

# Methodological Issues in Clinical Trials of Antidepressant Medications: Perspectives from Psychotherapy Outcome Research

Brandon A. Gaudiano<sup>a</sup> James D. Herbert<sup>b</sup>

<sup>a</sup>Department of Psychiatry and Human Behavior, Brown University School of Medicine and Butler Hospital, Providence, R.I., and <sup>b</sup>Department of Psychology, Drexel University, Philadelphia, Pa., USA

## Key Words

Antidepressants/active placebos, psychotherapy outcome research · Psychotherapy outcome research, depression · Randomized controlled trials, antidepressants/active placebos

## Abstract

Despite their widespread use, the specific efficacy of antidepressant medications has been a source of debate in recent years. Examination of the literature reveals that a significant proportion of the benefit produced in antidepressant trials is duplicated in pill placebo conditions. Furthermore, early trials utilizing active placebos, or medications that mimic the common side effects of antidepressants, showed even smaller differences as compared with active medications. We examine issues surrounding the use of placebo control conditions in antidepressant trials, including the pros and cons of active placebos. We conclude that similar challenges are faced by psychotherapy outcome researchers who have focused more on the separation of specific from nonspecific treatment factors and on the effects of researcher allegiance and patient expectancy on outcome. Within this context, recommendations for improving future antidepressant research are discussed.

Copyright © 2005 S. Karger AG, Basel

## Introduction

Depression has been referred to as the ‘common cold’ of mental health problems, based on its high prevalence rates and heterogeneous clinical features. Kessler et al. [1] recently reported that lifetime and 12-month prevalence rates of major depressive disorders are estimated to be 16 and 6%, respectively, with 72% of the lifetime cases also meeting criteria for other psychiatric diagnoses. The newer antidepressant medications, called selective serotonin reuptake inhibitors (SSRIs), are promoted as having improved side effect profiles as compared with older medications, although this issue is very much a matter of debate [2]. The popularity of these medications has contributed to the widespread use of pharmacotherapy in the treatment of depression, making SSRIs the most commonly prescribed class of psychotropics [3]. Over the years, however, critics have suggested that the specific benefit of antidepressants above and beyond placebo effects is relatively small or even nonexistent when examined under rigorously controlled conditions. This debate has even raged in the popular media recently [4]. Therefore, this paper will examine the evidence for the specific benefits of antidepressants as compared with both inert and active placebos (i.e., medications that mimic common side effects of antidepressants) and will make recommendations to improve future studies by paying particular attention to perspectives from psychotherapy outcome research.

## KARGER

Fax +41 61 306 12 34  
E-Mail [karger@karger.ch](mailto:karger@karger.ch)  
[www.karger.com](http://www.karger.com)

© 2005 S. Karger AG, Basel

Accessible online at:  
[www.karger.com/pps](http://www.karger.com/pps)

Brandon A. Gaudiano, PhD  
Psychosocial Research Program, Butler Hospital  
345 Blackstone Blvd.  
Providence, RI 02906 (USA)  
Tel. +1 401 455 6384, Fax +1 401 455 6235, E-Mail [Brandon\\_Gaudiano@brown.edu](mailto:Brandon_Gaudiano@brown.edu)

## Placebo Response in Antidepressant Trials

The inclusion of a pill placebo condition in trials investigating the efficacy of new drugs has been standard practice for over 50 years [5]. The epistemological aim of the typical clinical trial is to provide controlled conditions under which specific treatment effects can be separated. Without adequate comparison conditions, it is impossible to differentiate any specific effects of a drug from 'nonspecific' factors, such as chance variation, regression to the mean, healthcare provider attention, treatment credibility and rationale, persuasion, expectancy effects, allegiance effects, effort justification, spontaneous remission, demand characteristics, and so on [6, 7]. Placebo controls may be particularly important when investigating the benefits of pharmacotherapy for psychiatric problems (e.g., mood and anxiety disorders) which by their very nature pose difficulties in diagnosis and assessment relative to medical conditions with objective, pathognomonic biological markers.

Several meta-analyses have called into question the magnitude of benefits derived from antidepressant treatment beyond placebo effects. For example, Greenberg et al. [8] found only modest effect size differences between active medications and placebos ( $d = 0.19-0.25$ ) in trials comparing a new antidepressant, a standard antidepressant, and a placebo. Furthermore, effect sizes based on clinician ratings showed more improvement as compared with the gains found in patient reports which in contrast were not significantly different from placebo in the older established antidepressants [see 9 for differing conclusions].

In a more recent meta-analysis published in 1998, Kirsch and Sapirstein [10] compared the mean effect size changes in symptoms of depression across 19 double-blind antidepressant efficacy trials. The results demonstrated that placebos reproduced approximately 75% of the improvement found in active drugs. Furthermore, the authors assert that the explanation for the remaining 25% of improvement in the antidepressant condition may be attributable to an enhanced placebo response due to increased side effects experienced by patients taking the active drug or to other unidentified nonspecific factors.

