

Placebo Response in Clinical Trials

What We Can Learn from Antidepressant Trials

Tony Womack, John Potthoff, and Carol Udell

Placebo response is one of the most challenging phenomena in antidepressant clinical research. Depressed subjects—who by definition often feel alone and unworthy—are likely to respond positively to the attention paid and empathy shown by a physician in a clinical setting. It is not unusual for two-thirds of subjects with depression to improve while taking active medication and one-third to improve while taking placebo,¹ indicating that about a third of subjects benefit from the specific effects of antidepressant medication and about a third improve because of nonspecific placebo effects. In short, of those who respond while taking medication, at least half do so for reasons other than direct pharmacological effects. Although this result may be acceptable in treatment settings, it presents a unique challenge to clinical trial design and analysis in which a clear difference between true drug response and placebo response must be demonstrated.

Placebo response, of course, is not limited to subjects suffering from depression. In fact, probably no therapeutic endeavor is completely immune to the effects of the placebo response, but there seems to be little in common between different types of responses—except for the fact that they exist. Benedetti et al. explored the pain-placebo response phenomenon and concluded that placebo analgesia is mediated by endogenous opiates,²⁻³ that side effects to placebo seem to be conditioned by previous exposure to active drug, and that the mechanisms underlying placebo analgesia and placebo respiratory depression are independent of each other.⁴ Subjects with acute headache in the emergency room respond

equally well to placebo, ketorolac, or meperidine.⁵ Even studies of conditions subject to “objective” measurement of improvement, such as hypertension, have shown positive placebo response rates of up to 30%.⁶

This continuing research brings us only a tiny step closer to truly understanding the placebo response, and in doing so often raises more questions than it answers. Why, for example, did 47.7% of subjects suffering from treatment-refractory reactive arthritis in one study respond to placebo treatment?⁷ Why did placebo profoundly reduce the pain of duodenal ulcer subjects in another study while not appearing to aid in the healing process?⁸

Who are the subjects who respond to placebo, and why doesn't every subject respond? Attempts at identifying placebo responders have produced diametrically opposed results. In one study, the identifying factor most associated with a placebo responder was a sense of self-sufficiency,⁹ whereas in another the factor most likely to predict a positive placebo response was the bleakness of the subject's home environment.¹⁰

Attempts to decrease the impact of the placebo response by identifying surrogate markers have shown some progress, but also raise the question of the ultimate goal of our research. If an arthritic subject feels better, is more mobile, and has a better quality of life following treatment with a medication, does it truly matter that the subject's erythrocyte sedimentation rate is unchanged? The same question may be asked of depressed subjects: If the subject is no longer depressed, does it matter clinically what mechanism mediated the improvement? The answer to this question, of course,

The lessons from depression studies can give trial designers a new perspective on ways to differentiate placebo effect from true drug response.

brings us back to the heart of our quandary as researchers. Our mandate when researching a potential new drug is, for better or for worse, to demonstrate a robust difference between the placebo response and the active drug response while knowing that a certain percentage of the responses of all subjects is due to the placebo effect.

Regardless of the therapeutic area under investigation, we believe that our recommendations for clinical trial design can be used to increase the likelihood of demonstrating a true difference between active drug and placebo.

Separating placebo from drug responses

Determining which part of an observed response is due to placebo and which is due to drug is a complex process in any clinical trial using placebos. Clinicians often take an “either/or” position, assuming that a positive response to treatment can be credited either to the pharmacological effects or to a placebo response. In fact, the placebo response quite often contributes to the overall positive response to a “true” medication by supporting the characteristics of the reference drug with which it is compared. Moreover, the placebo response is not directly related to the act of taking a “dummy” pill, but rather emanates from a variety of factors:¹¹

- An expectation of success—by the subject as well as by the investigator/evaluator
- Operant conditioning—that is, an improvement in condition following a visit to the doctor, brought about by a history of improvement following such visits
- Persuasion and suggestibility
- Effects of social influences inherent in receiving treatment and in the authority of the doctor
- Possible therapeutic effects of repeated assessments
- Response bias in which the subject overstates the severity of the initial condition and/or the status of the improvement.

