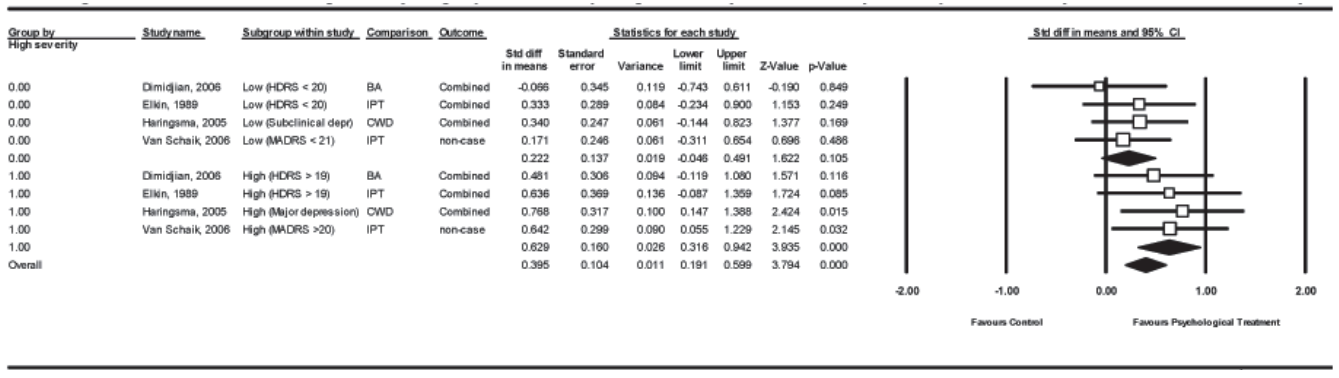
Figure 2: Effect sizes of low- and high-severity subgroups in studies reporting within-study moderation analyses of depression severity, nonefficacious interventions only.



Group by High severity	Study name	Subgroup within study	Comparison	Outcome	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
0.00	Barlow, 1986	Low (CES-D < 16)	Group	CES-D	-0.084	0.361	0.130	-0.791	0.624	-0.232	0.817
0.00	Dimidjian, 2006	Low (HDRS < 20)	CT	Combined	0.264	0.335	0.112	-0.394	0.921	0.786	0.432
0.00	Elkin, 1989	Low (HDRS < 20)	CBT	Combined	0.253	0.301	0.091	-0.337	0.843	0.840	0.401
0.00	Simpson, 2003	Low (BDI = 14-17)	DYN + CBT	non-case	0.432	0.308	0.095	-0.172	1.037	1.401	0.161
0.00					0.237	0.162	0.026	-0.080	0.555	1.464	0.143
1.00	Barlow, 1986	High (CES-D > 15)	Group	CES-D	-0.144	0.558	0.312	-1.239	0.950	-0.258	0.796
1.00	Dimidjian, 2006	High (HDRS > 19)	CT	Combined	0.289	0.307	0.094	-0.314	0.891	0.940	0.347
1.00	Elkin, 1989	High (HDRS > 19)	CBT	Combined	-0.113	0.386	0.149	-0.870	0.644	-0.292	0.770
1.00	Simpson, 2003	High (BDI > 24)	DYN + CBT	non-case	-0.159	0.426	0.182	-0.994	0.676	-0.372	0.710
1.00					0.037	0.196	0.038	-0.347	0.421	0.189	0.850
Overall					0.156	0.125	0.016	-0.089	0.401	1.249	0.212

CES-D = Center for Epidemiological Studies Depression Scale; HDRS = Hamilton Depression Rating Scale; BDI-I = Beck Depression Inventory; CT = cognitive therapy; CBT = cognitive behavior therapy; DYN = psychodynamic therapy; Std diff = standard difference.

Figure 3: Effect sizes of low- and high-severity subgroups in studies reporting within-study moderation analyses of depression severity, efficacious interventions only.



