Antidepressant Drug Effects and Depression Severity: A Patient-Level Meta-analysis

Jay C. Fournier; Robert J. DeRubeis; Steven D. Hollon; et al.

JAMA. 2010;303(1):47-53 (doi:10.1001/jama.2009.1943)

http://jama.ama-assn.org/cgi/content/full/303/1/47

REVIEW

Antidepressant Drug Effects and Depression Severity

A Patient-Level Meta-analysis

Jay C. Fournier, MA
Robert J. DeRubeis, PhD
Steven D. Hollon, PhD
Sona Dimidjian, PhD
Jay D. Amsterdam, MD
Richard C. Shelton, MD
Jan Fawcett, MD

Context Antidepressant medications represent the best established treatment for major depressive disorder, but there is little evidence that they have a specific pharmacological effect relative to pill placebo for patients with less severe depression.

Objective To estimate the relative benefit of medication vs placebo across a wide range of initial symptom severity in patients diagnosed with depression.

Data Sources PubMed, PsycINFO, and the Cochrane Library databases were searched from January 1980 through March 2009, along with references from meta-analyses and reviews.

Study Selection Randomized placebo-controlled trials of antidepressants approved by the Food and Drug Administration in the treatment of major or minor depressive disorder were selected. Studies were included if their authors provided the requisite original data, they comprised adult outpatients, they included a medication vs placebo comparison for at least 6 weeks, they did not exclude patients on the basis of a placebo washout period, and they used the Hamilton Depression Rating Scale (HDRS). Data from 6 studies (718 patients) were included.

Data Extraction Individual patient-level data were obtained from study authors.

Results Medication vs placebo differences varied substantially as a function of baseline severity. Among patients with HDRS scores below 23, Cohen *d* effect sizes for the difference between medication and placebo were estimated to be less than 0.20 (a standard definition of a small effect). Estimates of the magnitude of the superiority of medication over placebo increased with increases in baseline depression severity and crossed the threshold defined by the National Institute for Clinical Excellence for a clinically significant difference at a baseline HDRS score of 25.

Conclusions The magnitude of benefit of antidepressant medication compared with placebo increases with severity of depression symptoms and may be minimal or nonexistent, on average, in patients with mild or moderate symptoms. For patients with very severe depression, the benefit of medications over placebo is substantial.

JAMA. 2010;303(1):47-53 www.jama.com

ANTIDEPRESSANT MEDICATION (ADM) represents the current standard of treatment for major depressive disorder (MDD).[1] Antidepressant medication has been shown to be superior to placebo in thousands of controlled clinical trials over the past 5 decades.[2,3] The extent to which ADM outperforms placebo (which controls for nonpharmacological aspects of ADM) can be used to index the "true" pharmacological effect of ADM in clinical settings.

The randomized, double-blind, placebo-controlled trial is the gold standard for testing treatment efficacy and affords the opportunity to identify patient characteristics that predict differential pharmacological response. Baseline symptom severity is one dimension that may affect treatment outcome. Kirsch et al[4] and Khan et al[5] presented independent meta-analyses of randomized placebo-controlled trials based on data from the Food and Drug Administration (FDA) clinical trial database. Using mean scores and standard deviations on the Hamilton Depression Rating Scale (HDRS)[6] from each study, they examined the effect of baseline symptom severity on the relative efficacy of ADM vs placebo. Kirsch et al found that as the mean baseline HDRS score increased, the magnitude of HDRS change decreased for placebo but remained unchanged for ADM. Khan et al did not find a significant relationship between baseline scores and symptom change for the

Author Affiliations: Departments of Psychology (Mr Fournier and Dr DeRubeis) and Psychiatry (Dr Amsterdam), University of Pennsylvania, Philadelphia; Departments of Psychology (Dr Hollon) and Psychiatry (Dr Shelton), Vanderbilt University, Nashville, Tennessee; Department of Psychology, University of Colorado at Boulder (Dr Dimidjian); and Department of Psychiatry, University of New Mexico School of Medicine, Albuquerque (Dr Fawcett).
Corresponding Author: Jay C. Fournier, MA, Department of Psychology, University of Pennsylvania, 3720 Walnut St, Philadelphia, PA 19104 (jcf@sas.upenn.edu).

placebo condition but found greater symptom change in ADM as baseline HDRS scores increased. Thus, both studies found that the greater the baseline symptom severity, the greater the magnitude of the difference favoring ADM over placebo. Kirsch et al inferred from their findings that the minimum baseline HDRS score needed to achieve a clinically meaningful ADM/placebo difference is approximately 28 and that differences are negligible for lower baseline HDRS scores.

