



Published in final edited form as:

JAMA. 2010 January 6; 303(1): 47–53. doi:10.1001/jama.2009.1943.

Antidepressant Drug effects and Depression Severity: A Patient-Level Meta-Analysis

Jay C. Fournier, M.A., Robert J. DeRubeis, Ph.D, Steven D. Hollon, Ph.D., Sona Dimidjian, Ph.D., Jay D. Amsterdam, M.D., Richard C. Shelton, M.D., and Jan Fawcett, M.D.
Department of Psychology, University of Pennsylvania, Philadelphia, PA (JCF & RJD); Department of Psychology, Vanderbilt University, Nashville, TN (SDH); Department of Psychology, University of Colorado at Boulder (SD); Department of Psychiatry, University of Pennsylvania, Philadelphia, PA (JDA); Department of Psychiatry, Vanderbilt University, Nashville, TN (RCS); and Department of Psychiatry, University of New Mexico School of Medicine (JF)

Abstract

Context—Antidepressant medications represent the best established treatment for Major Depressive Disorder (MDD), 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 FDA approved antidepressants 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, included a medication vs placebo comparison for at least 6 weeks, did not exclude patients on the basis of a placebo washout period, and utilized the Hamilton Rating Scale for Depression. Data from six 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 Hamilton scores below 23, Cohen's d-type effect sizes for the difference between medication and placebo were estimated to be $< .20$ (a standard definition of a small effect). Estimates of the magnitude of the superiority of medication over placebo increased with increases in baseline Hamilton severity and crossed the NICE threshold for a clinically significant difference at a baseline score of 25.

Corresponding Author: Jay C. Fournier, Department of Psychology, University of Pennsylvania, Philadelphia, PA, 19104-6196, jcf@sas.upenn.edu.

JCF played a central role in the planning of the analyses, performed the analyses, composed (along with SDH and RJD) the first draft of the manuscript, and managed revisions of the manuscript. RJD jointly conceived the idea for this manuscript along with SDH, advised on all aspects of the analysis and interpretation, and offered substantial suggestions for revisions of the manuscript. SDH was involved with the development of the concept for the manuscript, provided input at several stages in the development of the analyses, and helped to compose the first draft of the manuscript. SD, JDA, RCS, and JF were instrumental in the conduct of the studies from which the data were obtained, and they provided consultation, substantive input as to the direction of the manuscript, and editorial suggestions at several stages during the development of the manuscript. All authors have seen and approved the final version of this manuscript. JCF and RJD had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

No other authors have any relevant conflicts of interest to declare.

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.

Introduction

Antidepressant medication (ADM) represents the current standard of treatment for Major Depressive Disorder (MDD).¹ ADM has been shown to be superior to placebo in thousands of controlled clinical trials over the past five decades.^{2, 3} The extent to which ADM outperforms placebo (which controls for non-pharmacological 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 upon data from the FDA clinical trial database. Using means and standard deviations on the Hamilton Rating Scale for Depression (HRSD)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 HRSD score increased, the magnitude of HRSD 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 placebo condition, but found greater symptom change in ADM as baseline HRSD 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 HRSD score needed to achieve a clinically meaningful ADM/placebo difference is approximately 28 and that differences are negligible for lower baseline HRSD scores.

One limitation to these meta-analyses is the range of baseline severity scores included in their constituent studies. In the Kirsch et al.⁴ analysis, only 1 of 35 studies comprised samples with baseline HRSD means lower than 23. As the authors noted, a score of 23 is characteristic of “very severe depression” according to the American Psychiatric Association Taskforce for the Handbook of Psychiatric Measures (who define mild depression as HRSD 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 Kahn et al.⁵ required a minimum entry score of 20 on the HRSD, 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 evidence 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 HRSD 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 Kahn et al. meta-analyses is that each included studies that utilized a placebo washout period. Typically, placebo washouts last from several days to two 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 HRSD) 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 utilize this feature.

In the present study we combined data from six 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 HRSD scores that ranged from the low-mid teens to the upper-30's.^{11–16} Unlike the data analyzed by Kirsch et al. and Kahn 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 six studies included in the present investigation provide data at the level of the individual patient. This allowed us to conduct a mega-analysis of drug-placebo differences as a function of baseline severity. A mega-analytic 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 Kahn et al., we hypothesized that ADM/placebo differences would become larger as baseline severity increased.

Methods

Search Strategy

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 RCTs.

