

## Antidepressants and Placebos: Secrets, Revelations, and Unanswered Questions

**Irving Kirsch**

University of Connecticut

**Alan Scoboria**

University of Connecticut

**Thomas J. Moore**

George Washington University

— We are very heartened by the thoughtful responses to our article. Unlike some of the responses to a previous meta-analysis of antidepressant drug effects ( [Kirsch & Sapirstein, 1998](#)), there is now unanimous agreement among commentators that the mean difference between response to antidepressant drugs and response to inert placebo is very small. It is so small that, despite sample sizes involving hundreds of participants, 57% of the trials funded by the pharmaceutical industry failed to show a significant difference between drug and placebo. Most of these negative data were not published (see [Thase, 2002](#)) and were accessible only by gaining access to U.S. Food and Drug Administration (FDA) documents.

— The small difference between the drug response and the placebo response has been a "dirty little secret" ( [Hollon, DeRubeis, Shelton, & Weiss, 2002](#)), known to researchers who conduct clinical trials, FDA reviewers, and a small group of critics who analyzed the published data and reached conclusions similar to ours (e.g., [Greenberg & Fisher, 1989](#)). It was not known to the general public, depressed patients, or even their physicians.<sup>1</sup> We are pleased that our effort facilitates dissemination of this information.

— The pharmaceutical company data that we have analyzed reveal a mean drug/placebo difference of less than 2 points on the Hamilton depression scale (HAM-D), a difference that is not clinically significant (see [Jacobson, Roberts, Berns, & Mcglinchey, 1999](#)). Another way of describing this difference is in terms of response rates. [Thase \(2002\)](#), for example, notes that 35% - 50% of patients respond to medication compared with 25% - 30% who respond to placebo (see also [Brown, 2002](#); [Hollon et al., 2002](#)). This is interpreted as indicating that 10% - 20% of depressed clinical trial patients show a true drug effect, in which case it follows that 80% - 90% of these patients do not.

— Response rate increments of 10% - 20% seem clinically important to some commentators (e.g., [Thase, 2002](#)) who have interpreted them as indicating the percentage of patients who show a strong response to antidepressants but would not be helped by placebo. This interpretation is mistaken ( [Moncrieff, 2002](#)). To classify patients as responders or nonresponders, one must establish a cutoff point (e.g., a 50% decrease in depressive symptoms; [Mulrow et al., 1999](#)). Consider two patients with baseline HAM-D scores of 20. One is randomized to the drug condition and shows a 10-point improvement on the HAM-D. The other is assigned to the placebo group and shows a 9-point improvement. The first would be classified as a responder and the second as a nonresponder, but in fact, the difference in response (1 point) is negligible.

— What makes this example particularly important is that these two patients are typical, rather than exceptions. A 50% drug response is the median in published clinical trials ( [Mulrow et al., 1999](#)), and a

10-point decrease was the mean drug response in the FDA data set. Thus, the 10%-20% difference between drug and placebo response rates could be due entirely to those patients showing a moderately strong response (just above criterion) to drug and those showing almost as strong a response (just below criterion) to placebo. It would be obtained if there were not even one patient with a strong true drug effect. <sup>2</sup> Of course, it could also be obtained if there were some patients with a larger true drug effect and others on which the drug had a negative effect or no effect at all. This would be true if there were a hidden moderator variable, a possibility raised by some reviewers and discussed by us below.

— We have suggested two possible explanations for the small differences between the drug response and the placebo response: Either drug and placebo effects are not additive (in which case, conventional clinical trials are an inappropriate means of evaluating drug effects because they could lead to the rejection of truly effective medications, the benefits of which are largely masked by placebo), or the drug effect is very small. The commentators have raised three additional possibilities. They suggest that (a) there may be flaws in the ways in which trials sponsored by the pharmaceutical companies are conducted, (b) the small mean differences between drug and placebo obscure strong drug effects produced by some antidepressants in a subset of depressed patients, and (c) the effects of medication may be more stable than those of placebo. In the present reply to the commentaries, we consider each of these possibilities. We also consider the suggestion that the use of antidepressants is justified, no matter how small the effect, as long as it is statistically significant ( [Brown, 2002](#); [Moerman, 2002](#); [Salamone, 2002](#); [Thase, 2002](#) ).

### The Adequacy of the Clinical Trials

— The medications evaluated in the clinical trials that comprise the FDA data set were sponsored by the companies that manufacture them and therefore stand to benefit financially from a positive outcome. So, if there were any biases in these trials, one would expect them to favor the investigated drug ( [Antonuccio, Burns, & Danton, 2002](#) ). Indeed, a quantitative review of the factors affecting response to antidepressants in clinical trials indicated that sponsorship of a trial by a drug company tended to produce effects favoring the sponsor's product ( [Freemantle, Anderson, & Young, 2000](#) ). Some of the commentators on our analysis of the FDA data set have suggested that the opposite may be the case. [Brown \(2002\)](#) suggests that zealous researchers inflate baseline scores so that mildly depressed potential participants are not screened out (see also [Thase, 2002](#) ). [Hollon et al. \(2002\)](#) maintain that low doses may be used to minimize side effects.