One limitation to these meta-analyses is the restricted range of baseline severity scores included in their constituent studies. In the analysis by Kirsch et al,[4] only 1 of 35 studies comprised samples with mean baseline HDRS scores lower than 23. As the authors noted, a score of 23 is characteristic of "very severe depression" according to the American Psychiatric Association's *Handbook of Psychiatric Measures* (which defines mild depression as HDRS scores from 8-13, moderate depression from 14-18, severe depression from 19-22, and very severe depression as ≥23).[7] Similarly, each of the studies included by Khan et al[5] required a minimum entry score of 20 on the HDRS, meaning that all patients could be classified as severe or very severe. It is likely that a sizable proportion of depressed individuals who start ADM in the community present with severity levels well below this value. In fact, a recent survey of depressed, treatment-seeking outpatients found that 71% of the 503 patients assessed had HDRS scores less than 22.[8] There has been a paucity of systematic investigations of the true effect of ADM in patients with less severe depression. Such data are scarce in the FDA database and in the published literature. This is partly the result of the inclusion criteria used for many FDA registration trials in which cutoff scores are imposed at baseline expressly to increase the sensitivity of ADM/placebo comparisons.

A second limitation of the Kirsch et al and Khan et al meta-analyses is that each included studies that used a placebo washout period. Typically, placebo washouts last from several days to 2 weeks, during which patients are administered a pill placebo in single-blind fashion. At the end of this period, patients who demonstrate an improvement of a particular magnitude (typically ≥20% on the HDRS) are excluded from the trial prior to randomization. The goal of this procedure is to increase the power to detect differences in efficacy between ADM and placebo by removing known placebo responders at the outset. Although it is not clear that placebo washouts actually enhance the statistical power of ADM/placebo comparisons,[9,10] this design feature severely limits the ability to generate accurate estimates of the placebo response rate. Because early placebo responders are removed from the trial before they can contribute data, the true rate of placebo response may be underestimated in trials that use this feature.

In the present study, we combined data from 6 large-scale, placebo-controlled trials that comprised patients with a broad range of baseline symptom severity.[11-16] Because most MDD studies incorporate a minimum baseline depressive severity score as an inclusion criterion, studies of minor depressive disorder (which do not typically have such strict thresholds) were included in this analysis as well. The entry criteria allowed patients to enter these studies with HDRS scores that ranged from the low teens to the upper 30s.[11-16] Unlike the data analyzed by Kirsch et al and Khan et al, which contained information only at the level of treatment group and thus could support only standard meta-analytic procedures, the databases from the 6 studies included in the present investigation provided data for a patient-level meta-analysis, also known as a mega-analysis. This approach is more appropriate and more powerful than a standard meta-analysis when original data are available and a fine-grained multivariate analysis is desired.[17] Based on the findings of Kirsch et al and Khan et al, we hypothesized that ADM/placebo differences would become larger as baseline severity increased.

METHODS

English-language articles from January 1980 through March 2009 were searched in the electronic databases PubMed and PsycINFO using the following search criteria: *antidepres** and *randomiz** and *placebo* and *depression* and (*treatment* or *trial*). The Cochrane Library was searched using the following terms as key words: *antidepres** and *placebo* and *depression*. No further restrictions were imposed on either search. We also examined the reference sections of meta-analyses and reviews to identify relevant randomized controlled trials.

The criteria for inclusion required studies to be randomized placebo-controlled trials of an FDA-approved antidepressant in the treatment of the full range of patients with major or minor depressive disorder (ie, studies that exclusively examined special populations or subtypes were excluded as were studies that exclusively examined patients diagnosed solely with dysthymia). The studies were restricted to adult outpatient samples; those that included children or adolescents below the age of 18 years were excluded. In addition, the studies had to include an ADM/placebo comparison of at least 6 weeks' duration and HDRS scores at intake and at the end of treatment. Studies were excluded if they excluded patients on the basis of a placebo washout period. The final inclusion criterion was that individual patient-level data had to be available for analysis.

Article Selection and Data Acquisition

The initial screening of the search results was supervised (S.D. and J.C.F.) and reviewed (J.C.F.) to ensure accuracy. All selected articles were read by 2 authors (J.C.F. and either S.D. or S.D.H.) to determine whether they met inclusion criteria (with an average κ of 0.82). Discrepancies were resolved by consensus.

The corresponding authors of studies meeting the inclusion criteria were contacted to verify that the study did not exclude patients on the basis of a placebo washout period and to ascertain whether individual patient-level data were available. Authors were initially asked to respond within 3 weeks, and additional time was provided to allow those making a positive response the opportunity to provide the requested data. FIGURE 1 displays the results of the search and data acquisition strategies.