Inclusion Criteria

The criteria for inclusion required studies: 1) to be randomized placebo-controlled trials of an FDA approved antidepressant in the treatment of Major or Minor Depressive Disorder; 2) to include the full range of patients diagnosed with Major Depressive Disorder or Minor Depressive Disorder (i.e., studies that exclusively examined special populations or subtypes were excluded); 3) not to exclude patients on the basis of a placebo washout period; 4) to consist of adult outpatient samples; 5) to include an ADM/placebo comparison of at least 6 weeks duration; 6) to include the HRSD at intake and at the end of treatment; and, 7) to make available to us individual subject level data.

Article Selection

The initial screening of the search results was supervised by SD and reviewed by JCF to ensure accuracy. All articles selected were read by two authors (JCF and either SD or SDH) to determine whether they met inclusion criteria (with an average Kappa of .82). Discrepancies were resolved by consensus.

Data Acquisition

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 subject level data was available. Authors were initially asked to respond within three 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 five MDD trials: Elkin et al.¹⁴, DeRubeis et al.¹², Dimidjian et al.¹³, Philipp et al.¹⁵, Wichers et al.¹⁶, and one 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 utilized the tricyclic antidepressant (TCA) imipramine^{14–16} and three utilized the selective serotonin reuptake inhibitor (SSRI) paroxetine.^{11–13} Table 1 lists characteristics that differ among the six 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 HRSD depression severity levels ranged from 10 to 39. In comparison to the 20 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 (see Supplemental Table).

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 employed a modified intent-to-treat approach whereby we used the most inclusive sample analyzed in the original publication of each of the six studies (see Table 1). To investigate the association between initial severity and symptom change scores in ADM vs placebo, we conducted Analyses of Covariance (ANCOVA) 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 (LOCF) to calculate the change score. Continuous variables were centered at their grand means, and non-significant higher-order interaction terms were removed from the models.

Results

Study Characteristics

Mean baseline depression severity scores and attrition rates for the six studies are displayed in Table 2. A 2 (treatment) \times 6 (study) Analysis of Variance (ANOVA) was conducted to examine differences in levels of intake depression severity. The study-by-treatment interaction was not significant and was removed from the model. Mean intake severity did not differ as a function of treatment condition, $F(1,711)=0.05$, $p=0.82$, but the six studies did evidence different mean intake severity levels, reflecting differences in inclusion criteria, $F(5,711)=79.56$, $p<0.001$. Attrition rates were compared in a logistic regression model examining the effects of study, treatment, and the study-by-treatment interaction. The study-by-treatment interaction term was not significant and was removed from the model. Attrition rates did not differ significantly as a function of treatment condition, $\chi^2(1, N=718)=0.47$, $p=.49$, but differences did emerge in the rates of attrition among the six studies, $\chi^2(5, N=718)=30.34$, $p<.001$ (see Table 2 for specific contrasts).

Baseline Severity and Symptom Change in ADM and Placebo

Pooling the data across the six studies, the severity X 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, $F(1,709)=9.31$, $p=0.002$ ¹ (the main effects of baseline severity, $F(1,709)=59.54$, $p<0.001$, and treatment, $F(1,709)=12.51$, $p<0.001$, were also significant). As displayed in Figure 2, the regression coefficient (i.e., the slope representing the relation between initial severity and change in symptoms) was positive for both ADM ($b=.70$, $t(709)=8.49$, $p<0.001$) and placebo ($b=.36$, $t(709)=3.87$, $p<.001$). The difference in the slopes of the two regression lines, $b=.34$, represents the interaction effect described above. The two regression lines converged near the lower end of the range of baseline severity scores and the magnitude of the difference between the treatments increased with increasing baseline depression severity. To illustrate the magnitude of the difference between the two treatments as a function of initial depression severity, we divided the sample into three groups, based on the characterizations of the HRSD scores offered by the APA task force, mild-to-moderate, HRSD 18 (N=180), severe, HRSD 19–22 (N=255), and very-severe, HRSD 23 (N=283).⁷ For patients in the mild-to-moderate range, the Cohen's d-type effect size was $d=.11$ (95% CI: $-.04$ to $.26$) and for patients in the severe range, $d=.17$ (95% CI: $.04$ to $.30$). Both values fall below the standard description of a small effect ($d=.20$).¹⁸ For patients in the very-severe group, $d=.47$ (95% CI: $.34$ – $.59$). This value falls between the small ($d=.20$) and medium ($d=.50$) effect size ranges. We also converted these d-type effect sizes into estimates of the number of patients needed to treat (NNT) to increase by one the number of patients in the treatment group who would have a better outcome than a randomly selected patient from the control group.²⁰ NNT values are 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 Great Britain has defined a threshold for clinical significance as an effect size of 0.50 or drug-placebo difference of 3 points on the HRSD.¹⁹ In the present data, this threshold was met for intake HRSD scores ≥ 25 , using the more liberal of the two criteria, ≥ 3 HRSD point difference. To examine the more conservative threshold defined by $d=.50$, we estimated Cohen's d-type effect sizes using least-squares means from the primary model described above. Drug-placebo differences were estimated to cross this threshold at an initial HRSD value of 27 (NNT=4). When we divide the sample into subgroups using these two thresholds, the superiority of medications over placebo is associated with a medium sized effect for patients with HRSD ≥ 25 ($d=.53$, 95% CI: $.36$ – $.70$) and a large effect for patients with HRSD ≥ 27 ($d=.81$, 95% CI: $.55$ – 1.07).

