Placebo Response in Studies of Major Depression: Variable, Substantial, and Growing

B. Timothy Walsh, MD
Stuart N. Seidman, MD
Robyn Sysko, BA
Madelyn Gould, PhD

Context Intense debate persists about the need for placebo-controlled groups in clinical trials of medications for major depressive disorder (MDD). There is continuing interest in the development of new medications, but because effective antidepressants are already available, ethical concerns have been raised about the need for placebo groups in new trials.

Objective To determine whether the characteristics of placebo control groups in antidepressant trials have changed over time.

Data Sources and Study Selection We searched MEDLINE and PsychLit for all controlled trials published in English between January 1981 and December 2000 in which adult outpatients with MDD were randomly assigned to receive medication or placebo. Seventy-five trials met our criteria for inclusion.

Data Extraction Data were extracted from the articles by 2 of the authors and discrepancies were resolved via discussion and additional review by a third author.

Data Synthesis The mean (SD) proportion of patients in the placebo group who responded was 29.7% (8.3%) (range, 12.5%-51.8%). Most studies examined more than a single active medication, and, in the active medication group with the greatest response, the mean (SD) proportion of patients responding was 50.1% (9.0%) (range, 31.6%-70.4%). Both the proportion of patients responding to placebo and the proportion responding to medication were significantly positively correlated with the year of publication (for placebo: n=75; r=0.45; 95% confidence interval [CI], 0.25-0.61; P<.001; for medication: n=75; r=0.26; 95% CI, 0.03-0.46; P=.02). The association between year of publication and response rate was more statistically robust for placebo than medication.

Conclusions The response to placebo in published trials of antidepressant medication for MDD is highly variable and often substantial and has increased significantly in recent years, as has the response to medication. These observations support the view that the inclusion of a placebo group has major scientific importance in trials of new antidepressant medications and indicate that efforts should continue to minimize the risks of such studies so that they may be conducted in an ethically acceptable manner.

JAMA. 2002;287:1840-1847

See also pp 1807 and 1853.
tion of efficacy compared with placebo eliminates the risk to public health of approving ineffective medications and because there is no evidence that treatment delay or assignment to placebo results in permanent harm. Along with these ethical arguments, investigators have suggested that, in recent years, it has become more difficult to demonstrate specific antidepressant effects compared with placebo, suggesting that the response to placebo may have increased. In this context of persistent ethical and scientific concerns about the use of placebo in studies of MDD, we examined data on the response to placebo and medication in controlled trials of adult outpatients with MDD published between 1981 and 2000 to determine whether the characteristics of such studies and the response to placebo have changed.

METHODS
Sources of Data and Criteria for Review
We collected peer-reviewed articles describing randomized, placebo-controlled trials of medication for outpatients with unipolar MDD published between 1981 and 2000. Studies published between 1981 and 1990 were taken primarily from an extensive, previously published literature review. To create a list of articles published between 1991 and 2000, we used similar methods. We specified key words, including placebo and generic names of all putative antidepressant medications, and conducted multiple computer searches using MEDLINE and PsycLit.

To be included in this review, articles were required to meet the following criteria: (1) reported in English; (2) published between January 1981 and December 2000; (3) primarily composed of outpatients with MDD according to research diagnostic criteria (RDC), Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Diagnostic and Statistical Manual of Mental Disorders, Revised Third Edition, Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, International Classification of Diseases, Ninth Revision (ICD-9), or International Classification of Diseases, Tenth Revision (ICD-10) criteria for major depression (we excluded trials that focused on bipolar disorder, but did include several early studies that contained a small number of bipolar depressed patients); (4) had at least 20 patients in the placebo group; (5) lasted at least 4 weeks; (6) randomly assigned patients to receive a putative antidepressant drug or drugs and placebo and assessed patients under double-blind conditions; and (7) reported the total number of patients assigned to placebo and medication group(s) and the number who had responded as determined by a reduction of at least 50% in their score on the Hamilton Rating Scale for Depression (HRSD) and/or a Clinical Global Impression (CGI) rating of markedly or moderately improved (CGI score of 1 or 2).