— [Brown's \(2002\)](#) suggestion is particularly troubling; it is tantamount to charging that clinical researchers have been fudging the data. Even more troubling is the news that this charge has been confirmed by a spokesperson for the FDA ( [Elias, 2002](#) ). Inflating baseline scores also inflates the apparent benefit produced by treatment. Thus, both the drug response and the placebo response might be overestimated. These data were the basis on which the medications were approved by the FDA. If they are suspect, then perhaps the decision to approve the medications should be reconsidered.

— [Hollon et al. \(2002\)](#) suggest that the manufacturers of fluoxetine may have used low doses in their clinical trials, so as to be able to claim low levels of side effects when marketing their product. As a result, the therapeutic effect might also have been underestimated. This possibility is contradicted by the data. Fluoxetine is prescribed in dosages ranging from 20 to 60 mg daily. The two dose-response studies submitted by Eli Lilly to the FDA evaluated fluoxetine doses of 20, 40, and 60 mg. In one of these trials (conducted on mildly depressed patients), no significant differences were found between doses or between

any dose and placebo. In the other study (conducted on moderately to severely depressed patients), the two lower doses were significantly more effective than the high dose, which was not significantly more effective than placebo. According to the package label for fluoxetine, a particularly high dose (80 mg) was used in other clinical trials submitted to the FDA.

### Moderator Variables

It has been suggested that some patients may be more responsive than others to medication or that some antidepressants are more effective than others ( [Brown, 2002](#); [Hollon et al., 2002](#); [Thase, 2002](#)). If this is the case, then drug/placebo differences, collapsed across patients and drugs, might underestimate the effect that some medications have on some patients.

Our analysis included six different antidepressant medications. Among those medications for which complete data were reported, the range of drug/placebo differences was between 1 and 3 points on the HAM-D. Even for those medications for which data on trials with negative results were withheld, the highest mean difference was 3.21 points. Thus, if there are any mean differences between these medications, they must be very small.

[Hollon et al. \(2002\)](#) express the concern that excluding estimates for the three medications for which scores for unsuccessful trials were not reported might result in the inadvertent omission of "several of the more powerful current medications" (¶ 10). They make specific reference to venlafaxine as a medication that might "produce the most robust effects" (¶ 10) because it works on multiple neurotransmitter systems. This concern is not well founded. First, venlafaxine was *not* one of the medications that were excluded in estimating effects across medication. Second, a meta-analysis of 105 clinical trials ( [Freemantle et al., 2000](#)) indicates that drugs that work on multiple neurotransmitter systems are not more effective than purely serotonergic medications. Third, the weighted mean drug/placebo difference for the three excluded medications was 2.31 HAM-D points. Because this is an overestimate, in which one third of the trials (those showing particularly poor results for the active drug) were excluded, the real drug/placebo difference for the excluded medications cannot be significantly more than that of the medications for which full data were reported.

If the medications do not produce significantly different responses, perhaps there are patient differences that mask more potent pharmacological effects. In support of this contention, [Hollon et al. \(2002\)](#) presented data from an unpublished study of paroxetine, in which patients were matched (post hoc) on rank order of change scores (i.e., the patient showing the least improvement in the active drug condition was matched with the patient showing the least improvement in the placebo condition, etc.). Drug/placebo differences between change scores of matched patients varied as a function of change score rank, but even among those pairs of participants showing the most change, the mean difference was only 3.39 points on the HAM-D.

[Hollon et al. \(2002\)](#) also report percentages of pairs of patients showing differences of 4 points or greater and differences of 6 points or greater. They interpret the existence of these differences as indicating that some patients show a greater drug effect than do others. However, measurement error always produces a distribution around a mean, and Hollon et al.'s data provide no reason to suspect any other explanation. Even if the true drug/placebo differences were always between 2 and 3 points, some of the observed differences would be greater and some would be smaller. Further, there is an exceptionally high

degree of error in the scores assessed by Hollon et al. Change scores are notoriously unreliable, even when pre- and posttest scores are highly reliable ( [Campbell & Kenny, 1999](#)). Differences between change scores are doubly unreliable. This unreliability is evident in the study described by Hollon et al., in which the pattern of differences between drug and placebo change scores varied substantially between the two sites at which the study was conducted.

☰ One method of coping with unreliability is to examine mean scores from large groups of participants rather than individual participants or (as in the case of [Hollon et al.'s \[2002\]](#) data) pairs of participants. Even with relatively large samples, differences in sample size produce differences in the reliability of the data. [Figure 1](#) displays drug/placebo differences in the FDA data set as a function of sample size.

☰ These differences vary widely in small samples, but they are substantially more reliable in large samples. In addition, there is a tendency for large samples to show smaller drug/placebo differences. As seen in [Figure 1](#), a mean difference of about 2 points on the HAM-D characterizes drug/placebo differences in the larger (and therefore, more reliable) studies.

☰ The most frequently proposed hypothetical moderator is baseline severity. The hypothesis is that more severely depressed patients show a greater response to medication and a smaller response to placebo. In the FDA data set, there was only one trial conducted on mildly depressed patients. Three trials were conducted on severely depressed, hospitalized inpatients, but the data from two of these trials were not reported because significant drug/placebo differences were not found. The vast bulk of the trials was conducted on patients judged to be moderately to severely depressed, and it is with these patients that the drug/placebo difference was approximately 2 points on the HAM-D. Thus, the drug/placebo difference appears to be negligible for most of the patients to whom antidepressants are prescribed.

