Original Investigation | META-ANALYSIS

Baseline Depression Severity as Moderator of Depression Outcomes Between Cognitive Behavioral Therapy vs Pharmacotherapy
An Individual Patient Data Meta-analysis

Erica S. Weitz, MA; Steven D. Hollon, PhD; Jos Twisk, PhD; Annemieke van Straten, PhD; Marcus J. H. Huibers, PhD; Daniel David, PhD; Robert J. DeRubeis, PhD; Sona Dimidjian, PhD; Boadie W. Dunlop, MD, MS; Ioana A. Cristea, PhD; Mahbobeh Faramarzi, PhD; Ulrich Hegerl, MD, PhD; Robin B. Jarrett, PhD; Farzan Kheirkhah, MD; Sidney H. Kennedy, MD; Roland Mergl, PhD; Jeanne Miranda, PhD; David C. Mohr, PhD; A. John Rush, MD; Zindel V. Segal, PhD; Juned Siddique, DrPH; Anne D. Simons, PhD; Jeffrey R. Vittengl, PhD; Pim Cuijpers, PhD

IMPORTANCE Current guidelines recommend treating severe depression with pharmacotherapy. Randomized clinical trials as well as traditional meta-analyses have considerable limitations in testing for moderators of treatment outcomes.

OBJECTIVES To conduct a systematic literature search, collect primary data from trials, and analyze baseline depression severity as a moderator of treatment outcomes between cognitive behavioral therapy (CBT) and antidepressant medication (ADM).

DATA SOURCES A total of 14 902 abstracts were examined from a comprehensive literature search in PubMed, PsycINFO, EMBASE, and Cochrane Registry of Controlled Trials from 1966 to January 1, 2014.

STUDY SELECTION Randomized clinical trials in which CBT and ADM were compared in patients with a *DSM*-defined depressive disorder were included.

DATA EXTRACTION AND SYNTHESIS Study authors were asked to provide primary data from their trial. Primary data from 16 of 24 identified trials (67%), with 1700 outpatients (794 from the CBT condition and 906 from the ADM condition), were included. Missing data were imputed with multiple imputation methods. Mixed-effects models adjusting for study-level differences were used to examine baseline depression severity as a moderator of treatment outcomes.

MAIN OUTCOMES AND MEASURES Seventeen-item Hamilton Rating Scale for Depression (HAM-D) and Beck Depression Inventory (BDI).

RESULTS There was a main effect of ADM over CBT on the HAM-D (β = −0.88; P = .03) and a nonsignificant trend on the BDI (β = −1.14; P = .08, statistical test for trend), but no significant differences in response (odds ratio [OR], 1.24; P = .12) or remission (OR, 1.18; P = .22). Mixed-effects models using the HAM-D indicated that baseline depression severity does not moderate reductions in depressive symptoms between CBT and ADM at outcome (β = 0.00; P = .96). Similar results were seen using the BDI. Baseline depression severity also did not moderate the likelihood of response (OR, 0.99; P = .77) or remission (OR, 1.00; P = .93) between CBT and ADM.

CONCLUSIONS AND RELEVANCE Baseline depression severity did not moderate differences between CBT and ADM on the HAM-D or BDI or in response or remission. This finding cannot be extrapolated to other psychotherapies, to individual ADMs, or to inpatients. However, it offers new and substantial evidence that is of relevance to researchers, physicians and therapists, and patients.

Supplemental content at jamapsychiatry.com

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Erica S. Weitz, MA, Department of Clinical Psychology and EMGO Institute for Health and Care Research, VU University Amsterdam, Van der Boechorststraat 1, 1081 BT Amsterdam, the Netherlands (e.weitz@vu.nl).

E1

There is no shortage of effective treatments for depression, including pharmacotherapy[1] and psychotherapies, of which cognitive behavioral therapy (CBT) is one of the best documented.[2,3] Previous meta-analyses[4,5] have shown that psychotherapies are at least as effective as pharmacotherapy, used as monotherapy, in treating depression of mild and moderate symptom severity (defined by cutoff scores on depressive symptom inventories). However, less is known about the relative efficacy of psychotherapy vs pharmacotherapy in severely depressed populations.[6,7]

Nonetheless, American Psychiatric Association[8] and British Association for Psychopharmacology[9] guidelines for the treatment of depression suggest that although psychotherapy is sufficient for treating mild depression, antidepressant medications (ADMs) should be used to treat severe depression in the context of major depressive disorder. This recommendation is mainly owing to the well-known findings of the National Institute of Mental Health Treatment of Depression Collaborative Research Program,[10] in which CBT was less effective than medications in the treatment of participants with severe depression. However, these differences were not observed in several other randomized clinical trials (RCTs) of acute-phase treatment.[11-15] One limitation of RCTs is that they often include too few patients and thus lack sufficient power to detect moderation of outcomes and thoroughly examine the efficacy of these treatments in severe depression.

Therefore, several meta-analyses aggregated the results of studies examining the effects of psychotherapy and pharmacotherapy on severe depression relative to control conditions. Two rigorous meta-analyses showed that psychotherapy[16] and pharmacotherapy[17] were more efficacious than control treatments for severe depression. However, the findings for psychotherapy should be interpreted with caution because they were based on study-level as opposed to patient-level pretreatment severity and smaller subgroup analysis. In addition, both meta-analyses provided information only on the effectiveness of a single treatment modality and did not address the crucial issue of the relative efficacy of psychotherapy and pharmacotherapy. A conventional meta-analysis[7] that directly compared psychotherapy and pharmacotherapy in severe depression showed no significant differences between the treatment groups. However, this finding was based on only 4 studies in the sample that reported baseline depression severity. Another meta-analysis[18] of CBT vs pharmacotherapy using individual patient data provided substantial information, but only 4 studies were included in the analysis, thus limiting power and representativeness.