Participants

The sample consisted of participants from the ADM and pill-placebo conditions of 5 MDD trials—DeRubeis et al,[12] Dimidjian et al,[13] Elkin et al,[14] Philipp et al,[15] Wichers et al[16]—and 1 minor depression trial, Barrett et al.[11] Full descriptions of the study designs, sample characteristics, treatment protocols, and primary outcome findings have been reported elsewhere.[11-16] Three studies used the tricyclic antidepressant imipramine[14-16] and 3 used the selective serotonin reuptake inhibitor paroxetine.[11-13] TABLE 1 lists characteristics that differ among the 6 studies. The pooled sample used in the current analyses included 434 patients in the ADM group and 284 patients in the placebo group. Individual baseline HDRS depression severity levels ranged from 10 to 39. In comparison with the 17 identified studies for which data were not available, the 6 included studies tended to have Jadad quality scores at the higher end of the range, to use flexible (as opposed to fixed) medication doses, and to provide more information about the samples in the original report (eTable, available at http://www.jama.com).

Statistical Analyses

Our primary statistical analysis investigated the relationship between baseline symptom severity and subsequent symptom change from intake to the end of acute treatment. We used a modified intent-to-treat approach whereby we used the most inclusive sample analyzed in the original publication of each of the 6 studies (Table 1). To investigate the association between initial severity and symptom change scores in ADM vs placebo, we conducted analyses of covariance that controlled for the effect of the study from which the data originated. For individuals who dropped out of treatment, we used the patient's last score prior to dropout (last observation carried forward) to calculate the change score. Continuous variables were centered at their grand means, and nonsignificant higher-order interaction terms were removed from the models. Level of significance was set at $P < .05$.

RESULTS

Study Characteristics

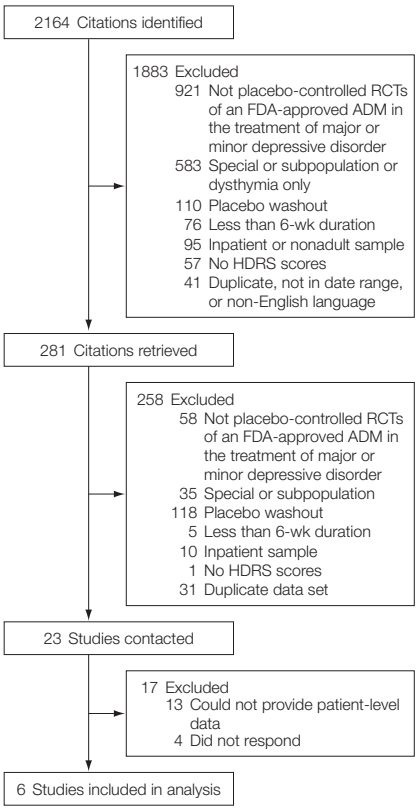
Mean baseline depression severity scores and attrition rates for the 6 studies are displayed in TABLE 2. A 2×6 (treatment × study) analysis of variance was conducted to examine differences in levels of intake depression severity. The study × treatment interaction was not significant and was removed from the model. Mean intake severity did not differ as a function of treatment condition ($F_{1,711} = 0.05$, $P = .82$), but the 6 studies did show different mean intake severity levels, reflecting differences in inclusion criteria ($F_{5,711} = 79.56$, $P < .001$). Attrition rates were compared in a logistic regression model examining the effects of study, treatment, and the study × treatment interaction. The study × treatment interaction term was not significant and was removed from the model. Attrition rates did not differ significantly as a function of treatment condition ($\chi^2_1 = 0.47$, $P = .49$), but differences did emerge in the rates of attrition among the 6 studies ($\chi^2_5 = 30.34$, $P < .001$) (Table 2).

Baseline Severity and Symptom Change in ADM and Placebo

Pooling the data across the 6 studies, the severity × treatment interaction (the statistic of primary interest in this investigation) was significant in a model that predicted depression change scores controlling for study of origin ($F_{1,709} = 9.31$, $P = .002$). The main effects of baseline severity ($F_{1,709} = 59.54$, $P < .001$) and treatment ($F_{1,709} = 12.51$, $P < .001$) were also significant.

As displayed in FIGURE 2, the regression coefficient (ie, the slope representing the relation between initial severity and change in symptoms) was positive for both ADM ($b = 0.70$, $t_{709} = 8.49$, $P < .001$) and placebo ($b = 0.36$, $t_{709} = 3.87$, $P < .001$). The difference in the slopes of the 2 regression lines, $b = 0.34$, represents the interaction effect described earlier in this section. The 2 regression lines converged near the lower end of the range of baseline severity scores

Figure 1. Study Selection and Data Acquisition



Reasons for exclusion describe the first reason for exclusion that was encountered during the review process. Several articles had multiple reasons for exclusion. RCTs indicates randomized controlled trials; FDA, US Food and Drug Administration; ADM, antidepressant medication; HDRS, Hamilton Depression Rating Scale.