☰ [Figure 2](#) displays the relation between baseline scores and improvement in the larger (  $N = 200$  or greater), and therefore more reliable, clinical trials described in the FDA data set.

☰ More severely depressed patients showed greater improvement, but this is true of patients treated with placebo as well as those treated with medication. This is due to regression toward the mean ( [Campbell & Kenny, 1999](#)), which almost always produces a correlation between baseline scores and change scores. <sup>3</sup>

3

☰ The slope of the regression line appears steeper for drug than for placebo, suggesting a greater drug/placebo difference among more severely depressed patients, but a regression analysis indicates that this difference is not statistically significant (  $p > .20$ ). Nevertheless, because tests of differences in slope have very low power, a greater drug/placebo difference among more severely depressed patients cannot be ruled out. Therefore, we calculated the drug/placebo differences observed in the six large studies with the lowest baseline depression scores (range = 17.21 - 24.25) and compared them to the differences observed in the six large studies with the highest baseline depression scores (range = 25.15 - 27.85). Mean drug/placebo differences were 1.46 points in studies with lower baseline scores and 2.56 in studies with higher baseline scores. Further, this is likely to be an overestimate of the drug/placebo difference for the more severely depressed patients because it does not include the unreported data from two trials conducted on severely depressed, hospitalized inpatients, neither of which showed a significant difference between drug and placebo. The bottom line is that even in studies with more severely depressed patients, there is a strong placebo response and a relatively small difference between drug and placebo.

Finally, it is possible that different patients respond to different medications ( [Hollon et al., 2002](#)). Although there are scant experimental data in support of this hypothesis, clinicians report that patients who do not respond to a particular antidepressant sometimes respond when given a different medication. The results of an early study on the prevention of nausea are pertinent to this issue ( [Wolf, Doering, Clark, & Hagans, 1957](#)). Patients were successively given seven different treatments for nausea, following which ipecac was administered. Response was defined as blocking the emetic effects of ipecac. There was substantial variability in response rates. Some participants responded to all treatments; some to none. Most responded to some treatments but not to others. Thus, the pattern of responding was similar to that observed clinically, following the administration of antidepressants: Patients who did not respond to a given medication often responded to another. In the [Wolf et al. \(1957\)](#) study, however, all seven treatments were known to be ineffective for the condition being treated. They were placebos. These data demonstrate that the response to placebo is inconsistent. Thus, finding that patients sometimes respond after switching to a different antidepressant is exactly what one would expect if antidepressants were nothing more than active placebos.

The most extensive study of the hypothesis that different patients respond to different medications is the [Freemantle et al. \(2000\)](#) meta-analysis. They reviewed 105 clinical trials in which serotonin uptake inhibitors (SSRIs) were compared with other types of antidepressant and used regression analysis to find predictors of differential outcome. Hypothesized predictors included pharmacological action (noradrenaline reuptake inhibition, serotonin reuptake inhibition, and 5-HT<sub>2</sub> receptor antagonism, dual action, and triple action), treatment setting (inpatient vs. outpatient), dose of the comparator drug, method of analysis (LOCF vs. completer data), age of the patient, measurement scale (HAM-D vs. any other), and sponsor of the trial. Only one of these factors bordered on significance: There was "a trend towards increased efficacy of the sponsor's drug" ( [Freemantle et al., 2000](#), p. 294).

[Freemantle et al. \(2000\)](#) did not use some of the factors that the commentators cited as potential predictors of differential outcome, and one can never rule out the possibility of undetected moderator variables. But if there are hidden moderators, the overall mean difference between drug and placebo (2 points on the HAM-D) constrains the conclusions that can be drawn from them. If the mean drug/placebo difference is greater than 2 points for a subset of medications or patients, then it must be less than 2 points for the others. For example, if the mean difference between drug and placebo is 4 points for half of the patients (which is still a rather small drug effect), then the mean effect of antidepressants on the other patients must be 0, and if it is more than 4 points for half the patients, then the medications must be causing harm to at least some others who would fare better on placebo.

### Is the Drug Response More Stable Than the Placebo Response?

Our analysis was limited to the acute efficacy studies submitted to the FDA. It is possible that short-term drug/placebo differences are negligible but that drug effects may be more stable than placebo effects, resulting in lower relapse rates ( [Brown, 2002](#); [Hollon et al., 2002](#); [Thase, 2002](#)), just as psychotherapy effects appear to be more stable than medication effects ( [Hollon, Shelton, & Loosen, 1991](#)). Before looking at the data, however, it is necessary to consider two issues: the way in which the outcome is described and the design of long-term studies.