Although traditional meta-analyses are useful in aggregating evidence, they are limited in their ability to test for moderation of outcomes. In 2 of 7 of the aforementioned meta-analyses,[7,16] the authors used the mean pretreatment depression scores of the full sample of the studies as an indication of severity. However, many studies have mean depression baseline scores in the moderate range even if the sample includes patients with severe depression, thus restricting the range of severity examined. The analysis is then limited to studies that include secondary comparisons of the severe sample or to studies that recruited a highly severe sample, both of which are rare.

Many of these concerns regarding conventional meta-analysis and RCTs can be addressed by using individual patient data meta-analysis (IPDMA), which includes raw data from RCTs. The use of IPDMA is a new technique in the mental health field, but it has been used successfully to examine acute and preventive treatments in medicine. Although conventional meta-analyses are appropriate for pooling outcomes, the large sample size of IPDMAs provides more power to accurately examine moderators of treatment outcomes.[19]

Therefore, we conducted an IPDMA to provide the best estimate of the efficacy of psychotherapy relative to pharmacotherapy for the treatment of severe depression. Cognitive behavioral therapy was chosen as the comparison for specificity because it is well researched and widely available and allows for better translation to routine practice.

Methods

Identification and Selection of Studies

Study searches were conducted using several methods. First, we used a database of studies of RCTs on the psychological treatment of adult depression. This database has been described elsewhere[7] and has been used in a series of earlier published meta-analyses (http://www.evidencebasedpsychotherapies.org). The database was developed by comprehensive literature searches (from 1966 to January 1, 2014). In these searches, 14 902 abstracts were examined from PubMed (n = 3864), PsycINFO (n = 2960), EMBASE (n = 4320), and Cochrane Registry of Controlled Trials (n = 3758). Studies examined were on the psychological treatment of depression in general. Earlier meta-analyses were searched for confirmation that no RCTs were previously missed. From 14 902 abstracts (10 992 after the removal of duplicates), we retrieved 1613 full-text articles for possible inclusion in the database.

We included RCTs in which CBT was compared with pharmacotherapy among patients with a primary diagnosis of a depressive disorder established by a standardized diagnostic interview. No language restrictions were applied. Only studies in which the patients met diagnostic criteria for depressive disorder (major depressive disorder or dysthymia) were included.[1] Studies used DSM-II, DSM-III, or DSM-IV diagnosis of depressive disorder. Cognitive behavioral therapies were required to be manualized and use cognitive restructuring as the main component of treatment.[20] Studies were excluded if they were aimed at relapse prevention or maintenance treatments or if they included adolescents or children younger than 18 years. Studies of inpatient populations were excluded because inpatients per definition receive more care than CBT or ADM alone and because patient characteristics likely differ between these 2 populations. Therefore, these studies were excluded to prevent high heterogeneity. Studies that included populations with comorbid general medical disorders were not excluded.

Authors of identified studies were invited via email to participate in the IPDMA and provide original data from their trial. If the authors did not respond to the request after 1 month, a second reminder email was sent and efforts to contact coauthors were made. If no response was received, we considered the data unavailable and did not include the study in the analysis.

Quality Assessment and Data Extraction

Study quality was assessed using 4 criteria from the Cochrane Collaboration's tool for evaluating the risk of bias.[21] This tool assesses whether there was adequate generation of randomization sequence, concealment of treatment allocation, masking of assessors, and appropriate methods for addressing missing data, which was denoted as positive when the analysis was completed in the intent-to-treat (ITT) sample, meaning that all randomized patients were included in the analysis. Only data in the published articles were used to determine the risk of bias. Two independent researchers (P.C. and an outside assessor) conducted quality assessments.

Meta-analysis

We performed a meta-analysis to examine differences between the 16 studies that provided data and the 8 studies that did not. First, we calculated the effect sizes indicating the difference between CBT and ADM at posttreatment based on data reported in the published articles. The effect sizes were calculated by subtracting the average score of the CBT group from the average score of the ADM group at posttest and then dividing the results by the pooled SD. If studies included only dichotomous outcomes without reporting means and SDs, we used the effect-size calculations outlined by Borenstein and colleagues.[22] Bias associated with small sample size was corrected using procedures described by Hedges and Olkin.[23]

Comparisons between CBT and ADM outcomes in studies that provided data vs those that did not were performed as a meta-regression analysis (Stata, version 13.1; StataCorp LP). The effect size was the dependent variable, and a predictor variable indicating whether we received the data was included. Patient and study characteristics were entered as covariates. In addition, the effects of publication bias on the included studies were inspected by examining a funnel plot produced by the trim and fill procedure by Duval and Tweedie[24] and by conducting a test by Egger et al[25] of the intercepts in a software program (Comprehensive Meta-Analysis, version 2.2.021; http://www.meta-analysis.com).

Individual Patient Data Analysis

Continuous scores on the HAM-D and BDI at baseline were used to determine baseline severity.[26-28] Fourteen of 16 studies contributed HAM-D scores (when studies provided HAM-D-21 scores; HAM-D-17 scores were calculated from individual items and used in all analyses), 4 studies contributed BDI-I scores, and 9 studies contributed BDI-II scores. Two studies were unable to contribute complete BDI scores and were removed from the BDI analysis. Ten of 16 studies contributed both HAM-D and BDI scores. The BDI-I scores were converted to BDI-II scores according to the measure's manual,[28] and the aggregated BDI-I and BDI-II scores are referred to below as the BDI. Full-sample analyses were based on the ITT sample, including all randomized patients, except for 3 trials that used a modified ITT or completer sample as cited in the published trials.[12,29,30] The details of the IPDMA are discussed further in the eMethods and eResults in the Supplement.