Data were extracted from the articles by 2 of the authors (S.S., R.S.) and discrepancies were resolved via discussion and additional review by a third author (B.T.W.). For each study, we calculated the response rate for the placebo and medication group(s) by dividing the number of patients in each group who had responded according to these criteria by the number of patients assigned to the group. In a few studies, only the number of patients randomized who returned for at least 1 visit was given, and this number was used as the denominator. If response data were provided only on study completers, we assumed that non-completers had not responded.

Statistical Methods
To examine the strength of the linear relationship between continuous variables, we used Pearson correlation coefficients (2-tailed) and linear regression. To examine differences between groups in continuous variables, we used the t statistic. To assess whether the proportion of studies that used certain design features (eg, placebo lead-in before randomization) had changed linearly with time, we determined the number of studies that used the feature during each of the four 5-year periods between 1981 and 2000 and calculated a χ² statistic for linearity. Tests of statistical significance regarding fractions of patients responding were also performed using an arcsin transformation of the fraction, specifically, 2 × arcsin. The results of the analyses using the untransformed and transformed data were virtually identical, and only results of analyses of untransformed response rates are reported herein. Effect size was calculated as the difference between the arcsin-transformed response rate on active medication and the arcsin-transformed response rate on placebo. Statistical calculations were performed using SPSS for Windows, version 10 (SPSS Inc, Chicago, Ill.).

RESULTS
Study Characteristics
Seventy-five trials of medication for adult outpatients with MDD meeting our criteria for review were identified. Data are presented as mean (SD). The trial length averaged 45.4 (17.9) days (range, 28-168 days) and increased significantly with year of publication (n=74; r=0.45; 95% confidence interval [CI], 0.25-0.61; P<.001). The average number of patients per group was 83.8 (59.7) (range, 20-336) for placebo and 85.0 (60.0) (range, 20-335) for medication; neither changed significantly with year of publication (placebo: n=75; r=0.12; 95% CI, −0.11 to 0.34; P=.30; medication: n=75; r=0.10; 95% CI, −0.13 to 0.32; P=.41). The diagnostic systems used have changed over time. For example, the proportion of studies using RDC declined from 28.6% between 1981 and 1985 to 0% between 1996 and 2000 (χ²=8.5, P=.004). The diagnostic systems generally require a minimum duration of 2 weeks of symptoms. Seventeen trials (22.7%) required a longer duration of illness, usually 4 weeks, to merit a diagnosis of MDD. Most trials (74.7%) used a 1- to 2-week initial placebo lead-in period; patients who responded during this time were not randomized. Neither the number of studies that required more than 2 weeks of symptoms nor the number ...
using placebo lead in changed significantly with time ($\chi^2_1 = 1.8$ and 1.7, and $P = .18$ and .19, respectively).

Seventy-four studies used the HRSD to measure severity of depression, and 67 reported which version was used to assess severity of depression at study entry: the 17-item version was used in 33 studies (49%), the 24-item version in 2 (2.9%), and 17-item version in 2 (2.9%). The proportion of studies using the 17-item version did not change significantly with time ($\chi^2_1 = 0.31$, $P = .58$). Sixty-eight studies required a minimum score on the HRSD for study entry. To compare the HRSD scores across studies using different versions of the HRSD, we prorated the HRSD score by dividing the score by the number of items in the HRSD version and multiplied by 17. Among the 63 studies that provided sufficient information to perform this calculation, the prorated minimum entry score averaged 16.7 (2.1) (range, 12.9-22.0) and increased with year of publication ($n=63; r=0.37; 95\% CI, 0.13-0.57; P = .003$). Forty-one of the studies provided explicit information on whether the use of concomitant medications (usually chloral hydrate or a benzodiazepine) was permitted during the trial for treatment of anxiety or insomnia. In 38 (84%) of these 45 studies, concomitant medications were permitted. The proportion of trials permitting concomitant medication did not change significantly with time ($\chi^2_1 = 0.72$, $P = .39$).