Baseline Severity and Symptom Change for Patients with MDD

In order to determine whether the pattern of results reported above was evident in patients diagnosed with MDD, data from the Barrett et al. study of Minor Depressive Disorder were removed and the models were re-run. The severity X 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 HRSD of 25.

¹Three of the studies included patients with lower levels of baseline severity; the other three imposed minimum baseline severity cut-offs of either 18 or 20 on the HRSD. In separate analyses with each of the two sets of studies, using the same models as for the main analysis, the same pattern as reported above was observed. In the lower cutoff group, the difference between treatments in the slopes of the regression lines was $.34$ in favor of ADM over placebo, the same as it was when all patients were included in the main analysis; that value was $.40$ for the three studies with higher cut-offs.

Baseline Severity and Symptom Change for Completers

In order to assess whether attrition might have biased the results, the primary analyses were repeated in a completers-only sample. Again the severity X treatment interaction was significant, $F(1,597)=5.62, p=0.02$. Among completers, the difference between ADM and placebo was estimated to cross the NICE threshold at an initial HRSD score of 24 (one point lower than that observed for the entire sample). We also repeated the primary analysis using data only from the three studies with the lowest dropout rates.^{12,13,15} Again, the interaction of interest was significant, $F(1, 452)=6.98, p < .01$.

Drug Class

Three of the studies utilized the SSRI paroxetine as the active ADM, whereas the other three studies utilized the TCA imipramine. In order 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 X drug class interaction was significant, $F(2,707)=4.41, p=0.01$. Specific contrasts revealed that the regression coefficient (i.e., the slope representing the relationship between initial severity and change in symptoms) was more positive for each medication class relative to placebo: imipramine, $F(1,707)=5.60, p=0.02$; and paroxetine, $F(1,707)=5.91, p=0.02$.

Discussion

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 severe baseline symptoms, whereas they were large for patients with very-severe symptoms. For baseline severity scores on the HRSD less than 25, estimates of the magnitudes of drug-placebo differences did not meet either of the two thresholds for clinical significance proposed by NICE.¹⁹ Conversely, for patients with the highest levels of baseline depression severity, ADM was markedly superior to placebo.

As documented in Zimmerman et al.’s analysis⁸ of published efficacy trials, as well as by Kirsch et al.’s⁴ and Kahn et al.’s⁵ analyses of studies submitted to the FDA, evidence concerning the effects of antidepressant medications 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 seen 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 cut-off scores of 18–20 or more typically have been imposed. As noted by Zimmerman et al., employing 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 HRSD 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 cut-off.²¹ We have no evidence that this occurred in the current datasets, but if it did, it should have worked against the hypothesis that severity moderates outcome. Furthermore, the inclusion of studies with different minimum severity levels should have mitigated any bias that such rater inflation might have caused. Third, the HRSD was used as the primary outcome measure for all analyses. The HRSD 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 utilize 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-by-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.³⁰ Finally, the results reported herein apply to acute treatment only and not to continuation or maintenance.