The studies included in our main analysis are listed in the eTable in the Supplement. We obtained data on randomized patients from 16 studies, combined the trials into 1 data set, and then imputed missing outcome data under a missing-at-random assumption, with missing data imputed using a software program (mi impute mvn in Stata, version 13.1; StataCorp LP). Using multiple imputation with a missing-at-random assumption tends to yield more unbiased results than using completer samples or mean imputation.[31] Overall, 30% of HAM-D posttreatment data and 18% of BDI posttreatment data were missing. Participants' missing outcome data were imputed 100 times using complete patient and study characteristics, such as baseline depression score, sex, length of treatment, and treatment group, as the predictor variables.[2] As a robustness check, we conducted analyses only among studies with complete data.
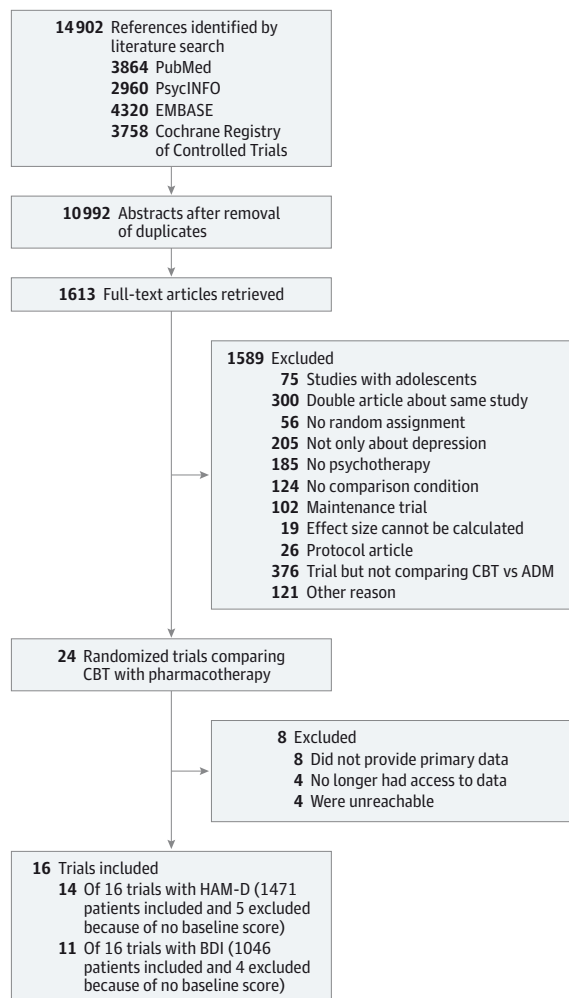
For patient-level data, we analyzed the effects of depression severity on treatment outcomes using a 1-step IPDMA approach that allows for the most sophisticated modeling of covariates.[21] It has better performance in terms of power and yields less biased estimates compared with 2-step IPDMAs, in which individual patient data are used to estimate the treatment × moderator interaction within each trial, followed by a standard inverse variance meta-analysis.[32-34]

We used multilevel linear and logistic regression and clustered on the study level to control for unobserved heterogeneity between studies. We used the default maximum likelihood algorithm in the software program (Stata, version 13.1; StataCorp LP). A 2-level multilevel linear regression with patient-level data as level 1 and with study-level data as level 2 was used in all further analyses.

The primary analysis concerned whether baseline depression severity was a moderator between CBT and ADM on depression outcomes. However, we first analyzed the effects of treatment group on depression outcomes while holding baseline severity constant. Posttreatment scores on the HAM-D or BDI were used as the outcome variable, and baseline depression score and treatment group were independent variables. To examine whether baseline depression severity was a moderator between CBT and ADM on depression outcomes, we added the interaction between baseline severity and treatment outcomes into the multilevel linear regression model. To examine the effects of patient and study variables on outcomes, we ran an adjusted model controlling for length of treatment, type of medication, demographic variables (age, sex, and marital status), and the risk of bias (sequence generation, allocation of concealment, masking, and ITT analysis). Finally, we ran the same analysis with only study completers. To examine clinically relevant outcomes, we ran the same models using response (50% reduction in scores on posttest HAM-D) and remission (score of ≤7 on posttest HAM-D) as outcomes.[35] The definition of remission did not include the duration in remission, which was not reported uniformly across studies.

To test the robustness of these findings, we examined a subset of the sample that met criteria for severe depression according to a less restrictive HAM-D standard of severity (HAM-D-17 score >19),[26] stricter criteria of the UK National Institute for Clinical Excellence (HAM-D-17 score >23),[36] and BDI-II cutoffs (score >28).[28] We ran multilevel linear regression models with posttreatment depression score as the dependent variable and intervention group as the independent variable.

Figure 1. Flowchart of Included Trials

ADM indicates antidepressant medication; BDI, Beck Depression Inventory; CBT, cognitive behavioral therapy; and HAM-D, Hamilton Rating Scale for Depression.

In addition, several sensitivity analyses examined the effects of certain subgroups of studies on the results. This included a subset in which trials were removed that included special populations (patients with multiple sclerosis or peripartum depression), placebo-controlled trials, and lower-quality scores to examine whether the inclusion of these studies[18] affected the results.

Results

Study Selection
Figure 1 shows the inclusion process. We originally retrieved 1613 full-text articles but excluded 1589 for various reasons. Twenty-four studies met the inclusion criteria for the IPDMA. Authors from 16 of 24 identified trials (67%), with 1700 outpatients, agreed to participate and provided data from their original study. Among 8 authors who did not provide data, 4 indicated that they no longer had access to data and 4 were unreachable.

Characteristics of Included Studies
In 6 studies, patients were exclusively recruited from clinical samples, 6 studies recruited patients (in part) through the community, and 4 studies used other recruitment methods. Thirteen studies recruited adults in general while 3 studies recruited specific populations (patients with multiple sclerosis, women who earned a low income, or women with infertility). Eleven studies were conducted in the United States, 2 in Canada, and 1 each from Germany, Romania, and Iran. In 9 studies, a selective serotonin reuptake inhibitor was used for pharmacotherapy, 4 studies prescribed a tricyclic antidepressant, and 3 studies used another antidepressant or a predefined protocol for deciding which medication to prescribe. One study[14] allowed for augmentation with lithium or desipramine hydrochloride and 2 studies[14,37] allowed a medication switch if patients experienced adverse effects. In 14 studies, CBT was given individually, 2 studies used group sessions, and 1 study used both methods. The number of CBT sessions ranged from 8 to 28 (mean, 15.4; mode, 20), and 11 studies used 16 to 20 sessions. Eleven trials reported measuring CBT adherence or competence by rating taped sessions, and 10 trials reported that therapists received regular supervision.