**Placebo Response**

Fifty-three studies provided information on the number of patients responding to placebo as indicated by a 50% or greater reduction in HRSD score. By this measure, the average proportion of patients per study who responded while receiving placebo was 30.0% (8.0%), and this proportion increased significantly with year of publication ($n=53; r=0.43; 95\% CI, 0.18-0.63; P = .001$). Forty-seven studies provided information on the number of patients responding to placebo as indicated by a CGI rating of 1 or 2. By this measure, the average proportion of patients per study who responded while receiving placebo was 32.8% (11.1%), and this proportion increased significantly with year of publication ($n=47; r=0.53; 95\% CI, 0.29-0.71; P < .001$).

We were interested in combining data from all studies to examine the relationship between study characteristics and placebo response rate. Twenty-two studies presented placebo response information only on the basis of CGI rating. To combine information from these studies with data from the other 53 studies that provided placebo response information on the basis of a 50% or greater reduction in HRSD score, we estimated the proportion responding according to HRSD as follows. Twenty-five studies reported both the number of patients receiving placebo who responded according to HRSD criteria and the number who responded according to CGI criteria. These proportions were highly correlated ($n=25; r=0.79; 95\% CI, 0.57-0.90; P < .001$). We estimated the proportion responding according to HRSD criteria using the equation for the best-fit straight line obtained from the 25 studies that provided data on response according to both measures (HRSD response rate = 0.658 × CGI response rate + 0.0832). According to this combined HRSD measure, 29.7% (8.3%) (range, 12.5%-51.8%) of patients assigned to placebo responded. The proportion of patients taking placebo responding by this measure was significantly correlated with year of publication ($n=75; r=0.45; 95\% CI, 0.25-0.61; P < .001$) (Figure), length of the trial ($n=74; r=0.27; 95\% CI, 0.04-0.47; P = .02$), and prorated minimum HRSD score required ($n=63; r=0.27; 95\% CI, 0.02-0.49; P = .03$). In a multiple linear regression model, with proportion of patients taking placebo responding according to the combined HRSD measure as the dependent variable and year of publication, length of trial, and prorated minimum HRSD score as independent variables, only year of publication remained significant ($\beta=0.366, t_{60}=2.7, P = .009$). There were no statistically significant associations between the proportion of patients taking placebo who responded and prorated average initial HRSD score ($n=61; r=0.10; 95\% CI, -0.16 to 0.34; P = .46$), average patient age ($n=61; r=0.11; 95\% CI, -0.15 to 0.35; P = .42$), fraction of patients who were women ($n=63; r=0.09; 95\% CI, -0.16 to 0.33; P = .49$), presence of a placebo lead-in phase ($t_{57}=0.33, P = .73$), minimum duration of symptoms required being greater than 2 weeks ($t_{57}=0.23, P = .82$), or concomitant medication use ($t_{57}=0.86, P = .39$).

**Medication Response**

To put the response to placebo in context, we examined the response to active medication in the same 75 studies. The 25 studies that provided information regarding placebo response on the basis of both a 50% reduction in HRSD score and CGI response contained similar information on medication response. Since most studies examined more than 1 active medication, we determined the best-fit straight line describing the relationship between CGI response and HRSD response to the first medication described in each of these studies: HRSD response rate = 0.744 × CGI response rate + 0.0805. The slope for this equation does not differ significantly from that obtained for placebo response rates ($t_{46}=0.56, P = .72$). For studies that provided response rates only using CGIs, we estimated the proportion responding according to HRSD criteria using this equation.