Despite differences in methods, our findings are consistent with those of both Kirsch et al.⁴ and Kahn et al.⁵ with regard to the finding that ADM/placebo differences increase as initial severity increases. We used individual patient data and included patients with less severe depressions, whereas both Kirsch et al. and Kahn et al. analyzed group means that largely excluded patients with HRSD scores below 20. Moreover, both Kirsch et al. and Kahn 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 three reviews is striking. However, there also were subtle differences in the pattern of findings across the three 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 Kahn 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.^{24, 25} 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.^{26, 27} 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 1950's researchers investigating a wide variety of medical and psychiatric conditions described a phenomenon whereby patients with higher baseline severity scores received more benefit from active treatments than from control conditions.^{28, 29} What makes our findings both surprising and compelling 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 antidepressant medication in clinical practice present with scores below these levels.

Prescribers, policymakers, 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 practitioners and the public. There is little mention of the fact that efficacy data 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 or those

obtained by Kirsch et al. and Khan et al., efforts should be made to clarify to providers and prospective patients that whereas ADMs can be quite potent with more severe depressions, there is little evidence to suggest that they produce specific pharmacological benefit for the majority of patients with less severe depressions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like thank the authors who shared their data with us for this project, with special thanks to Drs. Marieke Wichers (Maastricht University, School for Mental Health and Neuroscience), John Cornell, Ph.D. (University of Texas Health Science Center), and Karl-O. Hiller (Steiner Arzneimittel) for providing us with the raw data from their respective studies. Finally, we would like to thank all those who responded to our inquiries, even if data from their studies could not be made available.

RCS has served as a consultant to AstraZeneca, Eli Lilly & Co, Evotec AG, Forest Pharmaceuticals, Inc., Gideon Richter PLC, Janssen Pharmaceuticals, Merck & Co., Novartis Pharmaceuticals, Ostuka Pharmaceuticals, Pamlab, Inc, Pfizer, Inc. Repligen, Inc. Sierra Neuropharmaceuticals, and Wyeth, Inc. He has received speaking honoraria from AstraZeneca, Eli Lilly & Co., Forest Pharmaceuticals, Inc., GlaxoSmithKline PLC, Pamlab, Inc., Pfizer, Inc., and Wyeth, Inc. Finally, he received research and/or grant support from Bristol Myers Squibb, Eli Lilly & Co., Evotec AG, Forest Pharmaceuticals, Inc., GlaxoSmithKline PLC, Janssen Pharmaceuticals, Novartis Pharmaceuticals, Ostuka Pharmaceuticals, Pamlab, Inc. Pfizer, Inc. Repligen, Inc. and Wyeth, Inc.

JDA has served on the Speaker's Bureau of Wyeth Pharmaceuticals and Bristol Myers Squibb. He has received research support for Novartis, Lilly, Sanofi, Cephalon, and Forest Laboratories. He has served as a consultant for Bristol Myers Squibb.

JF has served as a consultant to Abbott Laboratories, Merck & Co., and Slack, Inc. He has received speaking honoraria from Eli Lilly & Co. He has served as a board member for the Berman Center and on the Scientific Advisory Boards for the non-profit advocacy organizations NARSAD and the Depression and Bipolar Support Alliance. He also has provided expert testimony on cases involving pharmaceutical companies including Banner Health and Alphapharm. He currently chairs the Mood Disorders Work Group for the revision of DSM-V.

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, USA. No funding source had direct oversight of the design or conduct of the study; collection, management, analysis, or interpretation of the data; or preparation, review, or approval of the manuscript.

References

- Hollon SD, Thase ME, Markowitz JC. Treatment and prevention of depression. *Psychological Science in the Public Interest*. 2002; 3(2):39–77.
- Mulrow, CD.; Williams, JW., Jr; Trivedi, M., et al. AHCPR Publication No 99-E014. Rockville, MD: U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research; 1999. Treatment of depression: Newer pharmacotherapies.
- Depression Guideline Panel. Clinical Practice Guideline No. 5. AHCPR Publication No. 93-0551. Rockville, MD: U.S. Department of Health and Human Services, Public Health Service, Agency for Health Care Policy and Research; 1993. Depression in primary care: Vol. 2. Treatment of major depression.
- 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. [PubMed: 18303940]
- 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. [PubMed: 11799341]
- Hamilton MA. A rating scale for depression. *J Neurol Neurosurg Psychiatry*. 1960; 23:56–62. [PubMed: 14399272]