Group by High severity	Study name	Subgroup within study	Comparison	Outcome	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
0.00	Dimidjian, 2006	Low (HDRS < 20)	BA	Combined	-0.066	0.345	0.119	-0.743	0.611	-0.190	0.849
0.00	Elkin, 1989	Low (HDRS < 20)	IPT	Combined	0.333	0.289	0.084	-0.234	0.900	1.153	0.249
0.00	Haringsma, 2005	Low (Subclinical depr)	CWD	Combined	0.340	0.247	0.061	-0.144	0.823	1.377	0.169
0.00	Van Schaik, 2006	Low (MADRS < 21)	IPT	non-case	0.171	0.246	0.061	-0.311	0.654	0.696	0.486
0.00					0.222	0.137	0.019	-0.046	0.491	1.622	0.105
1.00	Dimidjian, 2006	High (HDRS > 19)	BA	Combined	0.481	0.306	0.094	-0.119	1.080	1.571	0.116
1.00	Elkin, 1989	High (HDRS > 19)	IPT	Combined	0.636	0.369	0.136	-0.087	1.359	1.724	0.085
1.00	Haringsma, 2005	High (Major depression)	CWD	Combined	0.768	0.317	0.100	0.147	1.388	2.424	0.015
1.00	Van Schaik, 2006	High (MADRS >20)	IPT	non-case	0.642	0.299	0.090	0.055	1.229	2.145	0.032
1.00					0.629	0.160	0.026	0.316	0.942	3.935	0.000
Overall					0.395	0.104	0.011	0.191	0.599	3.794	0.000

HDRS = Hamilton Depression Rating Scale; depr = depression; MADRS = Montgomery Åsberg Depression Rating Scale; BA = behavioral activation; IPT = interpersonal psychotherapy; CWD = Coping With Depression course; Std diff = standard difference.

differences in effect size (p=.43; Figure 2), whereas comparing the high-severity and low-severity subgroups in the psychological treatment conditions that were found to be efficacious relative to controls did result in effect size differences. Here, the posttreatment effect size in the high-severity subgroups was higher (d=0.63; 95% CI [0.31, 0.94]; Z=3.94; p<.01) than the posttreatment effect size in the low-severity subgroups (d=0.22; 95% CI [-0.05, 0.49]; Z=1.62; p=.11) and the difference in effect size was significant (p=.05; Figure 3). Heterogeneity was low in all of the aforementioned subgroup analyses (Q=1.20, *ns*; I^2=0.00). Although the results of this analysis are based on a small number of comparisons, they suggest that when psychological treatment is efficacious, posttreatment effect sizes are higher for high-severity patients than for low-severity patients.

Discussion

The results of the metaregression analyses based on the total sample of studies provided no indication that posttreatment effect size was moderated by mean pretreatment depression levels. This finding was replicated with different outcome measures and did not appear to be the consequence of missing studies, nor did it change when controlling for relevant study characteristics. Furthermore, we found little evidence for the possibility that pooling different subgroups of studies in the total sample might obscure the finding of moderation relationships within certain subgroups, because pretreatment severity generally was not found to be a significant predictor of posttreatment effect size in smaller subsets of studies grouped according to type of treatment, control group, treatment format, depression diagnosis, and target group. The only exception to this general pattern of findings was that pretreatment mean HDRS score predicted posttreatment effect size in four smaller subsets of studies. However, in three of these subsets, this result was not replicated in the multiple metaregression analyses when adjusting for possible confounders, and none of the results were replicated in the larger sample of studies that used the BDI as outcome measure. We therefore think these findings should be interpreted with caution. In summary, using metaregression analyses, we found that mean pretreatment depression score was not a significant predictor of posttreatment effect size.

However, a nonsignificant predictor cannot be interpreted as evidence for the absence of a relationship between the predictor and the dependent variable (in this case, pretreatment mean depression score and posttreatment effect size, respectively), because by definition the null hypothesis of equal posttreatment effect sizes for high and low pretreatment depression levels can only be rejected and can never be proven. Moreover, the metaregression analyses based on mean pretreatment depression scores ignore within-study variability in initial severity that may moderate subsequent outcomes, and it was possible that by relying on meta-analyses based on pretreatment study means, we may not have detected evidence of moderation operating within a given sample. To address this possibility, we screened

for studies reporting on severity analyses within their population, so that we could summarize outcomes separately for subgroups of low-severity and high-severity participants. These comparisons are the ones most directly germane to our question, because they come closest to doing a mega-analysis on individual patient data.