Sixty studies (80%) assigned more than 1 group of patients to active medi-
cations (average number of active medication groups, 1.97 per study; range, 1-4), and many assigned different groups of patients to receive different doses or preparations of the same medication. For each study with more than 1 active medication group, the greatest actual or estimated response rate using HRSD criteria in one medication group was designated the maximum medication response rate for that study. The 2 classes of medication most frequently studied were the tricyclic antidepressants (TCAs, 43 studies) and the selective serotonin reuptake inhibitors (SSRIs, 33 studies). For each study that examined a TCA and/or an SSRI, we determined the TCA and/or the SSRI response rates; for studies that assigned different groups of patients to different preparations or doses within the same class, we used data from the group with the greatest response rate.

The average maximum proportion of patients responding to medication in a study was 50.1% (9.0%) (range, 31.6%-70.4%). The proportion of patients responding to TCAs was 46.9% (10.6%) (range, 27.5%-65.6%) and to SSRIs was 48.9% (10.3%) (range, 25.0%-70.4%). The proportion of patients responding to medication was correlated with year of publication (for maximum proportion of patients responding to a medication, \( n = 75; r = 0.26; 95\% \text{ CI}, 0.03-0.46; P = 0.02 \); for TCA response: \( n = 43; r = 0.29; 95\% \text{ CI}, -0.01 \text{ to } 0.54; P = 0.06 \); for SSRI response: \( n = 33; r = 0.47; 95\% \text{ CI}, 0.15-0.70; P = 0.006 \) (Figure). However, when significant covariates were entered into the regression models (trial length for maximum proportion responding and SSRI response; prorated initial HRSD score for TCA response), neither the relationship of medication response with year of publication nor with the covariate remained significant.

We calculated effect size, the difference between the response to medication and the response to placebo, using an arcsin transformation. The average effect size across studies for the maximum proportion of patients responding to a medication was 0.43 (0.22) (for TCA response, 0.38 (0.22); for SSRI response, 0.40 (0.24)). There were no statistically significant correlations between any of these effect sizes and year of publication (for maximum proportion of patients responding to a medication, \( n = 75; r = -0.17; 95\% \text{ CI}, -0.38 \text{ to } 0.06; P = 0.14 \); for TCA response: \( n = 43; r = -0.09; 95\% \text{ CI}, -0.38 \text{ to } 0.22; P = 0.56 \); for SSRI response: \( n = 33; r = -0.16; 95\% \text{ CI}, -0.48 \text{ to } 0.19; P = 0.38 \).

**Time Course**

We were interested in determining the week at which a significant difference in the mean HRSD score between medication and placebo groups was first noted. Fifty-three reports found a putative antidepressant to be superior to placebo, and 40 of these provided information, usually a graph, that allowed determination of the first week at which a statistically significant difference (\( P < 0.05 \)) between mean HRSD scores of medication and placebo groups was noted. If data on several medications or medication doses were included in one report, we used the week of first detectable medication vs placebo difference for the medication or dose with the greatest overall response rate. A significant drug effect was first noted, on average, 2.40 (1.48) weeks (range, 1-7 weeks) following randomization, and, in 92% (37 of 40 studies), by 4 weeks after randomization. In 19 studies that provided data on the time course of response to TCAs, a medication-placebo difference was noted in 2.53 (1.54) weeks (range, 1-6 weeks), and, in 19 studies with data on SSRIs, a medication-placebo difference was noted in 2.74 (1.63) weeks (range, 1-7 weeks).