7. American Psychiatric Association Task Force for the Handbook of Psychiatric Measures. Handbook of psychiatric measures. Washington, D.C: 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. [PubMed: 12454561]
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. [PubMed: 14978780]
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. [PubMed: 7945742]
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. [PubMed: 11350703]
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. [PubMed: 15809408]
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. [PubMed: 16881773]
14. Elkin I, Shea MT, Watkins JT, Imber SD, 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. [PubMed: 2684085]
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. *Br Med J*. 1999; 319(7224):1534–1538. [PubMed: 10591711]
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. [PubMed: 18496519]
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; 145:917–925. [PubMed: 9149663]
18. Cohen, J. *Statistical power analysis for the behavioral sciences*. 2. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
19. National Institute for Clinical Excellence. *Depression: Management of depression in primary and secondary care*. London: National Institute for Clinical Excellence; 2004.
20. Kraemer HC, Kupfer DJ. Size of treatment effects and their importance to clinical research and practice. *Biol Psychiatry*. 2006; 59(11):990–996. [PubMed: 16368078]
21. Landin R, DeBrotta DJ, DeVries TA, Potter WZ, Demitrack MA. The impact of restrictive entry criterion during the placebo lead-in period. *Biometrics*. 2000; 56:271–278. [PubMed: 10783806]
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. [PubMed: 15738740]
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. [PubMed: 15569884]
24. Baldwin D, Rudge S, Thomas S. Dysthymia: Options in pharmacotherapy. *CNS Drugs*. 1995; 4:422–431.
25. 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. [PubMed: 10616934]
26. Dunner DL. Acute and maintenance treatment of chronic depression. *J Clin Psychiatry*. Special Issue: Chronic major depression: A review and update. 2001; 62(Suppl6):10–16.
27. Khan A, Dager SR, Cohen S, Avery DH, et al. Chronicity of depressive episode in relation to antidepressant-placebo response. *Neuropsychopharmacology*. 1991; 4(2):125–130. [PubMed: 2025378]

28. 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. [PubMed: 14080094]
29. Fisher S, Lipman RS, Uhlenhuth EH, Rickels K, Park LC. Drug effects and initial severity of symptomatology. *Psychopharmacologia*. 1965; 14:57–60. [PubMed: 5318924]
30. 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. [PubMed: 18199864]
31. Boyer P, Montgomery S, Lepola U, et al. Efficacy, safety, and tolerability of fixed-dose desvenlafaxine 50 and 100 mg/day for major depressive disorder in a placebo-controlled trial. *Int Clin Psychopharmacol*. 2008; 23(5):243–253. [PubMed: 18703933]
32. Cohn CK, Robinson DS, Roberts DL, Schwiderski UE, O'Brien K, Ieni JR. Responders to antidepressant drug treatment: a study comparing nefazodone, imipramine, and placebo in patients with major depression. *J Clin Psychiatry*. 1996; 57 (Suppl 2):15–18. [PubMed: 8626358]