Table 1. Differences Between 6 Studies of Medications and Placebo for Depressed Outpatients

Characteristic	Barrett et al[11]	DeRubeis et al[12]	Dimidjian et al[13]	Elkin et al[14]	Philipp et al[15]	Wichers et al[16]
Disorder examined	Minor[a]	MDD	MDD	MDD	MDD	MDD
No. of intake evaluations[b]	1	2	1	2	1	2
No. of treatment sites	2	2	1	3	18	8
Medication used	Paroxetine	Paroxetine	Paroxetine	Imipramine	Imipramine[c]	Imipramine
Target dose, mg/d	20-40	50	50	150-250	100	100-200
Blinded evaluations[d]	Yes	Yes	Partial	Yes	Partial	Partial
HDRS version	17-Item	Modified 17-item	Modified 17-item	Modified 17-item	17-Item	17-Item
Minimum intake severity[e]	≥10	≥20	≥14	≥14	≥18	≥18
Sample analyzed[f]	F-ITT	F-ITT	M-ITT	M-ITT	M-ITT	Complied with protocol
Treatment duration, wk[g]	11	8	8	8	8	6

Abbreviations: F-ITT, full intent to treat; HDRS, Hamilton Depression Rating Scale; MDD, major depressive disorder; Minor, minor depressive disorder; M-ITT, modified intent to treat.
[a] The Barrett et al study also included patients diagnosed with dysthymia. These patients were not included in the current analyses.
[b] The Elkin et al, DeRubeis et al, and Wichers et al trials required participants to meet inclusion criteria in each of 2 consecutive evaluations that were held at least 1 week apart.
[c] The Philipp et al trial also included a *Hypericum* extract condition. Data from this condition were not included in the current analyses.
[d] "Yes" indicates that independent blind evaluators conducted evaluations of symptom severity at every assessment. "Partial" indicates that evaluations were conducted at each session by the treating pharmacotherapists. Treating pharmacotherapists were blind to treatment condition.
[e] Six patients from the Elkin et al sample registered scores less than 14 on the HDRS at intake (2 from the imipramine and 4 from the placebo conditions) and 1 patient from the DeRubeis et al trial registered a score less than 20 (in the paroxetine condition). These patients were retained in the present analyses.
[f] The Barrett et al and DeRubeis et al studies used a full intent-to-treat design whereby all patients randomized to treatment were included in the analysis. The Dimidjian et al, Elkin et al, and Philipp et al studies used a modified intent-to-treat approach whereby data from only those patients who attended at least 1 treatment session or who had 1 postbaseline score were included. The Wichers et al study included only those patients who met minimum compliance requirements for a protocol from a related research question (this sample did include treatment dropouts).
[g] Treatment in the Elkin et al trial was provided for 16 weeks. Because target doses were reached by the 8-week assessment, only data through week 8 were analyzed to improve comparability between the studies.

Table 2. Sample Size, Dropout Rates, and Baseline Depression Severity in 6 Studies That Compared Active Antidepressant Medications With Pill Placebo

Study	No. of Patients	Dropouts, No. (%)	Baseline HDRS Score, Mean (SD)
Barrett et al[11,a]			
Placebo	39	8 (21)	14.38 (3.98)
Medication	38	13 (34)	14.05 (2.86)
DeRubeis et al[12,b]			
Placebo	60	9 (15)	23.47 (2.73)
Medication	120	13 (11)	23.22 (2.70)
Dimidjian et al[13,b]			
Placebo	45	4 (9)	20.76 (4.59)
Medication	82	15 (18)	20.73 (3.94)
Elkin et al[14,a]			
Placebo	62	17 (27)	19.47 (4.60)
Medication	57	13 (23)	19.51 (4.62)
Philipp et al[15,c]			
Placebo	46	3 (7)	22.20 (4.15)
Medication	105	3 (3)	22.70 (4.03)
Wichers et al[16,a,b]			
Placebo	32	5 (16)	23.72 (2.81)
Medication	32	9 (28)	24.38 (3.84)

Abbreviation: HDRS, Hamilton Depression Rating Scale.
[a] Barrett et al, Elkin et al, and Wichers et al did not differ with respect to mean rates of attrition ($P < .05$) and comprised the group with the highest attrition rates.
[b] DeRubeis et al, Dimidjian et al, and Wichers et al did not differ with respect to mean rates of attrition ($P < .05$) and comprised the group with the second lowest rates. DeRubeis et al and Dimidjian et al differed from Barrett et al and Elkin et al, whereas Wichers et al did not.
[c] Philipp et al had the lowest dropout rate. It differed from all of the other studies.