What we found was that pretreatment levels of severity not only moderated the effects of psychotherapy but did so in a manner that was exactly the opposite of what is generally presumed in the field. In those studies that reported within-study severity analyses, posttreatment effect sizes not only differed as a function of pretreatment severity when psychotherapy was efficacious relative to controls but were actually higher for high-severity patients than for low-severity patients. This suggests not only that psychotherapy works (relative to controls) for patients with more severe depression but that it works to a greater extent for those patients than it does for patients with less severe depression. This finding appears to indicate that pretreatment severity moderates the efficacy of psychotherapy but that it does so in a direction opposite from the one that is often assumed (e.g., American Psychiatric Association, 2000).

Three out of the four studies included in the subgroup analysis that found a moderation effect compared efficacious treatments to stringent controls, such as pill placebo (Dimidjian et al., 2006; Elkin et al., 1989, 1995) or care as usual (Van Schaik et al., 2006). These studies differ from the modal study in this literature, the majority of which compared psychological treatment with waitlist control conditions. We think evidence of moderation is most likely to be found under exactly these conditions: when efficacious treatments are compared with stringent controls and the sample contains both more and less severely depressed patients. We speculate that nonspecific treatments may be sufficient for low-severity patients but that high-severity patients need treatments with specific effects beyond the simple provision of treatment in order to get fully well. This is in line with what has been found with regard to antidepressant medication, namely, that patients with less severe depressions are quite responsive to placebo and that true drug effects relative to placebo are likely to emerge only among patients with more severe depressions (Gelenberg et al., 2008). For example, a patient-level meta-analysis found that drug–placebo differences were negligible among patients with less severe depressions (scores of 20 or below on the HDRS) and that the drug did not begin to separate from the placebo until HDRS scores rose into the mid-20s (Fournier et al., 2010). Similar results in which drug-placebo differences increased as a function of initial severity have been found in meta-analyses based on sample means of data submitted to the U.S. Food and Drug Administration in support of drug approval (Khan, Leventhal, Khan, & Brown, 2002; Kirsch et al., 2008).

The clinical relevance of our findings might be better reflected in the effect size measure *number needed to treat*. This measure is defined as the number of patients one would have to treat to have one more successful outcome relative to the same number assigned to some control condition (Kraemer & Kupfer, 2006). Translating psychological treatment versus control group posttreatment effect sizes into number needed to treat, clinicians provide benefit for roughly every other patient they treat

compared with no treatment, regardless of severity. Moreover, translating effect sizes from the within-study severity analyses into number needed to treat suggests that clinicians need to treat eight less severely depressed patients with an efficacious psychological treatment to make a difference relative to more stringent nonspecific controls like pill placebos versus only three patients with more severe depression.

This study has several limitations. First, although we found a large number of studies meeting the inclusion criteria, not all of those studies could be included in the analyses, because of incomplete reporting of the requisite data. In addition, despite the comprehensive search strategy, the possibility remains that studies might have been missed. However, we consider it unlikely that we missed the number of studies necessary to substantially change the results of the simple metaregression analyses ($n \geq 230$). Second, the methodological limitations of the body of literature included in this study must be borne in mind when interpreting this study's results. For instance, it has been shown that the mean posttreatment effect size comparing psychological treatment with control conditions is likely to be overestimated because of publication bias (Cuijpers, Smit, Bohlmeijer, Hollon, & Andersson, 2010). Furthermore, the quality of the majority of studies included in this meta-analysis is suboptimal, with only a small number meeting the criteria for high-quality studies, and psychological treatment versus control condition posttreatment effect sizes have been shown to be significantly lower in high-quality than in low-quality studies (Cuijpers, van Straten, Bohlmeijer, Hollon, & Andersson, 2010). We did control for quality indicators in the multiple metaregression analyses, and rerunning the simple metaregression analyses in the subgroup of high-quality studies did not result in a different pattern of findings (although these latter analyses should be interpreted with caution because of the very small number of comparisons included [$n=7$]), but ideally the relationship between pretreatment severity and psychological treatment outcome would be assessed in a large number of high-quality studies only. Third, both outcome measures used in this study have been criticized; the BDI has been criticized for its emphasis on cognitive aspects of depression and for being susceptible to subjective bias as a self-report measure, and the HDRS has been criticized for its emphasis on anxiety and somatic symptoms of depression and other conceptual and psychometric flaws (Bagby, Ryder, Schuller, & Marshall, 2004; Gelenberg et al., 2008). Notwithstanding these methodological problems and differences between the two instruments, the patterns of results obtained by both measures were largely similar. Fourth, this meta-analysis included only studies on adult outpatients. Therefore, it is unclear how the results of this study generalize to other populations, such as adult inpatients or children and adolescents. Fifth, we used broad inclusion criteria for depression, including different types of depression (e.g., MDD, postpartum depression, minor depression), elevated levels of depression symptoms as measured with self-report scales, and depression within a medically defined population. Although we found that pretreatment severity did not predict psychological treatment outcome in the subgroup of studies that used diagnostic criteria to define depression as well as in the subgroup that used elevated levels of depression symptoms to define depression, it might still be the case that differential moderation relationships exist for different types of depression. Sixth, only