An additional problem in identifying a true clinical response to an antidepressant in particular is that the nature of the disease itself makes it difficult to differentiate between true remission and placebo response. Clinical depression is not a tightly defined diagnosis and has many variables that make clinical trials of new drugs difficult:

- The intersubject and intrasubject variability of symptoms
- The potential clinical differences between a single episode and chronic or recurrent episodes
- The lack of stability of symptoms over time (cyclic presentation of the symptoms).

The design of an antidepressant trial that will demonstrate a robust difference between the active drug and placebo is thus complicated when the heterogeneity of the depressed population is combined with the complexity of differentiating between the different types and stages of depression.

A number of studies have been performed to determine how to separate placebo and true clinical responses to antidepressants. The results indicate that antidepressant placebo responses occur early and are of short duration, whereas true drug responses occur later and last longer. Results of a study by Quitkin suggest a strong correlation between abrupt, transient improvements during the initial period of a study and a placebo response.¹² Such a conclusion is logical because at such an early

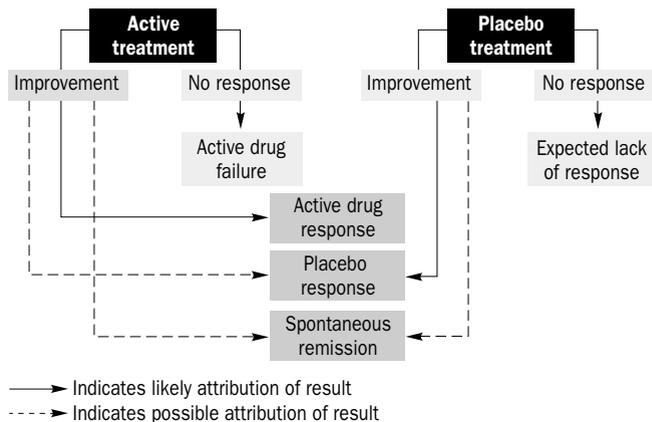


Figure 1. The possible attribution of a subject's response to treatment in a standard clinical trial.

stage of treatment, no “true” pharmaceutical effects are anticipated.¹³ Some qualifications of this finding have been found (see More on Early Positive Effects box), but the preponderance of evidence supports Quitkin's theory that early-onset, abrupt responses to either antidepressant or placebo treatment are most likely to be placebo responses that will prove to be non-persistent at some point in the future. Furthermore, some evidence suggests that even gradual responses during the first two weeks of treatment are likely to be due to spontaneous remission of disease rather than a true drug response. The currently accepted antidepressant clinical trial analysis plan would likely count both of these results as true drug effects if the subjects were receiving drugs. Our conclusion is that both attributions would be incorrect.

Addressing placebo effect in clinical trials

Reducing antidepressant placebo response through study design is extremely difficult, partly because depression is a complex and heterogeneous condition. The nature of the illness also lends itself to complicating attitudes on the part of patient and doctor. Although these attitudes play an important role in day-to-day treatment of depression, they inject significant “noise” into the analysis of a clinical trial in the form of either positive or negative placebo response. Unlike “silent” diseases such as hypertension or hypercholesterolemia, the depressed patient is usually very aware of having a problem. The attitude with which a patient approaches treatment can have both positive and negative implications on the outcome of that treatment. The patient's awareness of and attitude toward the problem must be considered part of the equation for finding a solution to the problem. Furthermore, in most cases, patients' beliefs about their condition and prognosis are readily influenced by their doctor's attitude.

Placebo washout period. Several attempts have been made to identify and limit this placebo response noise through clinical trial design. One common attempt has been to require a placebo washout period prior to randomization. During washout, all subjects receive placebo in a single-blind structure, in which the doctor is aware that the subject is receiving placebo but the subject is not. Subjects who show a predefined level of improvement within the washout period are not randomized into the study.