and the magnitude of the difference between the treatments increased with increasing baseline depression severity. To illustrate the magnitude of the difference between the 2 treatments as a function of initial depression severity, we divided the sample into 3 groups based on the characterizations of the HDRS scores offered by the American Psychiatric Association: mild to moderate, HDRS score of 18 or less (n=180); severe, HDRS score of 19 to 22 (n=255); and very severe, HDRS score of 23 or greater (n=283).[7] For patients in the mild to moderate range, the Cohen *d* effect size was $d=0.11$ (95% confidence interval, [CI], –0.18 to 0.41) and for patients in the severe range, $d=0.17$ (95% CI, –0.08 to 0.43). Both values were below the standard description of a small effect ($d=0.20$).[18] For patients in the very severe group, $d=0.47$ (95% CI, 0.22 to 0.71). This value was just below 0.50, the accepted cutoff for a medium effect size. We also converted these *d* effect sizes into estimates of the number of patients needed to treat (NNT) to increase by 1 the number of patients in the treatment group who would have a better outcome than a randomly selected patient from the control group.[19] Number-needed-to-treat values were estimated to be 16, 11, and 4 for the mild to moderate, severe, and very severe subgroups, respectively.

The National Institute for Clinical Excellence (NICE) of the National Health Service in England has defined

a threshold for clinical significance as an effect size of 0.50 or a drug/placebo difference of 3 points on the HDRS.[20] Using least-squares means from the primary model described earlier in this section, we estimated that this threshold was met for intake HDRS scores of 25 or greater, using the more liberal of the 2 criteria (a difference in HDRS scores of ≥3 points). To examine the more conservative threshold defined by d=0.50, we estimated Cohen d effect sizes, again using least-squares means estimates from the primary model. Drug/placebo differences were estimated to cross this threshold at an initial HDRS score of 27 (NNT=4). When we divided the sample into subgroups using these 2 thresholds, the superiority of medications over placebo was associated with a medium-sized effect for patients with HDRS scores of 25 or greater (d=0.53; 95% CI, 0.19 to 0.86) and a large effect for patients with HDRS of 27 or greater (d=0.81; 95% CI, 0.30 to 1.32).

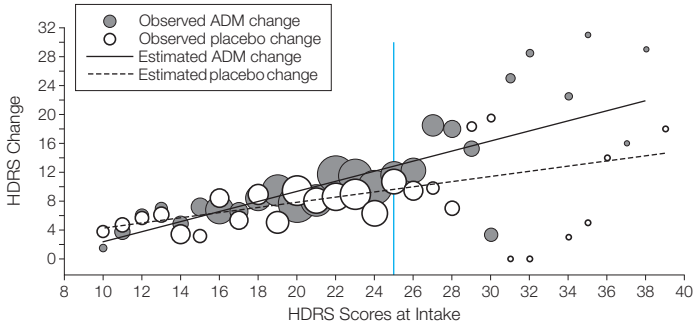
Baseline Severity and Symptom Change for Patients With MDD

To determine whether the pattern of results reported was evident in patients diagnosed with MDD, data from the Barrett et al[11] study of minor depressive disorder were removed and the models were rerun. The severity × treatment interaction was again significant ($F_{1,633}$=6.93, P=.009). As before, the ADM/placebo difference was estimated to cross the NICE criteria at an initial baseline HDRS score of 25.

Baseline Severity and Symptom Change for Completers

To assess whether attrition might have biased the results, the primary analyses were repeated in a completers-only sample. Again the severity × treatment interaction was significant ($F_{1,597}$=5.62, P=.02). Among completers, the difference between ADM and placebo was estimated to cross the NICE threshold at an initial HDRS score of 24 (1 point lower than that observed for the entire sample). We also repeated the primary analysis using data only from the 3 studies with the lowest dropout rates.[12,13,15] Again, the interaction of interest was significant ($F_{1,452}$=6.98, P<.01).

Figure 2. Observed and Estimated Change in HDRS Scores Following Treatment With ADM and Placebo



Circles represent observed (raw) mean change in depressive symptoms from intake to the end of treatment at each initial Hamilton Depression Rating Scale (HDRS) score for both the antidepressant medication (ADM) and placebo conditions. The size (area) of the circles is proportional to the number of data points that contributed to each mean. Regression lines represent estimates of change in depression symptoms from intake to end of treatment for ADM and placebo conditions as a function of baseline symptom severity. These regression lines were estimated from a model of the baseline severity × treatment interaction, controlling for the effects of the study from which the data originated. The National Institute for Clinical Excellence threshold for clinical significance (an HDRS point difference ≥3) was met for intake HDRS scores of 25 or greater, indicated by the blue line.