a small number of studies had a mean pretreatment depression level in the severe range. For example, in most of the antidepressant–placebo comparisons described by Kirsch et al. (2008), the pretreatment HDRS levels were above the highest levels in our study sample. This likely reflects a difference in patient severity in the respective pharmacotherapy and psychological treatment literatures and may reflect the widespread (but possibly erroneous) belief that psychological treatment is less efficacious for more severe patients. Future psychological treatment efficacy studies focusing on more severely depressed patients would contribute to further examining the question of whether pretreatment severity moderates its effects. Seventh, and most important, only a small number of studies reported within-study severity analyses. We strongly recommend that future studies routinely report tests for Severity x Treatment interactions, because severity has been suggested to be a key moderator. The question of moderation would be even better addressed by pooling data for individual patients from multiple randomized controlled trials in a mega-analysis (see, e.g., DeRubeis, Gelfand, Tang, & Simons, 1999).

Our findings contradict the general notion in the field, which is reflected in existing treatment guidelines, that psychological treatment has little effect on more severely depressed patients. Meta-analyses based on the small number of studies reporting within-study severity analyses suggested that more severely depressed patients actually show greater differential benefit than less severely depressed patients when treated with interventions shown to be efficacious relative to more stringent nonspecific controls. However, even if psychological treatment is proven to be more efficacious relative to control conditions for more severely depressed patients than for less severely depressed patients, this does not necessarily mean that treatment guidelines should be revised to recommend psychological treatment as a mono-treatment for patients with severe depression. The question of relative efficacy of treatments for this group of patients requires head-to-head comparisons between psychological treatments and other treatments, such as medication or the combination of psychological treatment and medication. In that regard, it should be noted that despite the widespread belief that drugs (or ECT) are required for patients with more severe depressions (American Psychiatric Association, 2000), psychological treatments for outpatient samples have been compared with medication treatment in placebo-controlled trials only four times. Although CBT did not separate from placebo in the more severely depressed samples in two studies, another type of psychotherapy performed at least as well as medications and was superior to pill placebo in each of those trials (interpersonal psychotherapy in Elkin et al., 1995; and behavioral activation in Dimidjian et al., 2006). Furthermore, questions have been raised about the adequacy with which the cognitive interventions were implemented in those two studies (Hollon, Thase, & Markowitz, 2002; Jacobson & Hollon, 1996a, 1996b), and cognitive therapy has been found to be comparable with medications and superior to pill placebo in the other two placebo-controlled trials in the literature (patients with atypical depression in Jarrett et al., 1999, and patients with more severe depressions in DeRubeis et al., 2005). More such comparisons are clearly needed in more severe depressions, but there is little reason to believe that psychological treatment does not

work for such patients or that it is necessarily inferior to medications when adequately implemented.

References

References marked with an asterisk indicate studies included in the meta-analysis.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2000). Practice guideline for the treatment of patients with major depressive disorder (revision). *American Journal of Psychiatry, 157*(Suppl. 4), 1–45.