The quality of the included studies based on the published reports varied (eTable in the Supplement). Seven studies reported adequate sequence generation and 6 studies reported allocation to conditions by an independent party. Eleven studies reported masking of outcome assessors, and ITT analyses were conducted in 12 studies. Five studies met all 4 quality criteria, 5 studies met 2 or 3 criteria, and the remaining 5 studies had lower-quality scores (0 or 1 of 4 criteria).
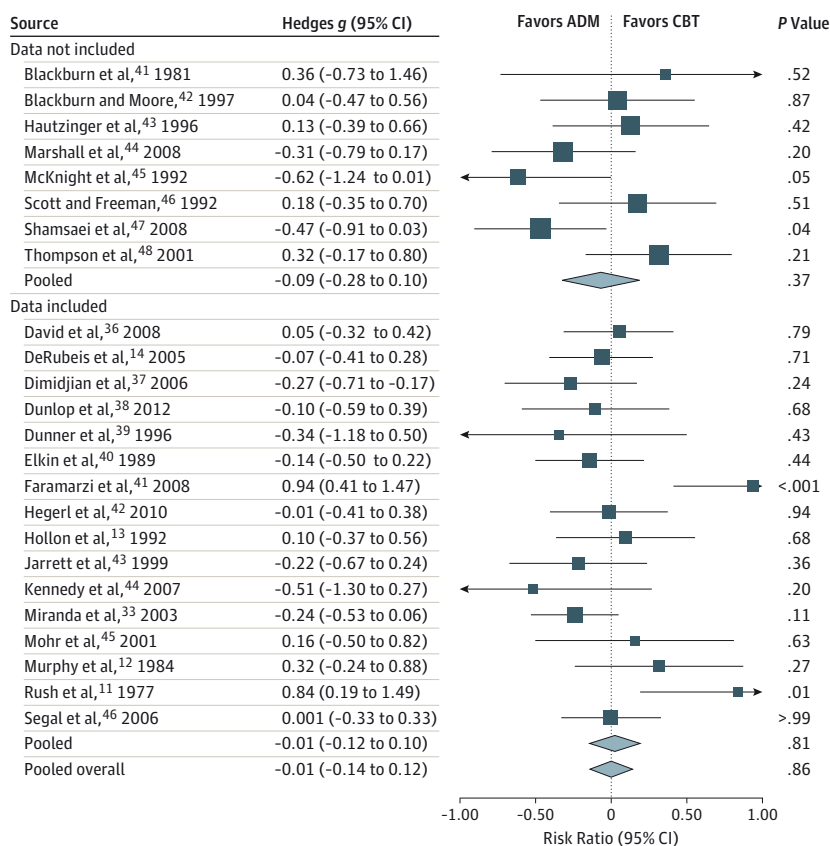
Traditional Meta-analysis Findings
Based on the results as published in all 24 articles, the difference in standardized depression scores at posttreatment between CBT and ADM was Hedges $g = -0.01$ (95% CI, -0.14 to 0.12), with low heterogeneity ($I^2 = 43\%$; 95% CI, 7%-64%).[23] There was no significant difference ($P = .54$) in the effect size between the 16 studies that provided data for our IPDMA analyses ($g = 0.01$; 95% CI, -14.00 to 0.17 and $I^2 = 46\%$; 95% CI, 4%-70%) and the 8 studies that did not ($g = -0.08$; 95% CI, -0.33 to 0.17 and $I^2 = 40\%$; 95% CI, 0%-73%). In addition, there were no indications of publication bias in the 24 studies according to the trim and fill procedure by Duval and Tweedie[24] (adjusted $g = -0.01$; 95% CI, -0.14 to 0.12) or the test by Egger ($P = .29$) (Figure 2).

The multivariable meta-regression analysis with the effect size as the dependent variable (Hedges g) and a dummy variable indicating whether a study was included in the IPDMA as the independent variable was not significant ($P = .88$) when controlling for differences in study quality and design (method of recruitment, type of medication [selective serotonin reuptake inhibitor, tricyclic antidepressant, or other]), treatment format of CBT (group or individual), and the number of intervention sessions. Again, this illustrated no differences in outcomes between studies that provided data and studies that did not.

Patient Characteristics
The sample included 1700 participants, 906 from the ADM condition and 794 from the CBT condition. Of the HAM-D out-

Figure 2. Direct Comparisons of Cognitive Behavioral Therapy (CBT) and Antidepressant Medication (ADM)

Source	Hedges g (95% CI)	Favors ADM	Favors CBT	P Value
Data not included				
Blackburn et al,[41] 1981	0.36 (−0.73 to 1.46)			.52
Blackburn and Moore,[42] 1997	0.04 (−0.47 to 0.56)			.87
Hautzinger et al,[43] 1996	0.13 (−0.39 to 0.66)			.42
Marshall et al,[44] 2008	−0.31 (−0.79 to 0.17)			.20
McKnight et al,[45] 1992	−0.62 (−1.24 to 0.01)			.05
Scott and Freeman,[46] 1992	0.18 (−0.35 to 0.70)			.51
Shamsaei et al,[47] 2008	−0.47 (−0.91 to 0.03)			.04
Thompson et al,[48] 2001	0.32 (−0.17 to 0.80)			.21
Pooled	−0.09 (−0.28 to 0.10)			.37
Data included				
David et al,[36] 2008	0.05 (−0.32 to 0.42)			.79
DeRubeis et al,[14] 2005	−0.07 (−0.41 to 0.28)			.71
Dimidjian et al,[37] 2006	−0.27 (−0.71 to -0.17)			.24
Dunlop et al,[38] 2012	−0.10 (−0.59 to 0.39)			.68
Dunner et al,[39] 1996	−0.34 (−1.18 to 0.50)			.43
Elkin et al,[40] 1989	−0.14 (−0.50 to 0.22)			.44
Faramarzi et al,[41] 2008	0.94 (0.41 to 1.47)			<.001
Hegerl et al,[42] 2010	−0.01 (−0.41 to 0.38)			.94
Hollon et al,[13] 1992	0.10 (−0.37 to 0.56)			.68
Jarrett et al,[43] 1999	−0.22 (−0.67 to 0.24)			.36
Kennedy et al,[44] 2007	−0.51 (−1.30 to 0.27)			.20
Miranda et al,[33] 2003	−0.24 (−0.53 to 0.06)			.11
Mohr et al,[45] 2001	0.16 (−0.50 to 0.82)			.63
Murphy et al,[12] 1984	0.32 (−0.24 to 0.88)			.27
Rush et al,[11] 1977	0.84 (0.19 to 1.49)			.01
Segal et al,[46] 2006	0.001 (−0.33 to 0.33)			>.99
Pooled	−0.01 (−0.12 to 0.10)			.81
Pooled overall	−0.01 (−0.14 to 0.12)			.86