Drug Class

Three of the studies used the selective serotonin reuptake inhibitor paroxetine as the active ADM, whereas the other 3 studies used the tricyclic antidepressant imipramine. To investigate whether baseline severity moderates treatment response in both drug classes, we conducted a secondary analysis in which we replaced the term representing ADM/placebo with a categorical variable representing medication type. As in the primary analysis, the severity × drug class interaction was significant ($F_{2,707}$=4.41, P=.01). Specific contrasts revealed that the regression coefficient (ie, the slope representing the relationship between initial severity and change in symptoms) was more positive for each medication class relative to placebo: imipramine, $F_{1,707}$=5.60, P=.02, and paroxetine, $F_{1,707}$=5.91, P=.02.

COMMENT

The present findings indicate that the efficacy of ADM treatment for depression varies considerably as a function of symptom severity. True drug effects (an advantage of ADM over placebo) were nonexistent to negligible among depressed patients with mild, moderate, and even severe baseline symptoms, whereas they were large for patients with very severe symptoms. For baseline severity scores on the HDRS less than 25, estimates of the magnitudes of drug/placebo differences did not meet either of the 2 thresholds for clinical significance proposed by NICE.[20] Conversely, for patients with the highest levels of baseline depression severity, ADM was markedly superior to placebo.

As documented in the analysis by Zimmerman et al[8] of published efficacy trials, as well as in the analyses by Kirsch et al[4] and Khan et al[5] of studies submitted to the FDA, evidence concerning the effects of ADM in patients with mild and moderate MDD has been sparse. Our findings add substantially to knowledge of the effects of ADM across the full range of symptom severity in patients diagnosed with depression. These findings are consistent with an understanding that has informed the entry criteria used in ADM registration trials, in which cutoff scores of 18 or greater typically have been imposed. As noted by Zimmerman et al,

using such cutoffs can be expected to exclude nearly half of all patients who meet diagnostic criteria for MDD.

We note several limitations of the present inquiry. First, all of the studies used in the current investigation imposed a minimum baseline severity criterion. Because only a small proportion of the patients registered baseline HDRS scores of 13 or lower, the results of the current investigation may not generalize to such individuals. Second, when a minimum score at intake is required for study entry, study diagnosticians sometimes inadvertently inflate the scores of patients whose true score is just below the cutoff.[21] We have no evidence that this occurred in the current data sets, but if it did, it should have worked against the hypothesis that severity moderates outcome. Moreover, the inclusion of studies with different minimum severity levels should have mitigated any bias that such rater inflation might have caused. Third, scores on the HDRS were used as the primary outcome measure for all analyses. The HDRS has been the most commonly used measure of depression symptom severity in clinical trials of ADM, but the measure's psychometric properties have been criticized.[22,23] Future efforts might use alternative symptom measures to examine the effects of baseline severity on treatment outcome. Fourth, because few studies in the literature report the magnitude of the baseline severity × treatment interaction effect, it is difficult to assess the role of publication bias in this report. For a detailed account of publication bias regarding the main effect of ADM, see Turner et al.[24] Finally, the results reported herein apply to acute treatment only and not to continuation or maintenance treatment.

Despite differences in methods, our findings are consistent with those of both Kirsch et al[4] and Khan et al[5] that ADM/placebo differences increase as initial severity increases. We used individual patient data and included patients with less severe depression, whereas both Kirsch et al and Khan et al analyzed group means that largely excluded patients with HDRS scores below 20. Moreover, both Kirsch et al and Khan et al included studies that screened out pill-placebo responders prior to randomization, whereas the studies from which our data were drawn did not.

Given these differences, the consistency of the primary finding across the 3 reviews is striking. However, there also were subtle differences in the pattern of findings across the 3 investigations that likely reflect additional differences in methodology. For example, using within-group effect sizes, Kirsch et al found that initial severity was unrelated to outcome among patients treated with ADM but negatively related to outcome among placebo patients, whereas using between-group comparisons, Khan et al found that initial severity predicted greater symptom change among ADM patients (as did we using individual patient data) but was unrelated with respect to placebo patients (whereas we found a small positive relationship). Given these inconsistencies, it would be premature to speculate regarding whether the increasing superiority of ADM relative to placebo as severity increases is due to an increasing efficacy of ADM or a declining efficacy of placebo. Such a distinction depends, in part, on the index of change that is chosen.

Several studies have demonstrated that ADM is superior to placebo for patients diagnosed with dysthymia, a condition partly defined by lower symptom levels relative to MDD.[25,26] The dysthymia studies indicate that ADM can produce a true drug effect in patients with mild or moderate depressive symptoms. However, dysthymia is by definition a chronic condition, and chronicity is known to be associated with poor response to placebo.[27,28] Thus, it may be the chronic nature of dysthymia that explains the advantage of ADM over placebo in this condition. Future work should examine whether chronicity moderates ADM/placebo differences across the range of baseline severity.