33. D'Amico MF, Roberts DL, Robinson DS, Schwiderski UE, Copp J. Placebo-controlled dose-ranging trial designs in phase II development of nefazodone. *Psychopharmacol Bull*. 1990; 26(1): 147–150. [PubMed: 2196622]
34. DeMartinis NA, Yeung PP, Entsuah R, Manley AL. A double-blind, placebo-controlled study of the efficacy and safety of desvenlafaxine succinate in the treatment of major depressive disorder. *J Clin Psychiatry*. 2007; 68(5):677–688. [PubMed: 17503976]
35. Feiger AD. A double-blind comparison of gepirone extended release, imipramine, and placebo in the treatment of outpatient major depression. *Psychopharmacol Bull*. 1996; 32(4):659–665. [PubMed: 8993088]
36. Feiger AD, Rickels K, Rynn MA, Zimbroff DL, Robinson DS. Selegiline transdermal system for the treatment of major depressive disorder: an 8-week, double-blind, placebo-controlled, flexible-dose titration trial. *J Clin Psychiatry*. 2006; 67(9):1354–1361. [PubMed: 17017821]
37. Gastpar M, Singer A, Zeller K. Comparative efficacy and safety of a once-daily dosage of hypericum extract STW3-VI and citalopram in patients with moderate depression: a double-blind, randomised, multicentre, placebo-controlled study. *Pharmacopsychiatry*. 2006; 39(2):66–75. [PubMed: 16555167]
38. Giller EJ, Bialos D, Harkness L, Riddle M. Assessing treatment response to the monoamine oxidase inhibitor isocarboxazid. *J Clin Psychiatry*. 1984; 45:44–48. [PubMed: 6376484]
39. Hollyman J, Freeling P, Paykel E, Bhat A, Sedgwick P. Double-blind placebo-controlled trial of amitriptyline among depressed patients in general practice. *J R Coll Gen Pract*. 1988; 38:393–397. [PubMed: 3076905]
40. Lecrubier Y, Guelfi JD. Efficacy of reversible inhibitors of monoamine oxidase-A in various forms of depression. *Acta Psychiatr Scand Suppl*. 1990; 360:18–23. [PubMed: 2248063]
41. Lieberman DZ, Montgomery SA, Tourian KA, et al. A pooled analysis of two placebo-controlled trials of desvenlafaxine in major depressive disorder. *Int Clin Psychopharmacol*. 2008; 23(4):188–197. [PubMed: 18545056]
42. Liebowitz MR, Manley AL, Padmanabhan SK, Ganguly R, Tummala R, Tourian KA. Efficacy, safety, and tolerability of desvenlafaxine 50 mg/day and 100 mg/day in outpatients with major depressive disorder. *Curr Med Res Opin*. 2008; 24(7):1877–1890. [PubMed: 18507895]
43. Liebowitz MR, Yeung PP, Entsuah R. A randomized, double-blind, placebo-controlled trial of desvenlafaxine succinate in adult outpatients with major depressive disorder. *J Clin Psychiatry*. 2007; 68(11):1663–1672. [PubMed: 18052559]
44. Mynors-Wallis LM, Gath DH, Lloyd-Thomas AR, Tomlinson D. Randomised controlled trial comparing problem solving treatment with amitriptyline and placebo for major depression in primary care. *BMJ*. 1995; 310(6977):441–445. [PubMed: 7873952]
45. Rickels K, Case WG. Trazodone in depressed outpatients. *Am J Psychiatry*. 1982; 139(6):803–806. [PubMed: 7044154]
46. Rickels K, Weise CC, Zal HM, Csanalosi I, Werblowsky J. Lofepamine and imipramine in unipolar depressed outpatients. A placebo controlled study. *Acta Psychiatr Scand*. 1982; 66(2): 109–120. [PubMed: 6753501]