*Arean, P. A., Perri, M. G., Nezu, A. M., Schein, R. L., Christopher, F., & Joseph, T. X. (1993). Comparative effectiveness of social problem-solving therapy and reminiscence therapy as treatments for depression in older adults. *Journal of Consulting and Clinical Psychology, 61,* 1003– 1010.

*Ayen, I., & Hautzinger, M. (2004). Kognitive Verhaltenstherapie Bei Depressionen Im Klimakterium: Eine Kontrollierte, Randomisierte Interventionsstudie. [Cognitive behavior therapy for depression in menopausal women: A controlled, randomized treatment study]. *Zeitschrift für Klinische Psychologie und Psychotherapie, 33,* 290–299.

Bagby, R. M., Ryder, A. G., Schuller, D. R., & Marshall, M. B. (2004). The Hamilton Depression Rating Scale: Has the gold standard become a lead weight? *American Journal of Psychiatry, 161,* 2163–2177.

*Barlow, J. P. (1986). *A group treatment for depression in the elderly.* (Unpublished doctoral dissertation). University of Houston, Houston, Texas.

Beck, A. T., Rush, A. J., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression.* New York, NY: Guilford Press.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory–II.* San Antonio, TX: Psychological Corporation.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4,* 561–571.

*Brown, R. A., & Lewinsohn, P. M. (1984). A psychoeducational approach to the treatment of depression: Comparison of group, individual and minimal contact procedures. *Journal of Consulting and Clinical Psychology, 52,* 774–783.

*Carrington, C. H. (1979). *A comparison of cognitive and analytically oriented brief treatment approaches to depression in Black women* (Unpublished doctoral disseration). University of Maryland, College Park, Maryland.

*Comas-Diaz, L. (1981). Effects of cognitive and behavioral group treatment on the depressive symptomatology of Puerto Rican women. *Journal of Consulting and Clinical Psychology, 49,* 627–632.

Cuijpers, P., Dekker, J., Hollon, S. D., & Andersson, G. (2009). Adding psychotherapy to pharmacotherapy in the treatment of depressive disorders in adults: A meta-analysis. *Journal of Clinical Psychiatry, 70,* 1219–1229.

Cuijpers, P., Smit, F., Bohlmeijer, E., Hollon, S. D., & Andersson, G. (2010). Is the efficacy of cognitive behavior therapy and other psychological treatments for 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). 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, S. D., & Andersson, G. (2010). 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., Smit, F., & Andersson, G. (2009). Is psychotherapy for depression equally effective in younger and older adults? A meta-regression analysis. *International Psychogeriatrics, 21,* 16–24.

Cuijpers, P., van Straten, A., van Oppen, P., & Andersson, G. (2008). 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). 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). Characteristics of effective psychological treatments of depression: A meta-regression analysis. *Psychotherapy Research, 18,* 225–236.

*Cullen, J. M. (2002). *Testing the effectiveness of behavioral activation therapy in the treatment of acute unipolar depression* (Unpublished doctoral dissertation). Western Michigan University, Kalamazoo, Michigan.

DeRubeis, R. J., Gelfand, L. A., Tang, T. Z., & Simons, A. D. (1999). Medication versus cognitive behavior therapy for severely depressed outpatients: Mega-analysis of four randomized comparisons. *American Journal of Psychiatry, 156,* 1007–1013.

DeRubeis, R. J., Hollon, S. D., Amsterdam, J. D., Shelton, R. C., Young, P. R., Salomon, R. M., . . . Gallop, R. (2005). Cognitive therapy vs. medications in the treatment of moderate to severe depression. *Archives of General Psychiatry, 62,* 409–416.

*Dimidjian, S., Hollon, S. D., Dobson, K. S., Schmaling, K. B., Kohlenberg, R. J., Addis, M. E., . . . Jacobson, N. S. (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.

Driessen, E., Cuijpers, P., de Maat, S. C. M., Abbass, A. A., de Jonghe, F., & Dekker, J. J. M. (2010). The efficacy of short-term psychodynamic psychotherapy for depression: A meta-analysis. *Clinical Psychology Review, 30,* 25–36.