−1.00   −0.50   0   0.50   1.00
Risk Ratio (95% CI)

Shown are Hedges g of studies included and not included in the individual patient data meta-analysis.

Table 1. Sample Size by Baseline Depression Severity at Posttest[a]

Baseline Severity	CBT		ADM		Cohen d (95% CI)
	No.	Posttreatment, Mean (SE)	No.	Posttreatment, Mean (SE)	
HAM-D score					
<14	62	7.89 (0.83)	74	6.08 (0.71)	0.29 (−0.63 to 0.05)
14 to <19	245	7.98 (0.41)	292	7.21 (0.45)	0.11 (−0.28 to 0.06)
≥19	344	9.89 (0.42)	450	9.07 (0.40)	0.10 (−0.24 to 0.04)
>23	104	10.62 (0.82)	152	9.79 (0.69)	0.10 (−0.35 to 0.15)
BDI score					
<29	233	10.07 (0.58)	254	9.28 (0.62)	0.08 (−0.26 to 0.09)
≥29	243	13.03 (0.73)	314	11.58 (0.67)	0.14 (−0.30 to 0.03)

Abbreviations: ADM, antidepressant medication; BDI, Beck Depression Inventory; CBT, cognitive behavioral therapy; HAM-D, Hamilton Rating Scale for Depression.

[a] The HAM-D cutoffs refer to mild, moderate, and severe according to American Psychiatric Association[8] standards, and severe according to National Institute for Clinical Excellence[36] standards at intake. The BDI cutoffs refer to less severe and severe intake scores.

come sample, 793 participants (54%) met criteria for severe depression using the more lenient HAM-D criterion, and 255 participants (17%) met the more stringent National Institute for Clinical Excellence[36] criterion. On the BDI, 509 participants (49%) met criteria for severe depression. The mean baseline scores were 19.18 on the HAM-D and 30.86 on the BDI. The mean age of the full sample was 37.38 years, 69% were female, 43% were married, and 52% were employed full-time. In total, 90% of our sample had a high school education (or 12 years of education), and 65% had a higher educational level.

IPDMA Findings

Table 1 lists the mean scores at posttreatment categorized by baseline depression severity. There was a significant main ef-

fect of ADM over CBT on the HAM-D (β = −0.88; P = .03) and a nonsignificant trend on the BDI (β = −1.14; P = .08, statistical test for trend), but no significant differences between ADM and CBT on clinically relevant outcomes of response (odds ratio [OR], 1.24; P = .12) or remission (OR, 1.18; P = .22). In total, 63% of patients in the ADM condition and 58% of patients in the CBT condition responded to treatment, and 51% of patients in the ADM condition and 47% of patients in the CBT condition met criteria for remission.

Table 2 summarizes the results of the primary analysis examining whether baseline depression severity is a moderator between treatments. When including the interaction effect in the model, the treatment effect (ADM vs CBT) did not differ as a function of severity (β = 0.00; P = .96 for interaction ef-

Table 2. Mixed-Effects Models of the Effects of Baseline Depression Severity and Treatment Group and Their Interaction on the HAM-D, the BDI, Response, and Remission[a]

Variable[b]	Full Sample		Adjusted Model		Study Completers Only	
	Statistic (SE) (n = 1466)	P Value	Statistic (SE) (n = 1170)	P Value	Statistic (SE) (n = 1036)	P Value
HAM-D						
Baseline severity	0.32 (0.07)	<.001	0.38 (0.08)	<.001	0.35 (0.07)	<.001
Treatment group	−0.88 (0.41)	.03	−1.01 (0.51)	.05	−0.91 (0.39)	.02
Baseline severity × treatment group	0.00 (0.08)	.96	−0.06 (0.09)	.50	−0.02 (0.08)	.85
Response						
Baseline severity	1.03 (0.02)	.15	1.01 (0.03)	.72	1.02 (0.02)	.44
Treatment group	1.24 (0.17)	.12	1.23 (0.20)	.22	1.27 (0.17)	.08
Baseline severity × treatment group	0.99 (0.03)	.77	1.01 (0.03)	.84	0.99 (0.03)	.83
Remission						
Baseline severity	0.93 (0.02)	<.001	0.90 (0.02)	<.001	0.93 (0.02)	<.001
Treatment group	1.18 (0.16)	.22	1.17 (0.19)	.33	1.23 (0.16)	.12
Baseline severity × treatment group	1.00 (0.03)	.93	1.03 (0.03)	.37	1.00 (0.03)	.90
BDI						
Baseline severity	0.18 (0.05)	<.001	0.24 (0.05)	<.001	0.19 (0.05)	<.001
Treatment group	−1.15 (0.66)	.08	−1.64 (0.70)	.02	−1.16 (0.67)	.08
Baseline severity × treatment group	0.00 (0.07)	.96	−0.02 (0.07)	.78	−0.01 (0.07)	.89

Abbreviations: ADM, antidepressant medication; BDI, Beck Depression Inventory; HAM-D, Hamilton Rating Scale for Depression.