More on Early Positive Effects and the Placebo Response

Studies indicate that antidepressant placebo responses occur early and are of short duration, whereas true drug responses occur later and last longer.¹² These observations are not without problems, however, because they do not explain those cases in which the placebo-positive effects are so persistent as to last long after treatment cessation. In another study, as many as 44% of placebo responders who were excluded from a depression study because of their positive (that is, placebo) response during the placebo washout period (defined as the initial screening period for excluding those with responses to placebo alone) remained improved for up to six months.¹⁴ The authors attribute this result to a combination of placebo response followed by spontaneous remission. Such an additive response seen in a subject could erroneously be ascribed to a "true" pharmacological effect of the study drug in a typical clinical trial.¹⁴

Further investigating Quitkin's findings, another group of investigators followed for six months the cases of 73 subjects who had shown a successful antidepressant response to 12 weeks of active drug treatment.¹⁵ Those subjects remained on the active drug to which they had shown improvement and received a biweekly Clinical Global

Impression of Improvement rating during the follow-up. Among subjects who were initially deemed a treatment success in a standard 12-week clinical trial and then ultimately relapsed, 83% had shown a placebo response as defined by Quitkin (that is, abrupt, early improvement) during the initial 6 weeks of study. Attempts to eliminate other possible contributory factors support the conclusion that subjects who show early-onset responses to treatment with antidepressant drugs are likely experiencing a placebo response rather than a sustainable drug response.

In another study, investigators attempted to further characterize the "early" placebo response defined by Quitkin's group by looking at the chronicity—that is, chronic versus acute onset—of the disease.¹⁶ Their results showed a nonsignificant trend indicating that placebo response tended to be slightly lower among chronically ill subjects compared with those with short-duration disease.

Working with other investigators, Quitkin used his definitions of a "true drug response" (delayed and persistent) and a "placebo response" (early and nonpersistent) to investigate long-term changes in subjects who achieved a score of ≤ 7 on the Hamilton Rating Scale for Depression (HRSD) and were judged not to be clinically depressed

for the final 3 weeks of a 12–14 week treatment regimen with fluoxetine.¹⁷ Subjects who met these criteria were randomized either to placebo or continuing fluoxetine treatment and were followed for 50 more weeks. A return of depressed symptoms defined relapse. Subjects with an initial true drug response relapsed significantly more frequently if they were switched to placebo than if they continued to receive fluoxetine ($p < .001$ for weeks 12–26, $p < .005$ for weeks 26–50, and $p < .41$ for weeks 50–62). Subjects with an initial placebo response pattern had similar outcomes regardless of long-term treatment ($p < .20$ for weeks 12–26, test invalid for weeks 26–50, and $p < .67$ for weeks 50–62). Subjects with a placebo response pattern relapsed more often when they continued to receive fluoxetine than subjects with a true drug response ($p < .01$ for weeks 12–26, $p < .10$ for weeks 26–50, and $p < .36$ for weeks 50–62). (It should be noted that the population size causes a lower confidence level in this final comparison.) Apparently, some of what is considered relapse in subjects is actually a loss of placebo effect rather than loss of true drug effect. Only by following subjects for an extended period of time can the true efficacy of a drug be determined.

Although the washout period is certainly logical and intuitively compelling, the preponderance of evidence in the published literature refutes the effectiveness of such a design in reducing the overall placebo response seen in a clinical trial. An analysis of 10 years of literature (1983–1992) and two trials specifically designed to test this trial design all led to the conclusion that a single-blind placebo washout period produces no significant changes in the overall reduction in depression ratings for subjects in the placebo groups, no significant difference in the antidepressant effectiveness, and no significant difference in the number of dropouts in the two types of studies.^{18–19} The evidence is presented, however, with no attempt at an explanation as to the reasons that such results were found.