**COMMENT**

This review of 75 placebo-controlled trials of medication for adult outpatients

![Figure](Figure.png)

For studies that did not provide data on response using this criterion, an estimated proportion was calculated based on the proportion of patients rated moderately or markedly improved on the Clinical Global Impression (see "Results" section). The lines are the best-fit straight lines describing the relationship between proportion of patients responding and year of publication for placebo (\( r = 0.45, n = 75, P = 0.001 \)), TCAs (\( r = 0.29, n = 43, P = 0.06 \)), and SSRIs (\( r = 0.47, n = 33, P = 0.006 \)).
PLACEBO RESPONSE IN DEPRESSION STUDIES

with MDD published between 1981 and 2000 leads to several conclusions relevant to considering the inclusion of placebo groups in future studies of antidepressant medication. First, the response to placebo across the trials varied, ranging from approximately 10% to more than 50%, and frequently was substantial; in approximately half of the studies, 30% or more of the patients assigned to placebo exhibited a clinically significant improvement. Second, in the last 2 decades, the proportion of patients responding to placebo has clearly increased, at the rate of approximately 7% per decade, and a similar increase has occurred in the fraction of patients responding to active medication. The effect size in published studies has not changed significantly, emphasizing the importance of examining unpublished data to determine whether the increased rate of placebo response may have contributed to the difficulty of establishing the efficacy of new antidepressants.18,100

The change in placebo response rate does not appear to be directly explained by changes in study characteristics. Although the diagnostic systems used have shifted with time as psychiatry has adopted the American Psychiatric Association’s editions of the Diagnostic and Statistical Manual of Mental Disorders, all the systems are close descendants of the RDC, and the specific criteria are extremely similar in the nature and duration of the symptoms required for diagnosis of MDD. Thus, there is little reason to think that the shift from RDC to Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition during the last 20 years is responsible for changes in response rates. The average age of patients has increased, but the relationship between age and placebo response rate was not significant. The length of randomized controlled trials has increased, and we found, as have others, that the proportion of patients responding to placebo increases with trial length.101-103 Presumably, this association reflects both the cumulative effects of the nonspecific interventions inherent in clinical trials and a longer period during which spontaneous recovery could occur. We also found that, after prorating for the number of items in the HRSDs used in different studies, more recent studies required a higher minimum HRSD score for study entry and that there was a significant positive correlation between minimum entry score and placebo response rate. However, in a multiple regression model that included year of publication, length of trial, and minimum required HRSD score, only year of publication was a significant predictor of placebo response. It is of note that the average placebo response rate of studies that used a placebo lead in did not differ significantly from that of studies that did not use a placebo lead in.104

Data on other potentially relevant characteristics, such as duration of illness or number of prior episodes of depression, were not reported in a sufficiently similar form across studies to permit evaluation of their impact on placebo response rate.

A recently published meta-analysis105 found little evidence that the course of patients receiving a placebo can be distinguished from the course of patients receiving no treatment, suggesting that the effect of placebo is usually no greater than the effect of the passage of time. This conclusion was not as clear-cut for studies examining more subjective outcomes such as pain, and the meta-analysis located only a few studies of the short-term treatment of MDD. Therefore, it is unclear whether placebo, when used in studies of MDD, has a greater effect than no treatment. In either case, the current review indicates that the impact of placebo in trials of MDD, even if it reflects no more than the natural course of untreated MDD, has changed in the last 2 decades. In addition, most studies permit the use of sedatives and anxiolytics, so that what is reported as response to placebo typically encompasses the impact of such agents.

Our review found that the response rates to both placebo and active medication in trials of MDD have increased significantly in the last 20 years and that only year of publication is a significant predictor of placebo response when other potentially relevant factors are taken into account. Some factor or factors associated with the level of placebo response must therefore have changed significantly during this period. Unfortunately, we were not able to determine the identity of these factors. The only characteristic of MDD available across studies was the pro-rated average initial HRSD score, which, despite an increase in the minimum HRSD score required for study entry, did not change significantly in the last 20 years. Because the HRSD was the sole measure of illness severity, was used in 1 of 3 different forms, and can be difficult to compare across studies,106 it is possible that the HRSD was not sufficiently sensitive to capture relevant changes in the characteristics of MDD.