[a] Statistics are β levels for the HAM-D and BDI and odds ratios for response and remission. Two-tailed P values are reported. Baseline severity is centered around the mean, and intercepts are not shown. Full-sample and adjusted-model analyses are conducted on the modified intent-to-treat model as defined by individual studies. Five participants were excluded in the HAM-D analysis and 4 participants in the BDI analysis for not having baseline depression severity scores.

[b] For treatment group, cognitive behavioral therapy is the reference.

fect) (eFigure in the Supplement). Nonsignificant differences between treatments as a function of severity were also obtained when response (OR, 0.99; P = .77) or remission (OR, 1.00; P = .93) was used as the outcome measure. Adjusting the model to control for study-level and patient-level characteristics did not alter this lack of interaction. In addition, no differences were detected in the speed of improvement between the 2 treatments when time (length of the intervention in weeks) was included in the interaction.

Further analyses on the BDI showed comparable nonsignificant results when interacting baseline depression severity and treatment group (Table 2). Additional analysis showed no significant differential treatment response between CBT and ADM when analyzing only the severe sample for HAM-D score exceeding 19 (β = −0.88; P = .12), HAM-D score exceeding 2 (β = −0.73; P = .48), and the BDI (β = −1.53; P = .14).

Sensitivity analyses found no evidence of moderation of outcomes as a function of baseline depression severity in several models. These findings are reported in the eResults in the Supplement.

Discussion

In this IPDMA, we found no evidence that baseline severity of depression, whether patient or clinician rated, moderated the effect of treatment on outcomes. That is, patients with more severe depression were no more likely to require medications to improve than patients with less severe depression, and these findings were robust in sensitivity analyses. There was a mod-

est (<1 point on the HAM-D) main effect of ADM over CBT on the continuous outcomes (HAM-D and BDI) but no evidence of any interaction, which provides new and important information for the debate about treatments for severe depression. Although guidelines[8,9] suggest that patients with severe depression require pharmacotherapy, we found no evidence that differences between ADM and CBT are moderated by baseline depression severity. Furthermore, robustness analysis on the severe sample alone showed no differential treatment response between CBT and ADM. Therefore, CBT may also be an effective first-line treatment for these patients.

However, there are some limitations to consider. The BDI and HAM-D outcome measures have been criticized. The BDI emphasizes cognitive aspects of depression and as a self-report measure may be prone to bias while the HAM-D contains some psychometric flaws and emphasizes anxiety and somatic symptoms.[8,38] Moreover, neither specifically addresses functional impairment. Nonetheless, these 2 depression measures are widely used in research and clinical practice. As such, they provide an understanding of treatment outcomes for depressive symptoms.

In addition, not all studies identified as meeting the inclusion criteria for the meta-analysis contributed data. Although we tested for and did not detect bias, it is possible that the included studies were not completely representative. Some studies had quality scores that were suboptimal. Determining quality from the published articles allowed for a consistent and conservative study approach. However, quality may be higher than reported. Sensitivity analyses were performed by removing lower-quality studies, which did not affect our findings.

Samples of RCTs for depression may also not be representative of patients with depression treated in primary and psychiatric care clinics,[39] which may be because of patients' willingness to accept randomization, because of their previous treatment experiences, or because study criteria may exclude patients with certain comorbid disorders. In addition, the studies included did not incorporate inpatient populations; therefore, these findings cannot be extrapolated to patients having severe depression with imminent suicidality or psychosis. Outcomes comparing CBT and ADM could have varied depending on the expertise and supervision of the therapists and psychiatrists and the adherence to treatment regimens; however, it was not possible to examine the contribution of the quality of treatment in this analysis.

Furthermore, these findings might not generalize to other psychotherapies or ADMs that were not represented in the included studies. They also may not pertain to combination treatments and may reflect data only from studies of acute outcomes. Prior exposure to CBT has been found to reduce rates of relapse relative to prior exposure to medication after treatment termination.[40] It would be important to determine whether that finding holds across the full range of initial depression severity.

Conclusions

While there are some study limitations, the defining strength of our meta-analysis is that it is the first investigation to date, to our knowledge, with sufficient power to examine baseline depression severity as a moderator of treatment outcomes between 2 active treatments. We found no evidence of any such interaction. While this IPDMA shows that pharmacotherapy provides minor improvement in the treatment of depression relative to CBT in terms of the continuous measures, there is no indication that differences between the modalities were moderated by the degree of baseline depression severity. Therefore, the data are insufficient to recommend ADM over CBT in outpatients based on baseline severity alone. More research is needed to examine whether other demographic and clinical characteristics moderate the differential response between CBT and ADM.

Author Affiliations: Department of Clinical Psychology and EMGO Institute for Health and Care Research, VU University Amsterdam, Amsterdam, the Netherlands (Weitz, van Straten, Huibers, Cuijpers); Department of Psychology, Vanderbilt University, Nashville, Tennessee (Hollon); Department of Epidemiology and Biostatistics, EMGO Institute for Health and Care Research, VU University Amsterdam, Amsterdam, the Netherlands (Twisk); Department of Clinical Psychology and Psychotherapy, Babes-Bolyai University, Cluj, Romania (David, Cristea); Department of Psychology, University of Pennsylvania, Philadelphia (DeRubeis); Department of Psychology and Neuroscience, University of Colorado, Boulder (Dimidjian); Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, Georgia (Dunlop); Fatemeh Zahra Infertility and Reproductive Health Research Center, Babol University of Medical Sciences, Babol, Iran (Faramarzi); Department of Psychiatry and Psychotherapy, University of Leipzig, Leipzig, Germany (Hegerl, Mergl); Department of Psychiatry, The University of Texas Southwestern Medical Center, Dallas (Jarrett); Department of Psychiatry, Faculty of Medicine, Babol University of Medical Sciences, Babol, Iran (Kheirkhah); Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada (Kennedy); Health Services Research Center, Neuropsychiatric Institute, University of California, Los Angeles (Miranda); Center for Behavioral Intervention Technologies, Feinberg School of Medicine, Northwestern University, Chicago, Illinois (Mohr); Duke–National University of Singapore Graduate Medical School, Singapore (Rush); Department of Psychology, University of Toronto Scarborough, Toronto, Ontario, Canada (Segal); Department of Preventive Medicine, Feinberg School of Medicine, Northwestern University, Chicago, Illinois (Siddique); Department of Psychology, University of Notre Dame, Notre Dame, Indiana (Simons); Department of Psychology, Truman State University, Kirksville, Missouri (Vittengl).