There are two ways in which long-term outcome can be described. One is in terms of relapse following initial improvement; the other is in terms of continued response to treatment. [Thase \(1999\)](#), for example,

reported 12-month relapse rates of 20% with venlafaxine and 34% with placebo in a pooled analysis of four clinical trials. Thus, there were 1.5 times more relapses with placebo than with medication. Described in terms of continued efficacy, however, these same data indicate that, after 12 months, 80% of the patients receiving medication remained in remission, compared with 66% of patients on placebo. Thus, 83% of the effect of medication was duplicated by placebo.<sup>4</sup> That this effect is similar to that seen in the short-term trials rules out the possibility that antidepressant drug effects might be substantially greater over a longer period of treatment.

— A second issue that must be considered in understanding the results of long-term studies is the design of the study. There are two experimental designs that have been used. The more common design is the relapse prevention trial, in which patients who have responded to medication are then randomized to either continue on medication or be switched to placebo. A less common method of assessing long-term response is with continuation or extension trials, in which patients who have responded to either drug or placebo during an acute efficacy trial are continued on the treatment to which they were originally assigned.

— The relapse prevention design is particularly biased against finding placebo effects. First, the clinical trials from which the patients are drawn typically begin with placebo washout periods, in which all patients who respond to placebo are eliminated from the study. Thus, participation in the relapse prevention trial is limited to patients who have not only responded to the active medication but have also failed to respond to placebo. Second, this design may exacerbate the problem of breaking blind on the basis of perceived side effects. It seems likely that a person who has been on an active medication and is then switched to placebo will be likely to detect the change.

— Despite this bias, relapse prevention trials often show large placebo responses. For example, one of the two 24-week relapse prevention trials reported to the FDA by the manufacturer of citalopram showed 2-year response rates of 69% for placebo and 90% for citalopram. The other showed response rates of 76% for placebo and 86% for citalopram. Thus, in these studies, the long-term placebo response among patients who had been successfully treated with an active antidepressant was between 77% and 88% of the medication response. Similar results were reported by [Walach and Maidhof \(1999\)](#) in a meta-analysis of relapse prevention trials published between 1973 and 1990. In these trials, 71% of the drug response was duplicated by placebo. Walach and Maidhof also examined response to treatment as a function of the duration of the trial. Their data indicate that responses to both drug and placebo decrease over time. Contrary to conventional wisdom, however, the correlation between duration of the trial and response to treatment was higher for active medication ( $r = -.84$ ) than for placebo ( $r = -.62$ ), suggesting a steeper decline in effectiveness for active drugs than for placebo.

— Continuation trials are relatively rare, but the data confirm the relative stability of the placebo response over time. As [Antonuccio et al. \(2002\)](#) note, the 18-month followup of the NIMH Collaborative Depression Study ( [Shea et al., 1992](#)) reported equivalent long-term outcomes for both drug and placebo, with a lower relapse rate among remitted patients in the placebo group. A more recent study compared hypericum, sertraline, and placebo ( [Hypericum Depression Trial Study Group, 2002](#)) and reported 26-week continuation data for patients who had responded to treatment during the 8-week clinical trial. Of the 79 patients who entered the continuation phase, none of the placebo patients relapsed, none of the sertraline patients relapsed, and only one taking hypericum relapsed.

## What Do We Do Now?

As many of the commentators have noted, the basic findings of our analysis are clear and undisputed. The mean difference between response to antidepressant medication and response to placebo is very small. This raises the question of what should be done in clinical practice. The response to both drug and placebo is substantial. How is this therapeutic benefit to be elicited if these medications, which may be little more than active placebos, are abandoned? Some commentators have argued that they should be given to depressed patients, even if their pharmacological effect is negligible. If nothing else, they can be the vehicle by means of which the placebo response is elicited ( [Hollon et al., 2002](#); [Moerman, 2002](#)).

This argument rests on the assumption that a genuine pharmacological effect has been proven. The emperor's new clothes may not be the elegant suit his subjects were led to expect, but at least they are made with real fabric ( [Thase, 2002](#)). We have not denied that antidepressant medication might have genuine pharmacological effects. In fact, we raised the possibility that its effects could be much larger and reliable than the data suggest. The problem is that the effects of antidepressant medication are unknown. They may be very large, vanishingly small, or completely nonexistent. The emperor may be wearing an elegant invisible suit, a fig leaf, or nothing at all.

It is true that, on average, there is a significantly greater response to medication than to placebo. What is not yet known is the reason for this difference in response. It may be a drug effect, but it may also be an enhanced placebo effect associated with the perception of side effects and the breaking of blind ( [Greenberg, 2002](#); [Greenberg & Fisher, 1989](#)).<sup>5</sup> It is known that antidepressants produce significantly more side effects than do those observed in inert placebo control groups ( [Mulrow et al., 1999](#)), patients assigned to conventional antidepressants are able to break blind to a significant degree ( [Rabkin et al., 1986](#)), and a drug effect has not been convincingly demonstrated in studies using active placebos ( [Moncrieff, Wessely, & Hardy, 2001](#)), despite the fact that the medications used as active placebos may not prevent patients from breaking blind ( [Hollon et al., 2002](#)). These data do not prove that the drug/placebo difference is due to the breaking of blind, but they do suggest this as a reasonable possibility, one that needs to be ruled out before the claim of even minimal drug effectiveness can be considered to have been substantiated ( [Greenberg, 2002](#)).