Subgroup analysis. Various methods of subgroup analysis have been undertaken to statistically separate drug effect from placebo response. One extensive analysis determined that women with a single episode of depression were the subjects most likely to show a placebo response, whereas women with recurrent disease were the least likely. The analysis showed also that placebo responders with their first depressed episode are likely to have a relatively low Hamilton Rating Scale for Depression (HRSD) score at baseline and will have a low rating of psychomotor retardation. Placebo responders with recurrent dis-

ease are likely to have a relatively low score on somatic anxiety at baseline.¹⁹

Therapeutic alliance. Although these characteristics are interesting and shed some light on the scope and intricacies of the overall problem, they offer little in the way of insight into clinical trial design. One area that does shed some light is referred to as *therapeutic alliance*.²⁰ Therapeutic alliance is part of a subject's situational treatment context and includes empathy, compassion, a helpful attitude, sympathetic listening, and making subjects feel that they are understood.

The shared doctor-subject goals, couched in this context, become the foundation upon which the treatment of a subject is built. When designing clinical trials, attention is generally paid to limiting the amount of psychotherapeutic treatment a psychopharmacological protocol may contain, but much less attention is usually paid to the effects of the average therapeutic alliance. It is unrealistic to try to eliminate this relationship. Even if it were possible to do so, a certain amount of therapeutic alliance seems necessary to assure compliance to the study regimen.

Subjects usually enter antidepressant studies with a significant personal history of interaction with both mental and medical health care professionals. Getting better following a visit to the doctor has become part of the expectations of these sub-

Differences between Centers

At the conclusion of two studies of an antidepressant that showed no efficacy by standard analysis, Niklson and associates completed a post hoc analysis to test the hypothesis that there is an important difference between treatment centers—that is, between private and primary care settings—that is not adequately controlled for in standard analysis.²⁴ The initial analysis of two clinical trials showed that neither

imipramine nor the new drug under investigation was significantly better than placebo at treating depression. By dividing the sites into two groups, however, one made up of sites that detected a difference between imipramine and placebo of at least two points on the HRSD scale following six weeks of treatment, the investigators found evidence that there were differences between the groups of centers

concerning important baseline characteristics of the subjects. They also found convincing evidence that different treatment strategies were being used by different centers, as demonstrated by a particularly large difference in the overall drop-out rate (24.9% in centers showing a 2-point difference versus 34.7% in those failing to do so).

jects. It is known that interpersonal therapy can be as efficacious as antidepressants,²¹ but less clearly understood is the point at which assessments and evaluations become as efficacious as interpersonal therapy. Depending on the requirements of the protocol, the initial assessment of a subject may take up to an hour or more. Follow-up visits may take nearly as long. The result is that a protocol that specifically prohibits interpersonal therapy during the trial, has, by virtue of its requirements, created a situation that very closely approximates the effect of such therapy. Subjects with a high degree of openness to new experiences, suggestibility, or a high need for social approval who are subjected to weekly or biweekly assessments—each of which involves spending a notable length of time with a clinician—can almost be expected to show a positive treatment response regardless of active drug or placebo treatment.

In addition, doctors are extremely adept at spotting behavioral clues that allow them to differentiate between subjects on active drug and those taking placebo.²² A doctor who believes that a particular subject is receiving placebo will pay very close attention to that subject in an attempt to head off any exacerbation of depression or other adverse events that might be brought about by placebo treatment. This extra attention could very likely increase the level of placebo response seen in that subject by alleviating his or her depressed mood—which has been shown to precede actual physiological improvement.²³

Type of treatment setting. Expanding the idea of the significance of doctor-subject alliance, it is interesting to compare treatment settings. Subjects seen in a more nurturing environment are likely to leave the visit feeling that they have gained something of value while in that environment. Such a feeling should likely lead to an improvement in the subjects' moods, and hence their conditions. In fact, a significant difference in placebo response has been observed in subjects receiving treatment in private psychiatric practices as opposed to those subjects receiving treatment in primary care settings.²⁰ Although stratification should prevent this sort of error, a small subject population combined with a preponderance of one type of site could result in spurious conclusions (see Differences between Centers box).