The general pattern of results reported in this work is not surprising. As early as the 1950s, researchers conducting controlled investigations of treatments for a wide variety of medical and psychiatric conditions described a phenomenon whereby patients with higher levels of severity showed greater differential (ie, specific) benefit from the active treatments.[29,30] What makes our findings surprising is the high level of depression symptom severity that appears to be required for clinically meaningful drug/placebo differences to emerge, particularly given the evidence that the majority of patients receiving ADM in clinical practice present with scores below these levels.

Prescribers, policy makers, and consumers may not be aware that the efficacy of medications largely has been established on the basis of studies that have included only those individuals with more severe forms of depression. This important feature of the evidence base is not reflected in the implicit messages present in the marketing of these medications to clinicians and the public. There is little mention of the fact that efficacy data often come from studies that exclude precisely those MDD patients who derive little specific pharmacological benefit from taking medications. Pending findings contrary to those reported here and those obtained by Kirsch et al and Khan et al, efforts should be made to clarify to clinicians and prospective patients that whereas ADM can have a substantial effect with more severe depressions, there is little evidence to suggest that they produce specific pharmacological benefit for the majority of patients with less severe acute depressions.

Author Contributions: Mr Fournier had full access to all of 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: Fournier, DeRubeis, Hollon, Dimidjian, Amsterdam, Shelton, Fawcett.
Acquisition of data: Fournier, DeRubeis, Hollon, Dimidjian, Amsterdam, Shelton, Fawcett.
Analysis and interpretation of data: Fournier, DeRubeis, Hollon, Dimidjian, Amsterdam, Shelton, Fawcett.

Drafting of the manuscript: Fournier, DeRubeis, Hollon.
Critical revision of the manuscript for important intellectual content: Fournier, DeRubeis, Hollon, Dimidjian, Amsterdam, Shelton, Fawcett.
Statistical analysis: Fournier, DeRubeis.
Obtained funding: DeRubeis, Hollon, Amsterdam.
Administrative, technical, or material support: Fournier, DeRubeis, Hollon, Dimidjian, Amsterdam, Shelton, Fawcett.
Financial Disclosures: Dr Amsterdam reported serving on the speakers' bureau of Wyeth Pharmaceuticals and Bristol Myers Squibb; receiving research support from Novartis, Eli Lilly, Sanofi, Cephalon, and Forest Laboratories; and serving as a consultant for Bristol Myers Squibb. Dr Shelton reported serving as a consultant to AstraZeneca, Eli Lilly, Evotec, Forest Pharmaceuticals, Gideon Richter, Janssen Pharmaceuticals, Merck, Novartis Pharmaceuticals, Ostuka Pharmaceuticals, Pamlab, Pfizer, Repligen, Sierra Neuropharmaceuticals, and Wyeth; receiving speaking honoraria from AstraZeneca, Eli Lilly, Forest Pharmaceuticals, GlaxoSmithKline, Pamlab, Pfizer, and Wyeth; and receiving research and/or grant support from Bristol Myers Squibb, Eli Lilly, Evotec, Forest Pharmaceuticals, GlaxoSmithKline, Janssen Pharmaceuticals, Novartis Pharmaceuticals, Ostuka Pharmaceuticals, Pamlab, Pfizer, Repligen, and Wyeth. Dr Fawcett reported serving as a consultant to Abbott Laboratories, Merck, and Slack; receiving speaking honoraria from Eli Lilly; and serving as a board member for the Berman Center and on the scientific advisory boards for the nonprofit advocacy organizations NARSAD and the Depression and Bipolar Support Alliance. Dr Fawcett also reported providing expert testimony on cases involving pharmaceutical companies including Banner Health and Alphapharm and currently chairing the Mood Disorders Work Group for the forthcoming revision of the *Diagnostic and Statistical Manual of Mental Disorders* (Fifth Edition). No other disclosures were reported.
Funding/Support: This research was supported by grants MH50129 (R10), MH55875 (R10), MH01697 (K02), MH01741 (K24), and MH060998 (R01) from the National Institute of Mental Health, Bethesda, Maryland.
Role of the Sponsor: The funding sources had no role in the design and conduct of the study; in the collection, analysis, and interpretation of the data; or in the preparation, review, or approval of the manuscript.
Online-Only Material: The eTable is available at http://www.jama.com.
Additional Contributions: We thank the authors who shared their data with us for this project, with special thanks to Marieke Wichers, PhD, Maastricht University, School for Mental Health and Neuroscience; John Cornell, PhD, University of Texas Health Science Center; and Karl-O. Hiller, PhD, Steiner Arzneimittel, for providing us with the raw data from their respective studies. Finally, we thank all those who responded to our inquiries, even if data from their studies could not be made available. None were compensated for their contributions.