In addition, the additivity assumption needs to be tested directly. The indirect data are mixed. As [Brown \(2002\)](#) notes, data showing different brain changes in placebo responders and drug responders ( [Leuchter, Cook, Witte, Morgan, & Abrams, 2002](#)) suggest that the effects may not be additive. Conversely, data showing similar brain changes in placebo responders and drug responders ( [Mayberg et al., 1999](#)) suggest a common mechanism, which is consistent with the additivity hypothesis. Additivity is also suggested by the parallel increases in response rates for SSRIs and placebos over the years ( [Walsh, Seidman, Sysko, & Gould, 2002](#)). Since the composition of the medication has not changed, the change must be due to other factors. We suspect that it is due to the dissemination of information (or misinformation) about the marvelous success of the newer antidepressants.

We have raised the possibility of using the balanced placebo design, perhaps with active placebos as an aid in preserving the manipulation, as the most straightforward test of the additivity hypothesis. [Hollon et al. \(2002\)](#) have raised some useful concerns about the success of using active placebos for this purpose. [Salamone \(2002\)](#) has provided suggestions that might address some of those concerns, and [Antonuccio et al. \(2002\)](#) have suggested an alternative (assessing the degree to which blind is broken) that should be implemented routinely in pharmacological research but that does not obviate the need for direct tests of the additivity assumption.

☰ There also are other questions that need to be answered. We need to know more about the placebo effect and its underlying mechanisms. [Rehm \(2002\)](#) has provided many insightful research suggestions aimed at addressing this issue, and we hope that clinical researchers will implement them. We also think that [Salamone's \(2002\)](#) suggestions for new studies of the dose-response relationship are worthwhile. The data sent to the FDA by the pharmaceutical companies included studies in which up to four different doses were evaluated, and except for the finding that lower doses of fluoxetine were more effective than a high dose, they failed to find reliable linear or quadratic differences between doses. It is possible, as [Salamone \(2002\)](#) suggests, that studies involving six or seven different doses would reveal some more complex relationships. In addition, there is a need for additional tests of possible moderator variables (chronicity, cortisol secretion, etc.).

☰ In the meantime, what are the alternatives for treating patients? Imagine having a choice between four treatments. Treatment A produces a large therapeutic response but also a large number of adverse effects, including diarrhea, nausea, anorexia, sweating, forgetfulness, bleeding, seizures, anxiety, mania, sleep disruption, and sexual dysfunction. Treatments B and C produce therapeutic responses that are almost as great as those produced by treatment A, but without the adverse effects. In fact, the side effects produced by Treatment B are beneficial (e.g., better general physical health). However, the therapeutic effects of Treatments B and C have been evaluated in relatively few studies. Treatment D has been assessed in many comparative studies, in which it has been found to be as effective as Treatment A in the short term and more effective in the long term. It does not produce adverse effects. Given a choice between these alternatives, which would you choose?

☰ Of course, these alternatives are not merely hypothetical. Treatment A corresponds to SSRIs, and the list of side effects is drawn from those that have been shown to be produced by these medications ([Antonuccio, Danton, DeNelsky, Greenberg, & Gordon, 1999](#); [Mulrow et al., 1999](#)). Treatment B is physical exercise, which has been reported to have lasting therapeutic benefits in the treatment of major depression ([Babyak et al., 2000](#)). It may be nothing more than a placebo, but if so, it is one with desirable rather than adverse side effects. Treatment C is bibliotherapy (e.g., [Burns, 1999](#)), another low-cost treatment with demonstrated effectiveness ([Jamison & Scogin, 1995](#); [Smith, Floyd, Jamison, & Scogin, 1997](#)) and little danger of side effects. Treatment D is psychotherapy. As noted by [Antonuccio et al. \(2002\)](#), "psychotherapy (particularly cognitive therapy, behavioral activation, and interpersonal therapy) compares favorably with medications in the short term, even when the depression is severe (e.g., [DeRubeis, Gelfand, Tang, & Simons, 1999](#)), and appears superior to medications in long-term comparative studies (Antonuccio et al. 1995; Hollon, Shelton, & Loosen, 1991)" (<http://journals.apa.org/prevention/volume5/pre525c.html#p24> ¶ 24). Given these data, antidepressant medication might best be considered a last resort, restricted to patients who refuse or fail to respond to other treatments.

☰ It may well be that the effects of psychotherapy are due to expectancy, conditioning, and other psychological factors that have been hypothesized to be the basis of placebo effect ([Kirsch, 1997](#)). Indeed, changing maladaptive expectations is an essential cornerstone of cognitive therapy. But this does not negate its effectiveness. The contention (e.g., [Salamone, 2002](#); [Thase, 2002](#)) that the logic of placebo-controlled evaluation should be extended to psychotherapy is mistaken ([Kirsch, 1978](#)). In drug research, placebos are used to distinguish the pharmacologically produced effects of substance administration from its psychologically produced effects. However, there are no pharmacologically produced effects of psychotherapy. The effects of a psychological intervention can only be due to psychological factors. Control conditions consisting of alternative psychological treatments (including those erroneously



called placebos) are useful in establishing which psychological factors are important in producing the effects of a particular therapy but not in evaluating the efficacy of that treatment. Like placebos, effective psychotherapies may accomplish their effects by changing expectations ( [Kirsch, 1985, 1990, 1999](#)), but unlike placebos, they do so without deception.