Response to adverse events. A final confounding factor in the therapeutic alliance is the subject's response to adverse events experienced during the study. It is commonly accepted that subjects will use the occurrence of adverse events as clues to whether they are on active medication or placebo. As antidepressants have progressed throughout the years, the side effect profile has improved; fewer and fewer subjects report

adverse events. As time passes in a clinical trial and subjects perceive no personal physiological changes, they may assume they are on placebo while they are in fact on active drug. Such a perception may have the effect of actually decreasing the effect of active drug, thereby decreasing the difference between the active drug response and the placebo response. On the other hand, placebo response can be conditioned.^{25–26} Subjects who believe they are experiencing an adverse event are very likely to perceive themselves as receiving the study drug. Should the subject in fact be receiving placebo, such a perception will again increase the measured placebo response and thereby decrease the overall measured true drug effect.

A practical approach to placebo response

The direct and indirect costs of the placebo effect in clinical trials of antidepressant medications may never be completely known. How many valuable drugs have been shelved due to an apparent lack of efficacy versus placebo? How many less or equally efficacious drugs are currently on the market because of the vagaries of the placebo effect? Some researchers have even gone so far as to suggest that as much as 75% of all antidepressant response to medication is attributable to the placebo effect, and some sponsors and investigators have suggested that the placebo response seen in clinical trials is actually increasing over the years.²⁷

Several practical recommendations could emerge from the evidence presented here. These recommendations are aimed at increasing the sensitivity of a study without compromising its specificity. For the greatest possible chance of achieving this goal, these recommendations should be practiced simultaneously throughout the planning and execution of the trial.

Assessing efficacy. Assessing clinical depression is a challenging task—differentiating between normal depressed mood, situational depression, dysthymia, and major depression is extremely difficult.²⁸ Most clinical trials of depression use the Hamilton Rating Scale for Depression (HRSD) and/or the Montgomery-Åsberg Depression Rating Scale (MADRS).^{29–30} Scores on the HRSD (also known as HAM-D) are most often used as the primary efficacy endpoint in clinical trials of antidepressants, though the scale may not be able to discriminate changes in depression among more mildly depressed subjects and thus may not be able to accurately assess efficacy in a clinical trial.³¹

Because clinical rating scales are used to measure the primary efficacy endpoints in most clinical trials of antidepres-

Using Depression Rating Scales to Measure Efficacy

Differentiating between normal depressed mood, situational depression, dysthymia, and major depression is extremely difficult.²⁸ There are many different depression rating scales, some of which are self-reporting and some clinician-administered. Symptoms assessed by depression scales are usually classified into three groups: affective (for example, sadness), cognitive (for example, suicidal thoughts), and somatic (for example, appetite loss). It has been our experience that self-report scales tend to be used when time and costs need to be kept low, but these scales lack the specificity of clinician-administered scales. Clinician-administered scales are able to capture nonverbal indicators of depression, and contributing environmental events, and are better able to evaluate the subjective severity of the depressive episode.

Most clinical trials of depression use the Hamilton Rating Scale for Depression (HRSD) and/or the Montgomery-Åsberg Depression Rating Scale (MADRS).^{29–30} The HRSD (also sometimes known as HAM-D) is most often used as the primary efficacy endpoint in clinical trials of antidepressants. An experienced rater administers the HRSD through a clinical interview. The HRSD was developed to assess the

severity of depression in previously diagnosed subjects and has high inter-rater reliability. The validity of the HRSD has been challenged because of its seeming weakness in assessing the cognitive and affective symptoms of depression. The scale tends to assess more thoroughly somatic symptoms, which are indicative of more severe depression. Thus, the scale may not be able to discriminate changes in depression among more mildly depressed subjects.³¹

The MADRS is similar to the HRSD in that it also is a clinician-administered scale that is used to assess the severity of depression among subjects diagnosed with depression. The MADRS assesses the full spectrum of depression except for motor retardation. The MADRS is often used as a secondary efficacy variable in clinical trials for antidepressants.³⁰

Because clinical rating scales are the primary efficacy endpoints in most clinical trials of antidepressants, it is extremely important that the scale be used in such a way as to maximize its sensitivity to treatment change over time. Using the HRSD as a primary efficacy endpoint can be problematic because of its limited ability to detect changes in symptom severity among mild to moderately depressed

subjects.³¹ Thus, treatment improvements in these subjects may not be accurately reflected in the data from a clinical trial.