The meta-analysis performed by Kirsch and Sapirstein [10] has been criticized extensively, most notably by Klein [11, 12] in a series of exchanges with Kirsch [13, 14]. Klein criticizes the study on several fronts, most notably arguing that the meta-analysis included a small, unrepresentative sample of antidepressant trials; that the use of completer-only analyses disadvantaged the drug group

due to differential dropout rates; that even if small differences were found, they were still clinically significant, especially when examining the responder status; that averaging sensitive and nonsensitive measures obscured true differences, and that the studies that were included were methodologically flawed. However, Kirsch responded that other meta-analyses including different samples of studies reached similar conclusions; that other meta-analyses using intention-to-treat data found similar results; that clinician-administered and other simple categorical measures are prone to bias and that between-group differences are smaller when self-report measures are examined; that Klein argues for the inclusion of measures that support his position and for the exclusion of those that do not, and that Klein does not criticize trials showing the superiority of antidepressants over placebos, but only those that show no differences.

Recently, Kirsch et al. [15] replicated and extended the results of the meta-analysis performed by Kirsch and Sapirstein [10] using US Food and Drug Administration data. The main advantage of using the Food and Drug Administration database is that it reduces the substantial problem of publication bias in conventional meta-analyses that rely only on published reports [i.e., the so-called 'file drawer problem'; 16]. Kirsch et al. [15] found only an 18% difference between drug and placebo, representing an average 2-point difference on the Hamilton Rating Scale for Depression (HAM-D), a semistructured clinical interview most commonly used to assess the treatment outcome.

Do the findings by Kirsch et al. [15] suggest that the glass is half empty or half full (actually 82 vs. 18% in this case)? Thase [17] concluded that even though nonspecific factors account for most of the variability in antidepressant trials, these meta-analyses still show a reliable difference between drugs and placebos, and even a small difference is clinically significant from a public health perspective. Furthermore, recent evidence suggests that venlafaxine, which is a serotonin-norepinephrine reuptake inhibitor, may be more efficacious than traditional SSRIs [18], suggesting that outcomes may vary between different antidepressants/classes.

A recent large controlled trial provided an example of the difficulty in demonstrating the efficacy of antidepressants over pill placebos. The Hypericum Depression Trial Study Group (HDTSG) [19] conducted a double-blind, randomized trial comparing St. John's wort (*Hypericum perforatum*), sertraline, or inert pill placebo in the treatment of 340 outpatients with major depression. Sertraline and St. John's wort were not significantly different from

placebo in changes on the HAMD by 8 weeks, nor did the two active conditions differ from placebo in rates of full responders. However, sertraline was significantly better than placebo on the Clinical Global Impressions (CGI) improvement scale, an interviewer-rated measure, at the end of treatment. Adverse side effect profiles for St. John's wort and sertraline differed from placebo, raising the possibility that rater bias contributed to the differences on interviewer ratings.

The HDTSG [19] trial holds important implications for pharmacotherapy research in depression. The study emphasizes the importance of including both active and inactive comparison treatments when testing antidepressants. The authors point out that if a placebo condition had not been included, it could have been concluded that St. John's wort was as effective as standard antidepressant treatment. In other words, a trial including only two active comparison treatments that fails to find group differences could indicate that (1) both treatments were effective or that (2) neither treatment was effective, leaving much ambiguity in the interpretation of results. The findings of the HDTSG study highlight the fact that designs including only active comparison conditions address the relative but not necessarily the absolute efficacy of the treatments. The study also provides strong support for the use of a multimodal assessment strategy in psychiatric outcome studies. Pharmacotherapy outcome studies tend to rely heavily on interviewer rating measures. The CGI is comprised of a single item that may be more prone to rater bias than the more extensive ratings derived from the HAMD or the other self-report measures of depression used in the study. Finally, the HDTSG study raises the possibility that rater bias will continue to be a problem even in methodologically stringent randomized controlled trials because of the side effects experienced in active medication versus inert pill placebo conditions.

### **Are Active Placebos the Answer?**

In a recent meta-analysis showing an increasing placebo response over the past 20 years in antidepressant trials, Walsh et al. [20] concluded that the use of a pill placebo arm in trials of investigational drugs should be continued. However, some have questioned the adequacy of inert pill placebo conditions in clinical trials. Much of the controversy stems from the possibility that patients and researchers, even in the context of a double-blind trial, may inadvertently become 'unblinded' due to medication side

effects, such as anticholinergic symptoms (e.g., dry mouth, drowsiness, decreased sweating, blurred vision, cardiac irregularities, nausea) [21]. In other words, unblinding represents the unintentional disclosure of information to the previously naïve assessor regarding which condition the participant was assigned to, thereby potentially biasing the ratings. This also applies to the participants themselves whose expectancies may change based on knowledge of their treatment condition.

#### *The Problems with Unblinding Effects*

Several studies have suggested problems with unblinding in drug trials, although they often have produced more questions than answers. In a study of patients maintained on neuroleptics, Double [22] found that although patients' guesses were no better than chance, raters were better than chance at detecting which patients were on active medication versus placebo. Bystritsky and Waikar [23] examined placebo-controlled trials for anxiety and depression and found that the patients generally were able to identify their study condition. White et al. [24] found similar results in their study, as an independent rater was able to guess accurately the treatment condition of patients with major depression receiving either an inert placebo or a tricyclic antidepressant (TCA). Recently, Piasecki et al. [25] assessed blind integrity in a double-blind trial comparing inert placebo to paroxetine. Reported side effects were significantly higher in the drug as compared with the placebo condition, and clinicians correctly guessed participants' conditions in 12 out of 13 cases after treatment.