---

## References

- Antonuccio, D. O., Burns, D. D., & Danton, W. G. (2002). Antidepressants: A triumph of marketing over science? *Prevention & Treatment*, 5, Article 25. Available on the World Wide Web: <http://www.journals.apa.org/prevention/volume5/pre525c.html>.
- Antonuccio, D. O., Danton, W. G., DeNelsky, G. Y., Greenberg, R. P., & Gordon, J. S. (1999). Raising questions about antidepressants. *Psychotherapy and Psychosomatics*, 68, 3-14. PsycINFO Article
- Babyak, M., Blumenthal, J. A., Herman, S., Khatri, P., Doraiswamy, M., Moore, K., Craighead, E., Baldewicz, T., & Krishnan, K. R. (2000). Exercise treatment for major depression: Maintenance of therapeutic benefit at 10 months. *Psychosomatic Medicine*, 62, 633-638. PsycINFO
- Brown, W. A. (2002). Are antidepressants as ineffective as they look? *Prevention & Treatment*, 5 <http://www.journals.apa.org/prevention/volume5/pre526c.html>.
- Burns, D. D. (1999). *Feeling good: The new mood therapy* (Rev. Ed.). New York: Avon.
- Campbell, D. T., & Kenny, D. A. (1999). *Primer on regression artifacts*. New York: Guilford Press.
- DeRubeis, R. J., Gelfand, L. A., Tang, T. Z., & Simons, A. D. (1999). Medications versus cognitive behavior therapy for severely depressed outpatients: Meta-analysis of four randomized comparisons. *American Journal of Psychiatry*, 156, 1007-1013. PsycINFO
- Elias, M. (2002, July 7). Study: Antidepressant barely better than placebo. *USA Today*, Retrieved July 8, 2002 from <http://usatoday.com/news/healthscience/health/drugs/2002-07-08-antidepressants.htm>
- Freemantle, N., Anderson, I. M., & Young, P. (2000). Predictive value of pharmacological activity for the relative efficacy of antidepressant drugs: Meta-regression analysis. *British Journal of Psychiatry*, 177, 292-302. PsycINFO Article
- Greenberg, R. P. (2002). Reflections on the emperor's new drugs. *Prevention & Treatment*, 5, Article 27. [Available on the World Wide Web: http://www.journals.apa.org/prevention/volume5/pre527c.html](http://www.journals.apa.org/prevention/volume5/pre527c.html).
- Greenberg, R. P., & Fisher, S. (1989). Examining antidepressant effectiveness: Findings, ambiguities, and some vexing puzzles. In S. Fisher & R. P. Greenberg (Eds.), *The limits of biological treatments for psychological distress: Comparisons with psychotherapy and placebo* (pp. 1-37). Hillsdale, NJ: Erlbaum.
- Hollon, S. D., DeRubeis, R. J., Shelton, R. C., & Weiss, B. (2002). The emperor's new drugs: Effect size and moderation effects. *Prevention & Treatment*, 5, Article 28. Available on the World Wide Web: <http://www.journals.apa.org/prevention/volume5/pre528c.html>.
- Hollon, S. D., Shelton, R. C., & Loosen, P. T. (1991). Cognitive therapy and pharmacotherapy for depression. *Journal of Consulting & Clinical Psychology*, 59, 88-99. PsycINFO Article
- Hypericum Depression Trial Study Group. (2002). Effect of *Hypericum Perforatum* (St John's wort) in major depressive disorder: A randomized controlled trial. *Journal of the American Medical Association*, 287, 1807-1814. PsycINFO Article
- Jacobson, N. S., Roberts, L. J., Berns, S. B., & McGlinchey, J. B. (1999). Methods defining and determining the clinical significance of treatment effects: Description, application, and alternatives. *Journal of Consulting & Clinical Psychology*, 67, 300-307. PsycINFO Article