Because there is much variability in the expression of symptoms among depressed subjects and scales are generally better at targeting a subset of symptoms, we suggest the use of multi-endpoint criteria—such as the HRSD or MADRS—to define clinical response. Further, recent work indicates that HRSD subscale scores are more sensitive in detecting differences in treatments (placebo versus active). Thus, it may be useful to include subscale hypotheses in future studies.³² In addition, because the symptoms of depression are highly variable over time, clinical trials should use more frequent rating assessments to better capture the cyclical nature of depression—for example, use average HRSD scores over a week to establish baseline and endpoint levels. As a possible secondary endpoint, trials could use the change in self-administered rating scales to supplement clinician-administered rating scales to further establish a true picture of the pattern of depression for each subject and to decrease the effect of the therapeutic alliance.

sants, it is extremely important to use the scale in such a way as to maximize its sensitivity to treatment change over time. Multi-endpoint criteria should be used with the scales to define clinical response because of the variability in the expression of symptoms among depressed subjects, and because scales are generally better at targeting a subset of symptoms. Subscale scores, found to be more sensitive in detecting differences in treatments (placebo versus active), should be included in future studies.³² Shorter, but more frequent, rating assessments can better capture the cyclical nature of depression. For example, use average HRSD scores over a week to establish baseline and endpoint levels. In addition, as a possible secondary endpoint, trials could use the change in self-administered rating scales to supplement clinician-administered rating scales. That could further establish a true picture of the pattern of depression for each subject and decrease the effect of the therapeutic alliance (see also *Using Depression Rating Scales to Measure Efficacy*).

It may be necessary to re-evaluate the use of rating scales in antidepressant trials. It is easy to agree that an HRSD change from 30 at baseline to 15 at endpoint is remarkable and, if commonplace, will contribute to an overall statistically significant improvement in the study population. But it is perhaps problematic to think of an individual subject as “cured” or “in remission” given that an HRSD score of 7 is generally considered at least

mildly depressed.²⁹ Can an analysis plan be devised to allow for more individual results? Or should new scales be attempted with remission being more directly related to the subject’s ability to function in the day-to-day world? The idea is intriguing, but it requires an in-depth review of its own.

Subject selection. As described earlier, depression is a very heterogeneous disease and is expressed in each individual by a collection of different symptom profiles.²⁸ For example, one depressed subject may present with decreased appetite and increased sleep, and another with increased appetite and decreased sleep. In light of this inherent symptom and subject variability, the study population should be as homogeneous as possible for an antidepressant trial to be successful. Careful attention must be given to appropriate selection of a study population taking into account population, nature of the disease, previous treatment, and the cyclical nature of depression.

Type of population (for example, elderly or young, women or men). Results obtained from a study involving subjects aged 70–90 might not be applicable to young depressed subjects. That is not only because the inherent components of the disease may be different, but also because of pharmacokinetic differences between the two groups. As many depression symptoms are somatic, their expression may be influenced by the age of the subject.

Nature of the disease (for example, chronic or single episode). As previously reported, single-episode disease seems more closely associated with placebo response than chronically manifested depression. To the extent that a homogeneous study population cannot be defined by the study protocol, stratification of the results for these variables should be performed in the statistical analysis.

Previous treatment. The response of subjects who have already participated in clinical trials that failed to elicit long-term improvement is unpredictable. A subject who has tried and failed to respond to new treatment may have an extremely wide range of responses, from a “negative” placebo response (“I’ve tried this before and it didn’t work”) to a positive placebo response (“This will certainly work this time”). Both responses are very sensitive to the evaluator’s input. And these subjects may have more chronic and/or advanced disease that could also introduce bias into the study result.