*Dunn, N. J., Rehm, L. P., Schillaci, J., Soucheck, J., Mehta, P., Ashton, C. M., . . . Hamilton, J. D. (2007). A randomized trial of self-management and psychoeducational group therapies for comorbid chronic posttraumatic stress disorder and depressive disorder. *Journal of Traumatic Stress, 20,* 221–237.

Elkin, I., Gibbons, R. D., Shea, M. T., Sotsky, S. M., Watkins, J. T., Pilkonis, P. A., & Hedeker, D. (1995). Initial severity and differential treatment outcome in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Journal of Consulting and Clinical Psychology, 63,* 841–847.

*Elkin, I., Shea, M. T., Watkins, J. T., Imber, S. D., Sotsky, S. M., Collins, J. F., . . . Parloff, M. B. (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program: General effectiveness of treatments. *Archives of General Psychiatry, 46,* 971–982.

*Floyd, M., Scogin, F., McKendree-Smith, N. L., Floyd, D. L., & Rokke, P. D. (2004). Cognitive therapy for depression: A comparison of individual psychotherapy and bibliotherapy for depressed older adults. *Behavior Modification, 28,* 297–318.

Fournier, J. C., DeRubeis, R. J., Hollon, S. D., Dimidjian, S., Amsterdam, J. D., Shelton, R. C., & Fawcett, J. (2010). Antidepressant drug effects and depression severity: A patient-level meta-analysis. *Journal of the American Medical Association, 303,* 47–53.

Fournier, J. C., DeRubeis, R. J., Shelton, R. C., Hollon, S. D., Amsterdam, J. D., & Gallop, R. (2009). Prediction of response to medication and cognitive therapy in the treatment of moderate to severe depression. *Journal of Consulting and Clinical Psychology, 77,* 775–787.

Gelenberg, A. J., Thase, M. E., Meyer, R. E., Goodwin, F. K., Katz, M. M., Kraemer, H. C., . . . Rosenbaum, J. F. (2008). The history and current state of antidepressant clinical trial design: A call to action for proof-of-concept studies. *Journal of Clinical Psychiatry, 69,* 1513–1528.

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

Hamilton, M. (1967). Development of a rating scale for primary depressive illness. *British Journal of Social and Clinical Psychology, 6,* 278–296.

*Haringsma, R., Engels, G. I., Cuijpers, P., & Spinhoven, P. (2006). Effectiveness of the Coping With Depression (CWD) course for older adults provided by the community-based mental health care system in the Netherlands: A randomized controlled field trial. *International Psychogeriatrics, 18,* 307–325.

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal,* 327, 557–560.

*Holden, J. M., Sagovsky, R., & Cox, J. L. (1989). Counseling in a general practice setting: Controlled study of health visitor intervention in treatment of postnatal depression. *British Medical Journal, 298,* 223–226.

Hollon, S. D., Thase, M. E., & Markowitz, J. C. (2002). Treatment and prevention of depression. *Psychological Science in the Public Interest, 3,* 39–77.

Jacobson, N. S., & Hollon, S. D. (1996a). Cognitive behavior therapy versus pharmacotherapy: Now that the jury's returned its verdict, it's time to present the rest of the evidence. *Journal of Consulting and Clinical Psychology, 64,* 74–80.

Jacobson, N. S., & Hollon, S. D. (1996b). Prospects for future comparisons between drugs and psychotherapy: Lessons from the CBT-versus-pharmacotherapy exchange. *Journal of Consulting and Clinical Psychology, 64,* 104–108.

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

Khan, A., Leventhal, R. M., Khan, S. R., & Brown, W. A. (2002). Severity of depression and response to antidepressants and placebo: An analysis of the Food and Drug Administration database. *Journal of Clinical Psychopharmacology, 22,* 40–45.

Kirsch, I., Deacon, B. J., Huedo-Medina, T. B., Scoboria, A., Moore, T. J., & Johnson, B. T. (2008). Initial severity and antidepressant benefits: A meta-analysis of data submitted to the Food and Drug Administration. *Public Library of Science Medicine, 5,* e45.