In recent years, as effective treatments for depression have become more widely available and socially acceptable, there has been a marked increase in the proportion of the population receiving treatment for depression.107 The factors responsible for this trend may also have contributed to the participation in clinical trials of individuals with milder, briefer, and more responsive forms of depression. In addition, the methods by which patients are recruited into therapeutic trials have changed. Twenty years ago, most patients were referred by other clinicians or learned of research studies while trying to obtain treatment from established centers. In more recent years, it has become commonplace for investigators to advertise the availability of research trials to the public. It is likely that clinically important characteristics of patients participating in treatment studies have changed as a result of these practices108 and have contributed to changes in placebo and medication response rates.

The variability of the response of patients with MDD to placebo across studies and the fact that the average rate of response has changed over time strongly argue against the use of a hist-
Placebo Response in Depression Studies

Our study has several limitations. We examined only published data. For a number of reasons, including the fact that most placebo-controlled trials of antidepressant medications are supported by the pharmaceutical industry, studies in which a significant drug-placebo difference is detected are presumably much more likely to be published than ones in which the responses to placebo and active medication are statistically indistinguishable. Therefore, it seems probable that the placebo response rates in the published studies we examined are lower, and the medication response rates higher, than those in unpublished studies. Several authors have recently begun to access data submitted to the FDA, much of which is unpublished. Until these and other unpublished studies are examined, uncertainty will persist about changes in the response to placebo and to medication over time. In addition, the variable delays in publication of results could bias the results regarding date of publication. Another limitation is that the results presented herein are based on somewhat arbitrary criteria for clinical response. We used a 50% decrease in HRSD score and/or a CGI improvement score of 1 or 2 as the criteria for response because they are routinely reported and generally accepted, for example, by the FDA. It would be of interest to know whether the results we describe are reflected in other measures of clinical response. In addition, because the focus of this review is on changes in the characteristics of studies of MDD during the last 2 decades, the unit of analysis is the study, not the patient. As a consequence, we have very limited ability to examine the relationship between patient characteristics and the likelihood of response. For example, we could examine only the relationship between average patient age and average response rate across studies, not the relationship between an individual’s age and the likelihood that he or she would respond.

In summary, our review of recent placebo-controlled trials of medication for MDD indicates that the response to placebo is variable but often substantial and has increased in recent years. These data support the view that the inclusion of a placebo group has major scientific importance in trials of new antidepressant medications and indicate that efforts should continue to minimize the risks of such studies so that they may be conducted in an ethically acceptable manner.

REFERENCES

The historical standard as a valid benchmark against which to measure the efficacy of new antidepressants. Since the response to placebo is variable, often substantial, and increasing, it is not surprising that in many randomized controlled trials the response associated with placebo is similar to that associated with an established antidepressant. Nonetheless, the observation that able to assess the change in effect size significant difference. In addition, suf-...
PLACEBO RESPONSE IN DEPRESSION STUDIES

PLACEBO response in depression studies is a topic of great interest in the field of psychiatry. This phenomenon is often studied in controlled trials to assess the efficacy of new treatments for depression. Placebo response can vary significantly among patients, and understanding its role is crucial for the development of effective treatment strategies.

Several factors contribute to placebo response. One of the most significant is the placebo effect, which is the improvement in symptoms observed in patients given a placebo treatment. This effect is not limited to depression but is observed in other areas of medicine as well.

In the context of depression, placebo response can influence the outcomes of clinical trials, potentially masking the true effectiveness of a new medication. This is why researchers and clinicians are keen on understanding the mechanisms behind placebo response and its impact on treatment outcomes.

Research in this area has shown that various factors, including patient expectations, the context of the clinical trial, and the characteristics of the placebo itself, can all influence placebo response. Efforts to standardize placebo administration and control for these variables are ongoing to ensure more accurate assessment of treatment efficacy.

Overall, the study of placebo response in depression studies is a complex and dynamic field that continues to evolve as new research findings are reported. Understanding placebo response is essential for improving the quality of clinical trials and developing more effective treatments for depression.