Although the inadvertent unblinding of raters or patients due to medication side effects or other factors may result in a bias in favor of the drug group in antidepressant trials, other explanations also are plausible. For example, Sharpe et al. [26] showed that unblinding often is confounded with treatment response. In other words, it is possible that participants figure out that they are in the active drug condition, because they are seeing improvements from taking the medication, in contrast to those who are taking the pill placebo. However, others have found that side effect profiles, but not treatment outcomes, were associated with unblinding, making this issue open to continued debate [27].

#### *The Use of Active Placebo Conditions*

Due to concerns about unblinding, some have suggested that so-called 'active' placebos, or drugs that mimic the common side effects of the active medication without providing the pharmacological agent theorized to be responsible for clinical improvement, should be included in

clinical trials [10, 21]. Providing a comparison condition, in which telltale medication side effects are kept constant between groups, could help to decrease this potential confound.

In an early review, Thomson [28] examined placebo-controlled TCA studies and found that fewer studies showed a significant difference between active placebo and TCA than between inert placebo and TCA. The author concluded that either atropine (the active placebo commonly used in these studies) has antidepressant qualities or that side effects amplify placebo responding.

More recently, meta-analyses have been conducted comparing inert versus active placebos. Moncrieff et al. [29], in their meta-analysis of nine clinical trials, found that all except one study showed a small or nonsignificant difference between antidepressants and active placebo. Moncrieff et al. [30] also conducted a similar meta-analysis for the Cochrane Review (i.e., an independent scholarly group that sponsors evidence-based reviews of treatments). When an outlying and questionable study was omitted from analyses, the effect size of antidepressants over active placebo was 0.17. The confidence interval included 0, indicating that the result was not statistically significant. Quitkin [31] agreed that the excluded study was an outlier. These results suggest that unblinding and expectancy effects may contribute to the differences typically found between antidepressants and inert placebos. However, few trials using active placebos exist for comparison. Furthermore, the ones conducted possess methodological limitations. At the present time, these results remain speculative but suggestive.

In an effort to explain these null results, it has been suggested that the active placebo commonly used, atropine, may itself have antidepressant effects [32, 33]. However, such post hoc explanations of null findings have questionable scientific merit. If an antidepressant fails to outperform a drug not considered to be an antidepressant, one cannot simply declare that the new drug possesses specific antidepressant qualities. This circular reasoning renders tests of antidepressants unfalsifiable by defining away the concept of placebo, when results are unfavorable to the active medication under study. For example, if an inert placebo condition had not been included in the St. John's wort study [19], the authors may have reached the erroneous conclusion that the supplement was an effective treatment in the study, because it was not different than the antidepressant. Neither drug was more effective than placebo in this trial. Perhaps atropine does indeed have antidepressant effects. However, evaluation of this hypothesis will require more than post hoc maneuvers, in

which a placebo condition is redefined as an active treatment, whenever an established treatment fails to outperform it.

Additional attempts to explain away null results also deserve discussion. For example, some argue that failed trials of established drugs are best explained by problems with 'assay sensitivity' (e.g., problems with sample selection, assessment, adherence) in a given study [34]. However, Otto and Nierenberg [35] asserted that requiring a new trial to replicate the results of a previous one in order to be considered valid constitutes a fundamental 'derailment' of the scientific method. Such assay sensitivity arguments are dangerous, because they assume that a trial should obtain certain preordained results before it even is conducted.

Other scientifically stronger arguments have been presented, contesting the need for the inclusion of active placebo conditions in antidepressant trials. Quitkin et al. [9] and Quitkin [31] argue that the reviews by Thomson [28] and Moncrieff et al. [29] are flawed and that their conclusions are contradicted by the data. Thomson [28] reported that antidepressants were shown to be superior in 59% of the trials using inert placebos versus only 14% of those using active placebos. However, Quitkin et al. [9] calculated the percentage of *responders* from these studies. They reported that the response rate to active placebo was about 30% which is comparable to the response consistently demonstrated in the literature using inert placebos. Furthermore, Quitkin et al. [9] argue that the studies using active placebos possessed serious methodological weaknesses that preclude definitive conclusions, including inadequate sample sizes, medication dosing, treatment duration, and diagnostic specificity.

How can two reviews of the same studies come to such completely different conclusions? Some of the confusion may be understood when examining the methodology used to examine the literature. Thomson [28] and Quitkin et al. [9] used different empirical approaches as compared with Moncrieff et al. [29, 30]. For example, Thomson [28] only examined whether or not the active placebo condition was significantly different from the active drug condition in each study. Quitkin et al. [9] examined a different categorical variable – responder status. It is well known that categorization procedures provide clearer differentiation between groups than examination of continuous variables [36]. Interestingly, Tedlow et al. [37] found that the relationship between baseline severity and outcome varied as a function of the type of measurement used (i.e., percent change vs. categorical response rate). However, meta-analysis also has its inherent flaws as a

data-analytic procedure [24], and two systematic reviews can produce different results, depending on the methods used [38]. For example, the results of a meta-analysis are only as good as the studies that are included. Also, meta-analyses that are overly inclusive tend to obscure differences found in individual studies because of increased variance.