Kraemer, H. C., & Kupfer, D. J. (2006). Size of treatment effects and their importance to clinical research and practice. *Biological Psychiatry,* 59, 990–996.

Kraemer, H. C., Wilson, G. T., Fairburn, C. G., & Agras, W. S. (2002). Mediators and moderators of treatment effects in randomized clinical trial. *Archives of General Psychiatry, 59,* 877–883.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational and behavioral treatment. *American Psychologist, 48,* 1181–1209.

*McKendree-Smith, N. L. (1998). *Cognitive and behavioral bibliotherapy for depression: An examination of efficacy and mediators and moderators of change* (Unpublished doctoral dissertation). University of Alabama, Tuscaloosa, Alabama.

*McLean, P. D., & Hakstian, A. R. (1979). Clinical depression: Comparative efficacy of outpatient treatments. *Journal of Consulting and Clinical Psychology, 47,* 818–836.

*Murphy, G. E., Carney, R. M., Knesevich, M. A., Wetsel, R. D., & Whitworth, P. (1995). Cognitive behavior therapy, relaxation training and tricyclic antidepressant medication in the treatment of depression. *Psychological Reports, 77,* 403–420.

*Nezu, A. M., & Perri, M. G. (1989). Social problem-solving therapy for unipolar depression: An initial dismantling investigation. *Journal of Consulting and Clinical Psychology, 57,* 408–413.

*Pecheur, D. R. (1980). *A comparison of the efficacy of secular and religious cognitive behavior modification in the treatment of depressed Christian college students* (Unpublished doctoral dissertation). Rosemead Graduate School of Professional Psychology, La Mirada, California.

*Rohen, N. (2002). *Analysis of efficacy and mediators of outcome in minimal-contact cognitive bibliotherapy used in the treatment of depressive symptoms* (Unpublished doctoral dissertation). University of Alabama, Tuscaloosa, Alabama.

*Selmi, P. M., Klein, M. H., Greist, J. H., Sorrell, S., & Erdman, H. P. (1990). Computer-administered cognitive–behavioral therapy for depression. *American Journal of Psychiatry, 147,* 51–56.

*Simpson, S., Corney, R., Fitzgerald, P., & Beecham, J. (2003). A randomized controlled trial to evaluate the effectiveness and cost-effectiveness of psychodynamic counseling for general practice patients with chronic depression. *Psychological Medicine, 33,* 229–239.

*Taylor, F. G., & Marshall, W. L. (1977). Experimental analysis of a cognitive– behavioral therapy for depression. *Cognitive Therapy and Research, 1,* 59–72.

*Teri, L., Logsdon, R. G., Uomoto, J., & McCurry, S. M. (1997). Behavioral treatment of depression in dementia patients: A controlled clinical trial. *Journals of Gerontology: Series B. Psychological Sciences and Social Sciences, 52,* 159–166.

*Usaf, S. O., & Kavanagh, D. J. (1990). Mechanisms of improvement in treatment for depression: Test of a self-efficacy and performance model. *Journal of Cognitive Psychotherapy: An International Quarterly, 4,* 51–70.

*Van Schaik, A., van Marwijk, H., Ader, H., van Dyck, R., de Haan, M., Penninx, B., . . . Beekman, A. (2006). Interpersonal psychotherapy for elderly patients in primary care. *American Journal of Geriatric Psychiatry, 14,* 777–786.

Wampold, B. E., Minami, T., Tierney, S. C., Baskin, T. W., & Bhati, K. S. (2005). Placebo is powerful: Estimating placebo effects in medicine and psychotherapy from randomized clinical trails. *Journal of Clinical Psychology, 61,* 835–854.

*Wilson, P. H., Goldin, J. C., & Charbonneau-Powis, M. (1983). Comparative efficacy of behavioral and cognitive treatments of depression. *Cognitive Therapy and Research, 7,* 111–124.

*Wright, J. H., Wright, A. S., Albano, A. M., Basco, M. R., Goldsmith, L. J., Raffield, T., & Otto, M. W. (2005). Computer-assisted cognitive therapy for depression: Maintaining efficacy while reducing therapist time. *American Journal of Psychiatry, 162,* 1158–1164.