- Jamison, C., & Scogin, F. (1995). Outcome of cognitive bibliotherapy with depressed adults. *Journal of Consulting & Clinical Psychology*, 63, 644-650. [PsycINFO](#) [Article](#)
- Khan, A., Warner, H. A., & Brown, W. A. (2000). Symptom reduction and suicide risk in patients treated with placebo in antidepressant clinical trials: An analysis of the Food and Drug Administration database. *Archives of General Psychiatry*, 57, 311-317. [PsycINFO](#) [Article](#)
- Kirsch, I. (1978). The placebo effect and the cognitive-behavioral revolution. *Cognitive Therapy and Research*, 2, 255-264. [PsycINFO](#)
- Kirsch, I. (1985). Response expectancy as a determinant of experience and behavior. *American Psychologist*, 40, 1189-1202. [PsycINFO](#) [Article](#)
- Kirsch, I. (1990). *Changing expectations: A key to effective psychotherapy*. Pacific Grove, CA: Brooks/Cole.
- Kirsch, I. (1997). Specifying nonspecifics: Psychological mechanisms of placebo effects. In A. Harrington (Ed.), *The placebo effect: An interdisciplinary exploration* (pp. 166-186). Cambridge, MA: Harvard University Press.
- Kirsch, I. (1999). *How expectancies shape experience*. Washington, DC: American Psychological Association.
- Kirsch, I., & Sapirstein, G. (1998). Listening to Prozac but hearing placebo: A meta analysis of antidepressant medication. *Prevention & Treatment*, 1, Article 0002a. Available on the World Wide Web: <http://www.journals.apa.org/prevention/volume1/pre0010002a.html>
- Leber, P. (1998, May 4). *Approvable action on Forrest Laboratories, Inc. NDA 20-822 Celexa (citalopram HBr) for the management of depression*. Memorandum to the Department of Health and Human Services, Public Health Service, Food and Drug Administration, Center for Drug Evaluation and Research. Washington, DC.
- Leuchter, A. F., Cook, I. A., Witte, E. A., Morgan, M., & Abrams, M. (2002). Changes in brain function of depressed subjects during treatment with placebo. *American Journal of Psychiatry*, 159, 122-129. [PsycINFO](#) [Article](#)
- Mayberg, H. S., Liotti, M., Brannan, S. K., McGinnis, S., Mahurin, R. K., Jerabek, P. A., Silva, A., Tekell, J. L., Martin, C. C., Lancaster, J. L., & Fox, P. T. (1999). Reciprocal limbic-cortical function and negative mood: Converging PET findings in depression and normal sadness. *American Journal of Psychiatry*, 156, 675-682. [PsycINFO](#)
- Moerman, D. E. (2002). "The Loaves and the Fishes": A Comment on "The Emperor's New Drugs: An Analysis of Antidepressant Medication Data Submitted to the U.S. Food and Drug Administration." *Prevention & Treatment*, 5, Article 29. Available on the World Wide Web: <http://www.journals.apa.org/prevention/volume5/pre529c.html>.
- Moncrieff, J. (2002). The antidepressant debate. *British Journal of Psychiatry*, 180, 193-194. [PsycINFO](#) [Article](#)
- Moncrieff, J., Wessely, S., & Hardy, R. (2001). Antidepressants using active placebos [Cochrane Review]. *Cochrane Database Systematic Review*, 2 CD003012.
- Mulrow, C. D., Williams, J. W., Jr., Trivedi, M., Chiquette, E., Aguilar, C., Cornell, J. E., Badgett, R., Noel, P. H., Lawrence, V., Lee, S., Luther, M., Ramirez, G., Richardson, W. S., & Stamm, K. (1999). *Treatment of depression: Newer pharmacotherapies. Evidence Report/Technology Assessment No. 7* (AHCPR Publication No. 99-E014; prepared by the San Antonio Evidence-based Practice Center based at the University of Texas Health Science Center at San Antonio under Contract 290-97-0012). Rockville, MD: Agency for Health Care Policy and Research.
- Muñoz, R. (2002). Comment on Kirsch, Moore, Scoboria, and Nicholls (2002). *Prevention & Treatment*, 5,

Article 30. Available on the World Wide Web:

<http://www.journals.apa.org/prevention/volume5/pre530c.html>.

Rabkin, J. G., Markowitz, J. S., Stewart, J. W., McGrath, P. J., Harrison, W., Quitkin, F. M., & Klein, D. F. (1986). How blind is blind? Assessment of patient and doctor medication guesses in a placebo-controlled trial of imipramine and phenelzine. *Psychiatry Research*, *19*, 75-86. PsycINFO

Rehm, L. P. (2002). How can we better disentangle placebo and drug effects? *Prevention & Treatment*, *5*, Article 31. Available on the World Wide Web:

<http://www.journals.apa.org/prevention/volume5/pre531c.html>

Salamone, J. D. (2002). Antidepressants and placebos: Conceptual problems and research strategies. *Prevention & Treatment*, *5*, Article 24. Available on the World Wide Web:

<http://www.journals.apa.org/prevention/volume5/pre524c.html>

Shea, M. T., Elkin, I., Imber, S. D., Sotsky, S. M., Watkins, J. T., Collins, J. F., Pilkonis, P. A., Beckham, E., Glass, D. R., Dolan, R. T., & Parloff, M. B. (1992). Course of depressive symptoms over follow-up: Findings from the National Institute of Mental Health Treatment of Depression Collaborative Research Program. *Archives of General Psychiatry*, *49*, 782-787. PsycINFO

Smith, N. M., Floyd, M. R., Jamison, C., & Scogin, F. (1997). Three-year follow-up of bibliotherapy for depression. *Journal of Consulting & Clinical Psychology*, *65*, 324-327. PsycINFO Article

Thase, M. E. (1999). How should efficacy be evaluated in randomized clinical trials of treatments for depression? *Journal of Clinical Psychiatry*, *60* (Suppl. 4), 23-31.

Thase, M. E. (2002). Antidepressant effects: The suit may be small, but the fabric is real. *Prevention & Treatment*, *5*, Article 32. Available on the World Wide Web:  
<http://www.journals.apa.org/prevention/volume5/pre532c.html>.

Walach, H., & Maidhof, C. (1999). Is the placebo effect dependent on time? A meta-analysis. In I. Kirsch (Ed.), *How expectancies shape experience* (pp. 321-332). Washington, DC: American Psychological Association.