47. Septien-Velez L, Pitrosky B, Padmanabhan SK, Germain JM, Tourian KA. A randomized, double-blind, placebo-controlled trial of desvenlafaxine succinate in the treatment of major depressive disorder. *Int Clin Psychopharmacol.* 2007; 22(6):338–347. [PubMed: 17917552]
48. UK Moclobemide Study Group. A multicentre comparative trial of moclobemide, imipramine and placebo in major depressive disorder. *Int Clin Psychopharmacol.* 1994; 9:109–113. [PubMed: 8056992]
49. Versiani M, Nardi AE, Mundim FD, Alves A, Schmid-Burgk W. Moclobemide, imipramine and placebo in the treatment of major depression. *Acta Psychiatr Scand Suppl.* 1990; 360:57–58. [PubMed: 2248073]
50. Versiani M, Oggero U, Alterwain P, Capponi R, et al. A double-blind comparative trial of moclobemide v. imipramine and placebo in major depressive episodes. *Br J Psychiatry.* 1989; 155(Suppl 6):72–77.

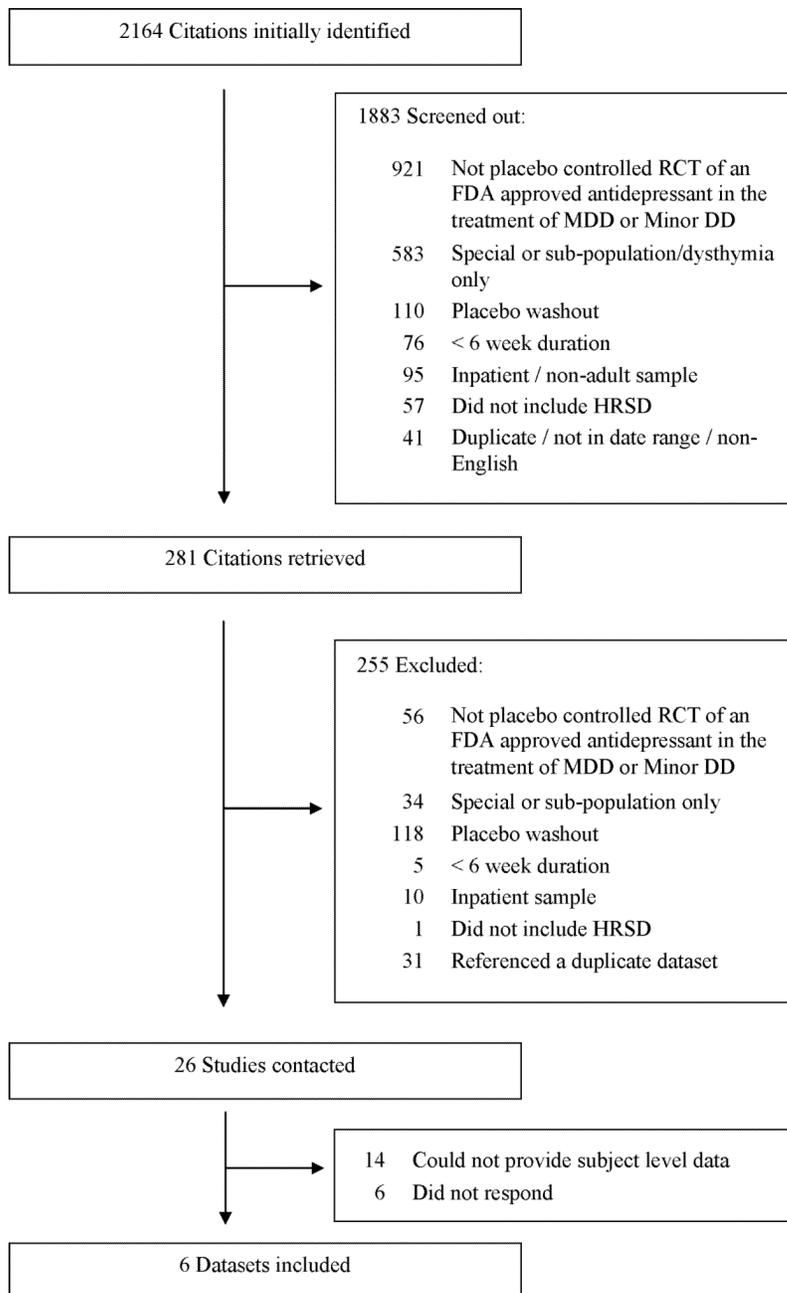


Figure 1. Study Selection and Data Acquisition. The reasons for exclusion describe the first reason for exclusion that was encountered during the review process. Several articles had multiple reasons for exclusion.