Author Contributions: Ms Weitz had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Study concept and design: Weitz, Hollon, Cuijpers.
Acquisition, analysis, or interpretation of data: All authors.
Drafting of the manuscript: Weitz, van Straten, Cuijpers.
Critical revision of the manuscript for important intellectual content: All authors.
Statistical analysis: Weitz, Twisk, Cuijpers.
Administrative, technical, or material support: All authors.
Study supervision: van Straten, Huibers, Cuijpers.

Conflict of Interest Disclosures: Dr Hollon reported having received research support from the National Institute of Mental Health (NIMH) (grants MH60713 and MH01697) (NIMH). Dr DeRubeis reported having received research support from the NIMH (grant MH60998). Dr Dunlop reported having received grant support from AstraZeneca, Bristol-Myers Squibb, Forest Laboratories, GlaxoSmithKline, NIMH, Otsuka Pharmaceutical, and Pfizer and reported receiving honoraria for consulting from Hoffmann-LaRoche, MedAvante, and Pfizer. Dr Hegerl reported serving as an advisory board member for Eli Lilly, Lundbeck, Otsuka Pharmaceutical, Takeda, and Servier; reported serving as a consultant for Nycomed (a Takeda company); and reported serving as a speaker for Bristol-Myers Squibb, MEDICE Arzneimittel, Novartis, and Roche Pharma. Dr Jarrett reported using data from grant MH-45043 from the NIMH, reported being a paid consultant to UpToDate, and reported that her medical center receives fees for cognitive therapy that she provides to patients. Dr Mergl reported having a consultancy agreement with Nycomed. Dr Mohr reported having received research support from the National Institutes of Health (grants R01 MH100482, R01 MH095753, P20 MH090318, and R34 MH095907) and reported having a consulting relationship with Otsuka Pharmaceutical. No other disclosures were reported.

REFERENCES

1. Mulrow CD, Williams JW Jr, Chiquette E, et al. Efficacy of newer medications for treating depression in primary care patients. *Am J Med*. 2000;108(1):54-64.

2. Churchill R, Hunot V, Corney R, et al. A systematic review of controlled trials of the effectiveness and cost-effectiveness of brief psychological treatments for depression. *Health Technol Assess*. 2001;5(35):1-173.

3. Butler AC, Chapman JE, Forman EM, Beck AT. The empirical status of cognitive-behavioral therapy: a review of meta-analyses. *Clin Psychol Rev*. 2006;26(1):17-31.

4. Dobson KS. A meta-analysis of the efficacy of cognitive therapy for depression. *J Consult Clin Psychol*. 1989;57(3):414-419.

5. De Maat S, Dekker J, Schoevers R, De Jonghe F. Relative efficacy of psychotherapy and pharmacotherapy in the treatment of depression: a meta-analysis. *Psychother Res*. 2006;16(5):566-578.

6. Robinson LA, Berman JS, Neimeyer RA. Psychotherapy for the treatment of depression: a comprehensive review of controlled outcome research. *Psychol Bull*. 1990;108(1):30-49.

7. Cuijpers P, van Straten A, van Oppen P, Andersson G. Are psychological and pharmacologic interventions equally effective in the treatment of adult depressive disorders? a meta-analysis of comparative studies. *J Clin Psychiatry*. 2008;69(11):1675-1685.

8. Gelenberg AJ, Freeman MP, Markowitz JC, et al. *Treatment of Patients With Major Depressive Disorder.* 3rd ed. Washington, DC: American Psychiatric Association; 2010.

9. Anderson I, Ferrier I, Baldwin R, et al. *Evidence-Based Guidelines for Treating Depressive Disorders With Antidepressants: A Revision of the 2000 British Association for Psychopharmacology Guidelines.* London, England: SAGE Publications Ltd; 2008:343-396.

10. Elkin I, Gibbons RD, Shea MT, et al. Initial severity and differential treatment outcome in the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *J Consult Clin Psychol.* 1995;63(5):841-847.

11. Rush AJ, Beck AT, Kovacs M, Hollon S. Comparative efficacy of cognitive therapy and pharmacotherapy in the treatment of depressed outpatients. *Cognit Ther Res.* 1977;1(1):17-37.

12. Murphy GE, Simons AD, Wetzel RD, Lustman PJ. Cognitive therapy and pharmacotherapy: singly and together in the treatment of depression. *Arch Gen Psychiatry.* 1984;41(1):33-41.

13. Hollon SD, DeRubeis RJ, Evans MD, et al. Cognitive therapy and pharmacotherapy for depression: singly and in combination. *Arch Gen Psychiatry.* 1992;49(10):774-781.

14. DeRubeis RJ, Hollon SD, Amsterdam JD, et al. Cognitive therapy vs medications in the treatment of moderate to severe depression. *Arch Gen Psychiatry.* 2005;62(4):409-416.

15. Siddique J, Chung JY, Brown CH, Miranda J. Comparative effectiveness of medication versus cognitive-behavioral therapy in a randomized controlled trial of low-income young minority women with depression. *J Consult Clin Psychol.* 2012;80(6):995-1006.

16. Driessen E, Cuijpers P, Hollon SD, Dekker JJM. Does pretreatment severity moderate the efficacy of psychological treatment of adult outpatient depression? a meta-analysis. *J Consult Clin Psychol.* 2010;78(5):668-680.