Walsh, B. T., Seidman, S. N., Sysko, R., & Gould, M. (2002). Placebo response in studies of major depression. Variable, substantial, and growing. *Journal of the American Medical Association*, *287*, 1840-1847. PsycINFO Article

Wolf, S., Doering, C. R., Clark, M. L., & Hagans, J. A. (1957). Chance distribution and the placebo "reactor." *Journal of Laboratory and Clinical Medicine*, *49*, 837-841.

---

<sup>1</sup>An internal memorandum by the Director of the Division of Neuropharmacological Drug Products indicates FDA awareness of this situation:

The Clinical Efficacy Trials subsection within the Clinical Pharmacology section not only describes the clinical trials providing evidence of citalopram's antidepressant effects, but make mention of adequate and well controlled clinical studies that failed to do so. I am mindful, based on prior discussions of the issue, that the Office Director is inclined toward the view that the provision of such information is of no practical value to either the patient or prescriber. I disagree. I believe it is useful for the prescriber, patient, and 3rd-party payer to know, without having to gain access to official FDA review documents, that citalopram's antidepressants (sic) effects were not detected in every controlled clinical trial intended to demonstrate those effects. I am aware that clinical studies often fail to document the efficacy of effective drugs, but I doubt that the public, or even the majority of the medical community, is aware of this fact. I am persuaded that they not only have a right to know but that they should know. Moreover, I believe that labeling that selectively describes positive studies and excludes mention of negative ones can be viewed as potentially "false and misleading" ([Leber, 1998](#), p. 11).

We agree that the public and the medical community should be informed of these data.

<sup>2</sup>The drug effect is conventionally interpreted as the difference between the response to the drug and the response to placebo. It is that part of the drug response that is due to the pharmacological action of the drug. In contrast, the drug response includes the effect of the drug, the placebo effect, spontaneous remission, regression to the mean, and any other factors that might contribute to changes observed following the administration of medication. Similarly, the placebo effect is that portion of the placebo response that is actually due to the administration of a placebo.

<sup>3</sup>Regression artifacts also affect the correlation between drug response and placebo response. Moerman reviewed prior meta-analyses showing that these correlations range between .43 and .90. Similarly, if the analysis is limited to the more reliable clinical trials in the FDA data set—those with samples of 200 or greater—the correlation between drug response and placebo response in the FDA data set is .72. However, because baseline scores are necessarily correlated with change scores (Campbell & Kenny, 1999), when the baseline scores of two groups are correlated (as they are in meta-analyses of drug and placebo responses), the change scores will also be correlated. For example, when normally distributed baseline scores in two samples are perfectly correlated with each other, and normally distributed posttest scores in these samples are uncorrelated either with each other or with baseline scores, the correlation between the two sets of change scores will be .50.

<sup>4</sup>The way in which this method of describing the data can be misleading can be illustrated by applying it to the [Khan, Warner, and Brown \(2000\)](#) analysis of the data submitted to the FDA. These data indicate that there were twice as many suicide attempts among patients given active medication than among those given placebo. This might be interpreted as indicating that medication doubled the risk of suicide attempts, but this method of describing the data is misleading. The actual rates were 0.8% with antidepressant medication and 0.4% with placebo. Thus, suicide attempts were absent in 99.2% of patients treated with drugs and 99.6% of patients treated with placebo, and as noted by Khan et al. (2000), the difference was not statistically significant.

<sup>5</sup>Contrary to the implication in [Muñoz's \(2002\)](#) commentary, this hypothesis does not suggest duplicity on the part of patients or researchers.

---

Correspondence concerning this article should be addressed to Irving Kirsch, Department of Psychology, University of Connecticut 406 Babbidge Road, U-20, Storrs, CT 06269-1020

E-mail: [irving.kirsch@uconn.edu](mailto:irving.kirsch@uconn.edu)

---

Figure 1. Drug/placebo differences as a function of sample size.

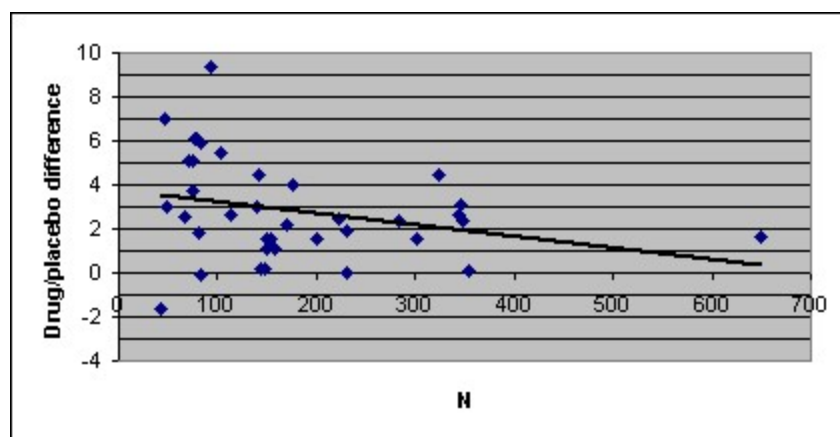


Figure 2. Regression toward the mean in response to drug and placebo.