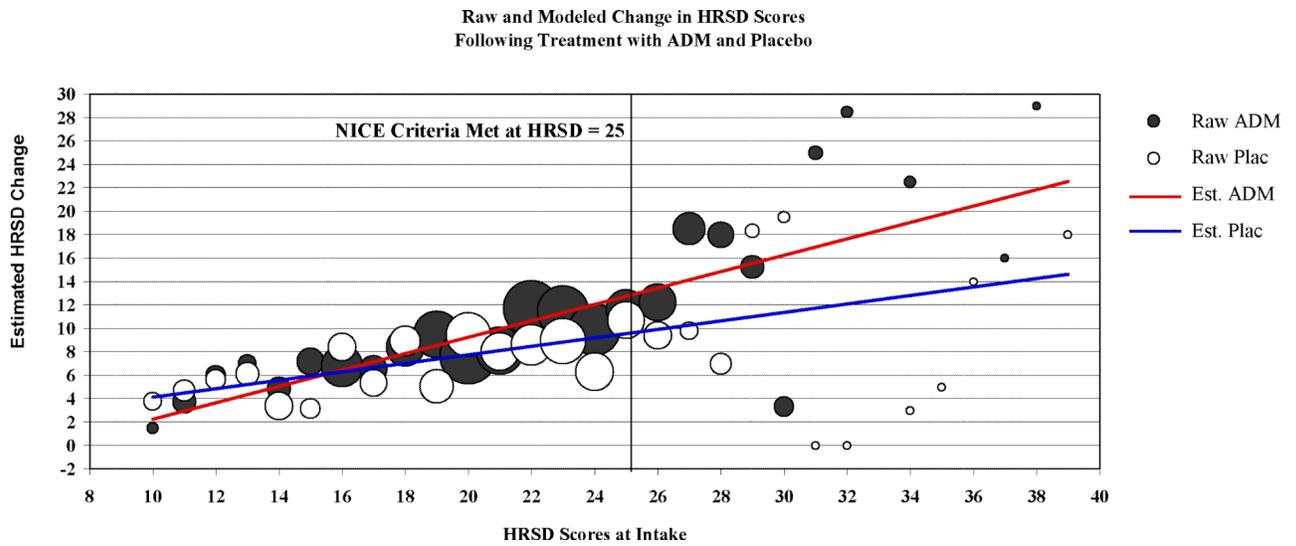


Figure 2. Circles represent raw mean change in depressive symptoms from intake to the end of treatment at each initial HRSD score for both the antidepressant medication (ADM) and placebo (Plac) conditions. The size (area) of the circle represents the number of data points that contributed to each mean. The regression lines represent the estimates of change in depression symptoms from intake to the end of treatment for ADM and Plac conditions as a function of baseline symptom severity. These regression lines were estimated from a model of the baseline severity-by-treatment interaction, controlling for the effects of the study from which the data originated.

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

Characteristic	Barrett et al.	Elkin et al.	DeRubeis et al.	Dimidjian et al.	Philipp et al.	Wichers et al.
Disorder examined	Minor ^f	MDD	MDD	MDD	MDD	MDD
# of intake evaluations ^a	1	2	2	1	1	2
# of treatment sites	2	3	2	1	18	8
Medication used	Paroxetine	Imipramine	Paroxetine	Paroxetine	Imipramine ^g	Imipramine
Target dose (mg/day)	20–40	150–250	50	50	100	100–200
Blinded evaluations ^b	Yes	Yes	Yes	Partial	Partial	Partial
HRSD Version	17-item	Modified 17-item	Modified 17-item	Modified 17-item	17-item	17-item
Minimum intake severity ^c	10	14	20	14	18	18
Sample analyzed ^d	Full-ITT	ITT	Full-ITT	ITT	Full-ITT	Complied with Protocol
Treatment duration (weeks) ^e	11	8	8	8	8	6

Note. Minor = Minor Depressive Disorder; MDD = Major Depressive Disorder; Entered = Entered Treatment; ITT = Intent to Treat

^aThe Elkin et al., DeRubeis et al., and Wichers et al. trials required participants to meet inclusion criteria in each of two consecutive evaluations that were held at least one week apart.

^bYes = Independent blind evaluators conducted evaluations of symptom severity at every assessment. Partial = Evaluations were conducted at each session by the treating pharmacotherapists. Treating pharmacotherapists were blind to treatment condition.

^cSix patients from the Elkin et al. sample registered scores less than 14 on the HRSD at intake (two from the imipramine and four from the placebo conditions) and one 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.

^dThe Barrett et al., DeRubeis et al., and Philipp et al. studies employed a full intent-to-treat design whereby all patients randomized to treatment were included in the analysis. Both the Elkin et al. and the Dimidjian et al. studies used a common intent-to-treat approach whereby data from only those patients who attended at least one treatment session 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).

^eTreatment 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 in order to improve comparability between the studies.

^fThe Barrett et al. study also included patients diagnosed with Dysthymia. These patients were not included in the current analyses.

^gThe Philipp et al. trial also included a hypericum extract condition. Data from this condition were not included in the current analyses.

Table 2

Sample Size, Dropout Rates, and Baseline Depression Severity in Six Studies that Compared Active Antidepressant Medications to Pill-Placebo

Study	Dropouts			Baseline HRSD Severity		
	N	N	%	Mean	SD	SD
Barrett et al. (2001) ^c						
Placebo	39	8	21	14.38	3.98	
Medication	38	13	34	14.05	2.86	
Elkin et al. (1989) ^c						
Placebo	62	17	27	19.47	4.60	
Medication	57	13	23	19.51	4.62	
DeRubeis et al. (2005) ^b						
Placebo	60	9	15	23.47	2.73	
Medication	120	13	11	23.22	2.70	
Dimidjian et al. (2006) ^b						
Placebo	45	4	9	20.76	4.59	
Medication	82	15	18	20.73	3.94	
Philipp et al. (1999) ^a						
Placebo	46	3	7	22.20	4.15	
Medication	105	3	3	22.70	4.03	
Wichers et al. (2009) ^{b,c}						
Placebo	32	5	16	23.72	2.81	
Medication	32	9	28	24.38	3.84	

Note. Studies with the same superscript do not differ with respect to mean rates of attrition at $p < .05$.