Cyclical nature of depression. One of the most difficult aspects of assessment of response to antidepressants is the occurrence

The doctor-patient relationship that is part of the treatment of any depressed patient is likely to be one of the contributing factors in the placebo response.

of “spontaneous” remission in subjects with major depression. Subjects often attain a near-normal symptom profile for periods of time while maintaining a diagnosis of recurrent major depression. Even more troubling are the day-to-day, or even hour-to-hour, variations in symptom expression and severity common among depressed subjects. These variations can completely obscure treatment differences in clinical trials. For example, a subject could be assessed for baseline on a particularly bad day, then be assessed at the endpoint on a particularly good day—and be labeled a responder when in fact no actual improvement has been achieved.

Study design proposal

Given the apparent failure of the single-blind placebo washout period at removing placebo responders from the clinical trial, we propose an alternative method. All subjects should be observed without changing their treatment for at least two weeks to ensure that subjects are stable and experiencing neither up- nor down-cycles of their conditions. Following this observation period, all stable and otherwise qualified subjects should then be randomized to either active drug or placebo treatment under double-blind conditions. The length of the double-blind phase of the study should be at least 8 weeks. All acute studies should be followed by an open-label extension of at least 12 weeks’ duration during which all eligible subjects receive active drug in order to fully see the effects of the active drug over time. The efficacy analysis should include only those subjects who showed no predefined improvement within the first 2 weeks of the trial, regardless of treatment received.

To control for the psychotherapeutic effect of repeated assessments, the amount of time an investigator is allowed to

spend with a subject during assessment and follow-up should be limited to 20 minutes per session if at all possible. Because the subjects’ beliefs about their treatment almost certainly play a role in the treatment outcome, the clinician’s response to subjects who say that they believe they are on placebo should be scripted to avoid any unconscious concurrence of the doctor who may also believe the subject is on placebo.

Sample size and statistical analysis. The sample size of a new clinical trial is crucial. Three factors determine the size of a placebo-controlled trial:

- The absolute antidepressant potential of the active compound being tested
- The size of the therapeutic response to placebo treatment
- The dropout rates and strategies used for treatment termination or continuation.

Assuming that efficacy has been defined as an absolute change of 3 points in the HRSD during treatment, about 100–140 subjects per treatment group will be necessary to show this difference by two-sided t-test with $p=.05$ and power $1-\beta=0.80$, given an average standard deviation of about 8 or 9 points.²⁴ A larger endpoint change will require a larger population. Even with the implementation of the suggestions made here, an adequate sample size will still be problematic to predict. Given our limited understanding of placebo response, an error on the side of caution may make the difference between a successful study and one that ends in disappointment.

Further thought and analysis should be given to the possibility of analyzing a study based on the average improvement shown by site. The idea holds possibility, but until further investigation is done, it will be risky to design a study around this premise.

Placebo washout period. Future study design should eliminate the placebo washout period by substituting an analysis plan that disregards early responders as defined by Quitkin.¹²

More effective trials

Clinical depression presents a unique challenge to clinical trial design and analysis because of the well-documented high placebo response rate and the inherent variability of symptom expression within and between depressed subjects. In an attempt to control the effects of placebo response on the trial’s outcome, most placebo-controlled trials have included little more than a single-blind washout period to eliminate placebo responders from the pool of subjects to be analyzed. This has clearly been a less than optimal approach to eliminating placebo responders from clinical trials. Although it seems that the placebo response cannot be completely eliminated from clinical trials, by using the suggestions included in this article, researchers may be able to minimize its effect and thereby increase the power of clinical trials in depression.