Quitkin et al. [9] do not specifically critique the methodology used by Moncrieff et al. [29], but do suggest that their conclusions are debatable because of the methodological flaws contained in the studies reviewed. In their 2002 meta-analysis, Moncrieff et al. [30] examined the methodological qualities of the studies under review. They found that in one study the raters assessed patients whom they guessed to be on the active drug as more improved and that in another study patients on antidepressants reported more side effects than those on the active placebo which could have resulted in the unblinding of raters. Therefore, it also appears likely that any methodological flaws could have produced biases against the active placebo condition. Most importantly, Moncrieff et al. [30] found that the study quality was inversely correlated with effect size ( $r = -0.78$ , with one outlying study excluded). In other words, studies with higher methodological rigor showed smaller differences between active placebos and antidepressants. This is consistent with the findings of the meta-analysis performed by Greenberg et al. [8], examining studies using inert placebos.

Furthermore, no compelling evidence exists that pill placebos are associated with any identifiable risk of harm to participants. For example, the use of inert placebos does not appear to lead to a higher suicide risk in antidepressant trials [39], and recent data actually suggest the opposite trend – higher rates of suicidality observed with SSRIs relative to placebos [40]. However, active placebos also have been criticized on ethical grounds. Because the drugs mimic common side effects, they not only provide no intentional therapeutic benefit, but also produce undesirable side effects. Therefore, although conceptually and methodologically appealing, studies designed to use active placebos would require more stringent safeguards and procedures for protecting participants. These types of ethical dilemmas may best be examined within a cost-versus-benefit context. Active placebos might well be a necessary precaution, at least in some trials to fully examine the efficacy of antidepressants which because of their widespread use constitute an important public health issue.

Finally, perhaps the biggest problem with incorporating active placebos in antidepressant trials is a purely practical one. The active placebos used in previous re-

search were mainly in trials examining TCAs. Drugs that mimic the problems associated with other popular classes of medication, such as SSRIs, need further development. Other possibilities for active placebos for SSRIs are peripherally acting anticholinergic and antihistamine medications that would produce side effects without the possibility of clinical improvement [33]. Some have suggested that active placebos are not needed, as other psychotropic drugs (e.g., neuroleptics, barbiturates, and benzodiazepines) have been compared with antidepressants for depression. Results have been fairly mixed, although some studies using neuroleptics or benzodiazepines have shown either equivalency or superiority to antidepressants [36, 41]. These investigations assured that patients would experience side effects. However, drugs such as neuroleptics and benzodiazepines already have demonstrated psychoactive properties with established efficacy for treating other conditions, making them methodologically undesirable for use as active ‘placebos’.

### Improving Antidepressant Trials

Although others [42–44] have made recommendations for improving antidepressant trials recently, the following suggestions focus specifically on how better to separate specific from nonspecific treatment effects and to reduce allegiance and expectancy effects. Separating the nonspecific from the specific active ingredients of effective psychiatric treatments traditionally has been the domain of psychotherapy outcome research. Psychotherapy studies also traditionally have had to contend with strong allegiance effects (i.e., researchers’ biases affecting the results from clinical trials) [45] and patient expectancies skewing results. Therefore, researchers who are involved in antidepressant medication trials may benefit from examining this literature on empirically supported psychosocial treatments [see 46 for a recent discussion in psychotherapy]. Within this context, the recommendations shown in table 1 were developed.

#### *Controlling for Nonspecific Factors*

Little attention is given in psychopharmacology trials as to how expectancy influences the outcome. Furthermore, unblinding may be confounded with treatment effect in studies showing drug superiority to inert placebos. Keeping raters blind to the study design and hypotheses offers an important degree of protection against such expectancy effects. Another option would be to conduct studies that manipulate patient/rater expectancy as an

**Table 1.** Potential strategies for separating specific from nonspecific effects in antidepressant trials

---

*Design issues*

Include active placebo arms in methodologically stringent trials  
Conduct dismantling studies based on underlying biological theories  
Manipulate patient/rater expectancy as an independent variable  
Increase the number of study arms within studies in large trials  
Include wait list control groups when possible  
Explore alternate experimental designs (e.g., 'sequential parallel-comparison design') [43]  
Conduct multisite trials to reduce allegiance biases  
Keep raters blind to study design *and* hypotheses when possible  
Provide full disclosure of financial or other competing interests in treatments under investigation  
Follow guidelines in the CONSORT statement [70] when designing and reporting trials

---

*Measurement issues*

Assess treatment outcome in a multi-modal fashion and examine convergence of measures  
Rate degree of blindness achieved in the study on standardized scales  
Ask patients and raters to guess treatment conditions, ask the reasons for their guesses, and obtain percentage confidence in guesses (at multiple time points)  
Administer 'reaction to treatment' questionnaires to assess expectancy effects (at multiple time points)

---

*Statistical issues*

Analyze differences between placebo responders and nonresponders based on side effect profiles  
Contrast early fleeting, early maintained, and late-onset improvement patterns  
Examine the disparity between patient and clinician ratings to identify potential confounds  
Contrast results from completer-only and intention-to-treat analyses  
Examine the clinical significance of results based on the methods of Jacobson and Truax [65]  
Contrast continuous and categorical measures of improvement

---

independent variable. Interestingly, Benedetti et al. [47] manipulated patients' knowledge of whether or not they were receiving treatment and found that hidden administration was less effective than open administration of psychopharmacological agents. Another approach would be to conduct a 2 (drug vs. placebo) by 2 (believe they are receiving drug vs. believe they are not receiving drug) factorial design that manipulates participant expectation [10]. In this design, which has been used in the study of alcohol intoxication, participants (1) are told they will receive the drug and do receive the drug; (2) are told they will not receive the drug but do receive the drug; (3) are told they will receive the drug but do not receive the drug,

and (4) are told they will not receive the drug and do not receive the drug. Such a design is experimentally elegant, but may be difficult to implement in clinical trials due to ethical concerns.