REFERENCES

1. Hollon SD, Thase ME, Markowitz JC. Treatment and prevention of depression. *Psychol Sci Public Interest*. 2002;3(2):39-77.
2. *Treatment of Depression: Newer Pharmacotherapies* [AHCPR publication No. 99-E014]. Mulrow CD, Williams JW Jr, Trivedi M, et al. Rockville, MD: US Dept of Health and Human Services; 1999.
3. *Depression in Primary Care: Vol 2, Treatment of Major Depression* [Clinical Practice Guideline No. 5; AHCPR Publication No. 93-0551]. Depression Guideline Panel. Rockville, MD: US Dept of Health and Human Services; 1993.
4. Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med*. 2008; 5(2):e45.
5. Khan A, Leventhal RM, Khan SR, Brown WA. Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. *J Clin Psychopharmacol*. 2002; 22(1):40-45.
6. Hamilton MA. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960;23:56-62.
7. American Psychiatric Association Task Force for the Handbook of Psychiatric Measures. *Handbook of Psychiatric Measures.* Washington, DC: American Psychiatric Association; 2000.
8. Zimmerman M, Posternak MA, Chelminski I. Symptom severity and exclusion from antidepressant efficacy trials. *J Clin Psychopharmacol*. 2002;22(6): 610-614.
9. Lee S, Walker JR, Jakul L, Sexton K. Does elimination of placebo responders in a placebo run-in increase the treatment effect in randomized clinical trials? a meta-analytic evaluation. *Depress Anxiety*. 2004; 19(1):10-19.
10. Trivedi MH, Rush J. Does a placebo run-in or a placebo treatment cell affect the efficacy of antidepressant medications? *Neuropsychopharmacology*. 1994;11(1):33-43.
11. Barrett JE, Williams JW Jr, Oxman TE, et al. Treatment of dysthymia and minor depression in primary care: a randomized trial in patients aged 18 to 59 years. *J Fam Pract*. 2001;50(5):405-412.
12. 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.
13. Dimidjian S, Hollon SD, Dobson KS, et al. Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *J Consult Clin Psychol*. 2006;74(4):658-670.
14. Elkin I, Shea MT, Watkins JT, et al. National Institute of Mental Health Treatment of Depression Collaborative Research Program: general effectiveness of treatments. *Arch Gen Psychiatry*. 1989;46(11): 971-982.
15. Philipp M, Kohnen R, Hiller KO. Hypericum extract versus imipramine or placebo in patients with moderate depression: randomised multicentre study of treatment for eight weeks. *BMJ*. 1999;319(7224): 1534-1538.
16. Wichers MC, Barge-Schaapveld DQ, Nicolson NA, et al. Reduced stress-sensitivity or increased reward experience: the psychological mechanism of response to antidepressant medication. *Neuropsychopharmacology*. 2009;34(4):923-931.
17. Steinberg KK, Smith SJ, Stroup DF, et al. Comparison of effect estimates from a meta-analysis of summary data from published studies and from a meta-analysis using individual patient data for ovarian cancer studies. *Am J Epidemiol*. 1997;145(10): 917-925.
18. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
19. Kraemer HC, Kupfer DJ. Size of treatment effects and their importance to clinical research and practice. *Biol Psychiatry*. 2006;59(11):990-996.
20. National Institute for Clinical Excellence. *Depression: Management of Depression in Primary and Secondary Care.* London, England: National Institute for Clinical Excellence; 2004.
21. Landin R, DeBrota DJ, DeVries TA, Potter WZ, Demitrack MA. The impact of restrictive entry criterion during the placebo lead-in period. *Biometrics*. 2000;56(1):271-278.
22. Zimmerman M, Posternak MA, Chelminski I. Is it time to replace the Hamilton Depression Rating Scale as the primary outcome measure in treatment studies of depression? *J Clin Psychopharmacol*. 2005; 25(2):105-110.
23. 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.
24. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med*. 2008;358(3):252-260.
25. Baldwin D, Rudge S, Thomas S. Dysthymia: options in pharmacotherapy. *CNS Drugs*. 1995;4: 422-431.
26. de Lima MS, Hotoph M, Wessely S. The efficacy of drug treatments for dysthymia: a systematic review and meta-analysis. *Psychol Med*. 1999;29 (6):1273-1289.
27. Dunner DL. Acute and maintenance treatment of chronic depression. *J Clin Psychiatry*. 2001;62 (suppl 6):10-16.
28. Khan A, Dager SR, Cohen S, Avery DH, Scherzo B, Dunner DL. Chronicity of depressive episode in relation to antidepressant-placebo response. *Neuropsychopharmacology*. 1991;4(2):125-130.
29. Benjamin LS. Statistical treatment of the law of initial values (LIV) in autonomic research: a review and recommendation. *Psychosom Med*. 1963;25(6): 556-566.
30. Fisher S, Lipman RS, Uhlenhuth EH, Rickels K, Park LC. Drug effects and initial severity of symptomatology. *Psychopharmacologia*. 1965;7(1):57-60.