17. Fournier JC, DeRubeis RJ, Hollon SD, et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA.* 2010; 303(1):47-53.

18. DeRubeis RJ, Gelfand LA, Tang TZ, Simons AD. Medications versus cognitive behavior therapy for severely depressed outpatients: mega-analysis of four randomized comparisons. *Am J Psychiatry.* 1999;156(7):1007-1013.

19. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ.* 2010;340: c221.

20. Cuijpers P, Berking M, Andersson G, Quigley L, Kleiboer A, Dobson KS. A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. *Can J Psychiatry.* 2013;58(7):376-385.

21. Higgins JP, Green S, eds. Chapter 8: assessing risk of bias in included studies. In: Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0. http://handbook.cochrane.org/. Updated March 2011. Accessed August 10, 2015.

22. Borenstein M, Hedges LV, Rothstein HR. *Introduction to Meta-analysis.* Sussex, England: Wiley Inc; 2009.

23. Hedges L, Olkin I. *Statistical Methods for Meta-analysis.* Orlando, FL: Academic Press; 1985.

24. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics.* 2000;56(2):455-463.

25. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ.* 1997;315(7109):629-634.

26. Hamilton M. A rating scale for depression. *J Neurol Neurosurg Psychiatry.* 1960;23:56-62.

27. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry.* 1961;4:561-571.

28. Beck AT, Steer RA, Brown GK. *The Beck Depression Inventory.* 2nd ed. San Antonio, TX: Psychological Corp; 1996.

29. Dunner DL, Schmaling KB, Hendrickson H, Becker J, Lehman A, Bea C. Cognitive therapy versus fluoxetine in the treatment of dysthymic disorder. *Depression.* 1996;4(1):34-41.

30. Faramarzi M, Alipor A, Esmaelzadeh S, Kheirkhah F, Poladi K, Pash H. Treatment of depression and anxiety in infertile women: cognitive behavioral therapy versus fluoxetine. *J Affect Disord.* 2008;108(1-2):159-164.

31. Donders AR, van der Heijden GJ, Stijnen T, Moons KG. Review: a gentle introduction to imputation of missing values. *J Clin Epidemiol.* 2006;59(10):1087-1091.

32. Bower P, Kontopantelis E, Sutton A, et al. Influence of initial severity of depression on effectiveness of low intensity interventions: meta-analysis of individual patient data. *BMJ.* 2013; 346:f540.

33. Stewart LA, Parmar MK. Meta-analysis of the literature or of individual patient data: is there a difference? *Lancet.* 1993;341(8842):418-422.

34. Vickers AJ, Cronin AM, Maschino AC, et al; Acupuncture Trialists' Collaboration. Individual patient data meta-analysis of acupuncture for chronic pain: protocol of the Acupuncture Trialists' Collaboration. *Trials.* 2010;11:90.

35. Frank E, Prien RF, Jarrett RB, et al. Conceptualization and rationale for consensus definitions of terms in major depressive disorder: remission, recovery, relapse, and recurrence. *Arch Gen Psychiatry.* 1991;48(9):851-855.

36. National Institute for Clinical Excellence. *The Treatment and Management of Depression in Adults.* London, England: National Institute for Clinical Excellence; 2009.

37. Miranda J, Chung JY, Green BL, et al. Treating depression in predominantly low-income young minority women: a randomized controlled trial. *JAMA.* 2003;290(1):57-65.

38. Bagby RM, Ryder AG, Schuller DR, Marshall MB. The Hamilton Depression Rating Scale: has the gold standard become a lead weight? *Am J Psychiatry.* 2004;161(12):2163-2177.

39. Wisniewski SR, Rush AJ, Nierenberg AA, et al. Can phase III trial results of antidepressant medications be generalized to clinical practice? a STAR*D report. *Am J Psychiatry.* 2009;166(5): 599-607.

40. Cuijpers P, Hollon SD, van Straten A, Bockting C, Berking M, Andersson G. Does cognitive behaviour therapy have an enduring effect that is superior to keeping patients on continuation pharmacotherapy? a meta-analysis. *BMJ Open.* 2013;3(4):e002542. doi:10.1136/bmjopen-2012 -002542.

41. Blackburn IM, Bishop S, Glen AI, Whalley LJ, Christie JE. The efficacy of cognitive therapy in depression: a treatment trial using cognitive therapy and pharmacotherapy, each alone and in combination. *Br J Psychiatry.* 1981;139:181-189.

42. Blackburn IM, Moore RG. Controlled acute and follow-up trial of cognitive therapy and pharmacotherapy in out-patients with recurrent depression. *Br J Psychiatry.* 1997;171:328-334.

43. Hautzinger M, de Jong-Meyer R, Treiber R, Rudolf GA, Thien U. Wirksamkeit Kognitiever Verhaltenstherapie, Pharmakotherapie und deren Kombination bei nicht-endogenen, unipolaren Depressionen. *Zeitschr Klin Psychol.* 1996;25:130-145.

44. Marshall MB, Zuroff DC, McBride C, Bagby RM. Self-criticism predicts differential response to treatment for major depression. *J Clin Psychol.* 2008;64(3):231-244.

45. McKnight DL, Nelson-Gray RO, Barnhill J. Dexamethasone suppression test and response to cognitive therapy and antidepressant medication. *Behav Ther.* 1992;23:99-111.

46. Scott AI, Freeman CP. Edinburgh primary care depression study: treatment outcome, patient satisfaction, and cost after 16 weeks. *BMJ.* 1992; 304(6831):883-887.

47. Shamsaei F, Rahimi A, Zarabian MK, Sedehi M. Efficacy of pharmacotherapy and cognitive therapy, alone and in combination in major depressive disorder. *Hong Kong J Psychiatry.* 2008;18:76-80.

48. Thompson LW, Coon DW, Gallagher-Thompson D, Sommer BR, Koin D. Comparison of desipramine and cognitive/behavioral therapy in the treatment of elderly outpatients with mild-to-moderate depression. *Am J Geriatr Psychiatry.* 2001;9(3):225-240.