An alternative to the above design that would be less ethically problematic would involve the use of active placebos. Active placebos could help control for expectancy by keeping side effects constant between groups. The analysis of appropriate comparison conditions in behavior therapy by Lohr et al. [6] may be a useful model here. These authors assert that the use of simple comparison conditions in psychosocial treatment research, such as wait list controls, is necessary but not sufficient to control for the strong influencing effects of nonspecific and common treatment factors. They recommend the inclusion of more stringent comparison conditions that can separate the incidental and characteristic elements of an investigational treatment. Furthermore, they assert that this type of investigation should be guided by an examination of the underlying theory of psychopathology on which the treatment is based.

Applied to pharmacotherapy, biological theories posit that the pharmacological characteristics of the investigational agents are responsible for improvement [48, 49]. A strong test of antidepressant medications for depression would include comparison conditions that mimic all the theoretically important elements of pharmacotherapy (e.g., expectation for improvement, doctor involvement and contact, effort justification, credible treatment rationale, etc.) minus the actual pharmacological agents posited to treat depression. Such designs in psychotherapy are commonly referred to as dismantling or component analysis studies, because they attempt to isolate the 'active ingredient' of treatments [50]. To date, it appears that these issues have been better examined in psychosocial treatment efficacy trials in contrast to pharmacotherapy research [see 51 for a recent example with generalized anxiety disorder]. However, as the above discussion of antidepressant trials demonstrates, the thorough examination of and experimental control for nonspecific and common treatment factors must be an equally important focus of research on drug therapies.

Furthermore, Lohr et al. [6] propose that additional treatment arms should be included *within* studies, because comparisons of conditions between studies are frequently unreliable. Along this vein, Kirsch and Sapirstein [10] argue for the inclusion of the following conditions in antidepressant trials: active placebos, inert placebos, active medication, and wait list controls for natural history. However, before active placebos can be incorporated into

modern antidepressant trials, work will need to be done to find appropriate drugs for this use, especially in trials with SSRIs.

#### *Measuring Bias Systematically and Statistically*

All studies are imperfect, and no design, either ideally or practically, can reduce all sources of bias and error. Conflict of interest is a pervasive problem in psychopharmacology [52]. For example, Kjærgard and Als-Nielsen [53] found that clinical trials favored the investigational treatment, if the researchers declared financial competing interests. The inclusion of active placebos or other experimental controls may help to reduce unblinding and expectancy effects, but other steps also may be necessary to reduce potential confounds. Specifically, making better attempts to measure and analyze bias is essential to interpreting results accurately and assuring generalizability. Common biasing effects include those of unblinding, allegiance, and expectancy.

Kirsch and Lynn [54] outlined the potential role of expectancy effects in psychotherapy outcome. To date, too few psychopharmacology studies have systematically examined and reported on these factors. However, more researchers are beginning to address these concerns in the literature. For example, 'reaction to treatment' questionnaires can be incorporated into antidepressant trials. These measures are increasingly being used in psychotherapy research to assess patients' expectancies about treatment and the believability of the treatment rationale [55]. Treatment credibility questionnaires are given typically early on in treatment, making them unlikely to be confounded with treatment effect.

Asking raters and patients to identify who is in which group, or which group will likely show greater improvement, can provide simple yet indispensable information when interpreting results [56]. Furthermore, asking participants how they came to their conclusions could prove useful. For example, do they believe they are in the medication group because of side effects experienced or observed improvement? Blindness assessments taken at multiple points that indicate percentage of confidence may provide added protection [25]. To aid in measurement, Even et al. [57] devised the blindness assessment and protection checklist, a seven-item checklist for evaluating the blindness protection of antidepressant trials. Analyzing the relationship between these variables and treatment outcome can help investigators examine the effectiveness of the controls used in the study.

Newer statistical techniques also may prove useful. Petkova et al. [58] developed a method to quantify rater

bias in antidepressant trials by contrasting patient and clinician ratings. Katz and Deveaugh-Geiss [59] recommended that the outcomes of subgroups of participants in the placebo condition be compared based on their side effect profiles. Ross et al. [60] suggested a two-step data analysis process that does not assume that placebo responders are a subset of medication responders. They developed a model to assess early fleeting, early maintained, and late-onset clinical improvement and suggested that this method can differentiate drug from placebo responding. However, because it is based on categorical analyses, it should be supplemented with information from continuous measures to ensure convergence of evidence. In addition, Stassen et al. [61] have adapted survival analysis techniques to investigate onset of improvement in antidepressant trials. They found that more participants dropped out due to lack of improvement (in contrast to side effects, etc.) in the placebo as compared with the antidepressant condition. Surprisingly, the time course of improvement for responders was identical regardless of treatment condition.

We recommend that researchers conduct and report both completers only and intention-to-treat (i.e., including all those who were assigned to a condition) analyses. High or differential dropout rates between groups may produce unrealistically positive results of overall benefits or may artificially inflate the superiority of one condition over another [see 62 for an example in psychotherapy research]. Kirsch et al. [15] found that the placebo response in completer-only analyses was significantly greater than that obtained in intention-to-treat analyses. Others have found the opposite relationship [36].

Finally, we suggest that antidepressant trials make better attempts to assess treatment outcomes in a multimodal fashion (e.g., self-report, clinician ratings, collateral information, behavioral ratings, and physiological indices) to examine convergence of data, as is fundamental to modern psychotherapy trials [63]. To date, multimodal approaches to assessment have most successfully been used in the study of anxiety disorders [64], and more thought needs to be given to the use of this strategy in studying depression. As discussed previously, outcome trials, especially those investigating pharmacological agents, tend to rely too heavily on simple interviewer ratings of outcome that can be biased, such as with the CGI. More valid and reliable improvement criteria, such as methods based on the work by Jacobson and Truax [65] related to 'clinical significance', deserve increased attention in pharmacological trials as well. Other commonly used instruments such as the HAMD include symptom

items that relate to changes in other nondepressive disorders or medical conditions that may skew results [36]. Some suggest that Bech's six-item subscale of the HAMD may be a more sensitive in determining changes in core depressive symptoms [66].

## Conclusions

It is important to note that the controversy over the specific efficacy of treatments for depression is not confined to pharmacotherapy. The 'mechanisms of action' debate has raged in the psychosocial intervention literature over some of the most common and successful interventions. Over a decade and a half after the publication of the seminal book by Beck et al. [67] on the use of cognitive therapy for depression, Jacobson et al. [68] conducted a dismantling study that implicated behavioral activation,

not cognitive techniques, as the active ingredient in the treatment. Modern psychotherapy research is based largely on the designs and theories initially developed for pharmacological trials. Perhaps it is time for researchers of pharmacological treatments to pay attention to the refinements in methodology that psychotherapy researchers have made to help disentangle specific from nonspecific treatment effects [see 69 for a comprehensive review]. Only future research using improved methodology will allow us to piece together the puzzle of the optimal treatment of depression.

## Acknowledgments

We would like to thank David Antonuccio, PhD, for his helpful feedback on an earlier draft of the manuscript. This paper was supported in part by Grant No. R01 MH052232 from the National Institute of Mental Health (NIMH), Bethesda, Md., USA.

## References

- 1 Kessler RC, Berglund P, Demler O, Jin R, Koretz D, Merikangas KR, Rush AJ, Walters EE, Wang PS: The epidemiology of major depressive disorder: Results from the National Comorbidity Survey Replication (NCS-R). *JAMA* 2003;289:3095–3105.
- 2 Antonuccio DO, Danton WG, DeNelsky GY, Greenberg RP, Gordon JS: Raising questions about antidepressants. *Psychother Psychosom* 1999;68:3–14.
- 3 Department of Health: Prescription Cost Analysis 1997. London, HMSO, 1998.
- 4 Gaudiano BA, Herbert JD: The antidepressant-placebo debate in the media: Balanced coverage or placebo hype? *Sci Rev Ment Health Pract* 2003;2:74–77.
- 5 Leber P: The use of placebo control groups in the assessment of psychiatric drugs: An historical context. *Biol Psychiatry* 2000;47:699–706.
- 6 Lohr JM, DeMaio C, McGlynn FD: Specific and nonspecific treatment factors in the experimental analysis of behavioral treatment efficacy. *Behav Modif* 2003;27:322–368.
- 7 White L, Tursky B, Schwartz GE: Placebo: Theory, Research, and Mechanisms. New York, Guilford, 1985.
- 8 Greenberg RP, Bornstein RF, Greenberg MD, Fisher S: A meta-analysis of antidepressant outcome under 'blinder' conditions. *J Consult Clin Psychol* 1992;60:664–669.
- 9 Quitkin FM, Rabkin JG, Gerald J, Davis JM, Klein DF: Validity of clinical trials of antidepressants. *Am J Psychiatry* 2000;157:327–337.
- 10 Kirsch I, Sapirstein G: Listening to Prozac but hearing placebo: A meta-analysis of antidepressant medication. *Prevention & Treatment* 1998;1:Article 0002a. Retrieved from: <http://journals.apa.org/prevention/volume1/pre0010002a.html>.
- 11 Klein DF: Listening to meta-analysis but hearing bias. *Prevention & Treatment* 1998;1:Article 0006c. Retrieved from: <http://journals.apa.org/prevention/volume1/pre0010006c.html>.
- 12 Klein DF: Reply to Kirsch's rejoinder regarding antidepressant meta-analysis. *Prevention & Treatment* 1998;1:Article 0008r. Retrieved from: <http://journals.apa.org/prevention/volume1/pre0010008r.html>.
- 13 Kirsch I: Reducing noise and hearing placebo more clearly. *Prevention & Treatment* 1998;1:Article 0007r. Retrieved from: <http://journals.apa.org/prevention/volume1/pre0010009r.html>.
- 14 Kirsch I: On the importance of reading carefully: A response to Klein. *Prevention & Treatment* 1998;1:Article 0009r. Retrieved from: <http://journals.apa.org/prevention/volume1/pre0010009r.html>.
- 15 Kirsch I, Moore TJ, Scoboria A, Nicholls SS: The emperor's new drugs: An analysis of antidepressant medication data submitted to the U.S. Food and Drug Administration. *Prevention & Treatment* 2002;5:Article 23. Retrieved from: <http://journals.apa.org/prevention/volume5/pre0050023a>.
- 16 Rosenthal R: The 'file drawer problem' and tolerance for null results. *Psychol Bull* 1979;86:638–641.
- 17 Thase ME: Small effects are not trivial from a public health perspective. *Psychiatric Times* 2003;19(9). Retrieved from: <http://www.psychiatristimes.com/p020909.html>.
- 18 Smith D, Dempster C, Glanville J, Freemantle N, Anderson I: Efficacy and tolerability of venlafaxine compared with selective serotonin reuptake inhibitors and other antidepressants: A meta-analysis. *Br J Psychiatry* 2002;180:396–404.
- 19 Hypericum Depression Trial Study Group: Effect of *Hypericum perforatum* (St John's wort) in major depressive disorder: A randomized controlled trial. *JAMA* 2002;287:1807–1814.
- 20 Walsh BT, Seidman SN, Sysko R, Gould M: Placebo response in studies of major depression: Variable, substantial, and growing. *JAMA* 2002;287:1840–1847.
- 21 Greenberg RP, Fisher S: Examining antidepressant effectiveness: Findings, ambiguities, and some vexing puzzles; in Fisher S (ed): *The Limits of Biological Treatment for Psychological Distress: Comparisons with Psychotherapy and Placebo*. Hillsdale, Erlbaum, 1989, pp 1–37.
- 22 Double DB: Unblinding in trials of the withdrawal of anticholinergic agents in patients maintained on neuroleptics. *J Nerv Ment Dis* 1995;183:599–602.
- 23 Bystritsky A, Waikar SV: Inert placebo versus active medication: Patient blindability in clinical pharmacological trials. *J Nerv Ment Dis* 1994;182:485–487.



- 24 White K, Kando J, Park T, Waternaux C, Brown WA: Side effects and the 'blindability' of clinical drug trials. *Am J Psychiatry* 1992; 149:1730-1731.
- 25 Piasecki MP, Antonuccio DO, Steinagel GM, Kohlenberg BS, Kafadar K: Penetrating the blind in a study of an SSRI. *J Behav Ther Exp Psychiatry* 2002;33:67-71.
- 26 Sharpe L, Ryan B, Allard S, Sensky T: Testing for the integrity of blinding in clinical trials: How valid are forced choice paradigms? *Psychother Psychosom* 2003;72:128-131.
- 27 Basoglu M, Marks I, Livanou M, Swinson R: Double-blindness procedures, rater blindness, and ratings of outcome: Observations from a controlled trial. *Arch Gen Psychiatry* 1997;54: 744-748.
- 28 Thomson R: Side effects and placebo amplification. *Br J Psychiatry* 1982;140:64-68.
- 29 Moncrieff J, Wessely S, Hardy R: Meta-analysis of trials comparing antidepressants with active placebos. *Br J Psychiatry* 1998;172:227-231.
- 30 Moncrieff J, Wessely S, Hardy R: Active placebos versus antidepressants for depression (Cochrane Review). *The Cochrane Library* 2002, issue 1. Oxford, Update Software.
- 31 Quitkin FM: Placebos, drug effects, and study design: A clinician's guide. *Am J Psychiatry* 1999;156:829-836.
- 32 Cooney JM: Active placebos in antidepressant trials. *Br J Psychiatry* 1998;172:541-542.
- 33 Salamone JD: A critique of recent studies on placebo effects of antidepressants: Importance of research on active placebos. *Psychopharmacology* 2000;152:1-6.
- 34 Klein DF: Flawed meta-analyses comparing psychotherapy with pharmacology. *Am J Psychiatry* 2000;157:1204-1211.
- 35 Otto MW, Nierenberg AA: Assay sensitivity, failed clinical trials, and the conduct of science. *Psychother Psychosom* 2002;71:241-243.
- 36 Moncrieff J: Are antidepressants overrated? A review of methodological problems in antidepressant trials. *J Nerv Ment Dis* 2001;198: 288-295.
- 37 Tedlow J, Fava M, Uebelacker J, Nierenberg AA, Alpert JE, Rosenbaum J: Outcome definitions and predictors in depression. *Psychother Psychosom* 1998;67:266-270.
- 38 Sensky T: The utility of systematic reviews: The case of psychological debriefing after trauma. *Psychother Psychosom* 2003;72:171-175.
- 39 Khan A, Warner HA, Brown WA: Symptom reduction and suicide risk in patients treated with placebo in antidepressant clinical trials: An analysis of the Food and Drug Administration database. *Arch Gen Psychiatry* 2000;57: 311-317.
- 40 Healy D: Lines of evidence on the risks of suicide with selective serotonin reuptake inhibitors. *Psychother Psychosom* 2003;72:71-79.
- 41 Schatzberg AF, Cole JO: Benzodiazepines in depressive disorders. *Arch Gen Psychiatry* 1978;35:1359-1365.
- 42 Klein DF, Thase ME, Endicott J, Adler L, Glick I, Kalali A, Leventer S, Mattes J, Ross P, Bystritsky A: Improving clinical trials: American Society of Clinical Psychopharmacology recommendations. *Arch Gen Psychiatry* 2002; 59:272-278.
- 43 Fava M, Evins AE, Dorer DJ, Shoenfeld DA: The problem of the placebo response in clinical trials for psychiatric disorders: Culprits, possible remedies, and a novel study design approach. *Psychother Psychosom* 2003;72:115-127.
- 44 TenHave TR, Coyne J, Salzer M, Katz I: Research to improve the quality of care for depression: Alternatives to the simple randomized clinical trial. *Gen Hosp Psychiatry* 2003; 25:115-123.
- 45 Luborsky L, Diguier L, Seligman DA, Rosenthal R, Krause ED, Johnson S, Halperin G, Bishop M, Berman JS, Schweizer E: The researcher's own therapy allegiances: A 'wild card' in comparisons of treatment efficacy. *Clin Psychol Sci Pract* 1999;6:95-106.
- 46 Herbert JD: The science and practice of empirically supported treatments. *Behav Modif* 2003;27:412-430.
- 47 Benedetti F, Maggi G, Lopiano L: Open versus hidden medical treatments: The patient's knowledge about a therapy affects the therapy outcome. *Prevention & Treatment* 2003;6:Article 1. Retrieved from: <http://journals.apa.org/prevention/volume6/pre0060001a.html>.
- 48 Miller JL, Delgado PL, Salomon RM, Berman R, Krystal JH, Heninger GR, Charney DS: Clinical and biochemical effects of catecholamine depletion on antidepressant-induced remission of depression. *Arch Gen Psychiatry* 1996;53:117-128.
- 49 Young LT, Bakish D, Beaulieu S: The neurobiology of treatment response to antidepressants and mood stabilizing medications. *J Psychiatry Neurosci* 2002;27:260-265.
- 50 Kazdin AE: *Research Design in Clinical Psychology*, ed 3. Boston, Allyn & Bacon, 1998.
- 51 Borkovec TD, Newman MG, Pincus AL, Lytle R: A component analysis of cognitive-behavioral therapy for generalized anxiety disorder and the role of interpersonal problems. *J Consult Clin Psychol* 2002;70:288-298.
- 52 Fava GA: Conflict in interest in psychopharmacology: Can Dr. Jekyll still control Mr. Hyde? *Psychother Psychosom* 2004;73:1-4.
- 53 Kjaergard LL, Als-Nielsen B: Association between competing interests and authors' conclusions: Epidemiological study of randomized clinical trials published in the BMJ. *BMJ* 2002; 325:249.
- 54 Kirsch I, Lynn SJ: Automaticity in clinical psychology. *Am Psychol* 1999;54:504-515.
- 55 Holt CS, Heimberg RG: The Reaction to Treatment Questionnaire: Measuring treatment credibility and outcome expectancies. *Behav Ther* 1990;13:213-214.
- 56 Antonuccio DO, Burns DD, Danton WG: Antidepressants: A triumph of marketing over science? *Prevention & Treatment* 2002;5:Article 25. Retrieved from: <http://journals.apa.org/prevention/volume5/pre0050025c.html>.
- 57 Even C, Siobud-Dorocant E, Dardennes RM: Critical approach to antidepressant trials: Blindness protection is necessary, feasible, and measurable. *Br J Psychiatry* 2000;177:47-51.
- 58 Petkova E, Quitkin FM, McGrath PJ, Stewart JW, Klein DF: A method to quantify rater bias in antidepressant trials. *Neuropsychopharmacology* 2000;22:559-565.
- 59 Katz RJ, Deveaugh-Geiss J: The issue of side effects and the 'blindability' of clinical drug trials. *Am J Psychiatry* 1994;150:1759-1760.
- 60 Ross DC, Quitkin FM, Klein DF: A typological model for estimation of drug and placebo effects in depression. *J Clin Psychopharmacol* 2002;22:414-418.
- 61 Stassen HH, Delini-Stula A, Angst J: Time course of improvement under antidepressant treatment: A survival-analytical approach. *Eur Neuropsychopharmacol* 1993;3:127-135.
- 62 Herbert JD, Rheingold AA, Gaudiano BA, Harwell VD: Standard versus extended cognitive behavior therapy for social anxiety disorder: A randomized-controlled trial. *Behav Cogn Psychother* 2004;32:131-147.
- 63 Haynes SN, O'Brien WH: *Principles and Practice of Behavioral Assessment*. New York, Kluwer/Plenum Press, 2000.
- 64 Herbert JD, Rheingold AA, Brandsma LL: Assessment of social anxiety and social phobia; in Hofmann SG, DiBartolo PM (eds): *From Social Anxiety to Social Phobia: Multiple Perspectives*. New York, Guilford, 2001, pp 20-45.
- 65 Jacobson NS, Truax P: Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol* 1991;59:12-19.
- 66 Möller HJ: Methodological aspects in the assessment of severity of depression by the Hamilton Depression Scale. *Eur Arch Psychiatry Clin Neurosci* 2001;251(Suppl 2):II13-II20.
- 67 Beck AT, Rush AJ, Shaw BF, Emery G: *Cognitive Therapy of Depression*. New York, Guilford, 1979.
- 68 Jacobson NS, Dobson KS, Truax PA, Addis ME, Koerner K, Gollan JK, Gortner E, Prince SE: A component analysis of cognitive-behavioral treatment for depression. *J Consult Clin Psychol* 1996;64:295-304.
- 69 Kendall PC, Holmbeck G, Verduin T: *Methodology, design, and evaluation in psychotherapy research*; in Lambert MJ (ed): *Bergin and Garfield's Handbook of Psychotherapy and Behavior Change*, ed 5. New York, Wiley, 2004, p 43.
- 70 Moher D, Schulz KF, Altman DG: The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. *Ann Intern Med* 2001;134:657-662.