References

1. W.A. Brown, “Placebo as a Treatment for Depression,” *Neuropsychopharmacology*, 10, 265–269 (1994).
2. F. Benedetti, M. Amanzio, C. Casadio, A. Olario, and G. Maggi, “Blockade of Nocebo Hyperalgesia by Cholecystokinin Antagonist

- Proglumide," *Pain*, 71, 135–140 (1997).
3. F. Benedetti, M. Amanzio, S. Baldi, C. Casadio, and G. Maggi, "Inducing Placebo Respiratory Depressant Responses in Humans via Opioid Receptors," *European Journal of Neuroscience*, 11, 625–631 (1999).
 4. F. Benedetti, M. Amanzio, S. Baldi, C. Casadio, A. Cavallo, M. Mancuso, E. Ruffini, A. Oliaro, and G. Maggi, "The Specific Effects of Prior Opioid Exposure on Placebo Analgesia and Placebo Respiratory Depression," *Pain*, 75, 313–319 (1998).
 5. R.N. Harden, R.H. Gracely, T. Carter, and G. Warner, "The Placebo Effect in Acute Headache Management: Ketorolac, Meperidine, and Saline in the Emergency Department," *Headache*, 36, 352–356 (1996).
 6. R.A. Preston, B.J. Materson, D.J. Reda, and D.W. Williams, "Placebo-Associated Blood Pressure Response and Adverse Effects in the Treatment of Hypertension," *Archives of Internal Medicine*, 160, 1449–1454 (2000).
 7. D.O. Clegg, D.J. Reda, M.H. Weisman, J.J. Cush, F.B. Vasey, H.R. Schumacher, Jr., E. Budiman-Mak, D.J. Balestra, W.D. Blackburn, G.W. Cannon, R.D. Inman, F.P. Alepa, E. Mejias, M.R. Cohen, R. Makkena, M.L. Mahowald, J. Higashida, S.L. Silverman, N. Parhami, J. Buxbaum, C.M. Haakenson, R.H. Ward, B.J. Manaster, R.J. Anderson, W.G. Henderson, et al., "Comparison of Sulfasalazine and Placebo in the Treatment of Reactive Arthritis (Reiter's Syndrome)," *Arthritis and Rheumatism*, 39 (12) 2021–2027 (1996).
 8. M.J. Collen, M.R. Hanan, J.A. Maher, M. Rent, et al., "Cimetidine vs Placebo in Duodenal Ulcer Therapy," *Digestive Diseases and Sciences*, 25 (10) 744–749 (1980).
 9. C.G. Moertel, S.R. Taylor, A. Roth, and F.A.J. Tyce, "Who Responds to Sugar Pills?" *Mayo Clinic Proceedings*, 51, 96–100 (1976).
 10. L.E. Sanchez, J.L. Armenteros, A.M. Small, M. Campbell, and P.B. Adams, "Placebo Response in Aggressive Children with Conduct Disorder," *Psychopharmacology Bulletin*, 30 (2) 209–213 (1994).
 11. A.K. Shapiro and L. Morris, "The Placebo Effect in Medical and Psychological Therapies," in *Handbook of Psychotherapy and Behavior Change*, 2nd ed., S.L. Garfield and A.E. Bergin, Eds. (John Wiley & Sons, New York, 1978), pp. 369–410.
 12. F.M. Quitkin, P.J. McGrath, J.G. Rabkin, J.W. Stewart, W. Harrison, D.C. Ross, E. Tricamo, J. Fleiss, J. Markowitz, and D.F. Klein, "Different Types of Placebo Response in Subjects Receiving Antidepressants," *American Journal of Psychiatry*, 148, 197–203 (1991).
 13. J.M. Gorman, *The Essential Guide to Psychiatric Drugs*, 3rd ed. (St. Martins Mass Market Paper, New York, 1997).
 14. J.G. Rabkin, P.F. McGrath, J.W. Stewart, W. Harrison, J.S. Markowitz, and F. Quitkin, "Follow Up of Subjects Who Improved during Placebo Wash Out," *Journal of Clinical Psychopharmacology*, 6, 274–278 (1986).
 15. W. Harrison, J.W. Stewart, P.J. McGrath, E. Tricamo, and F.M. Quitkin, "Is Loss of Antidepressant Effect during Continuation Therapy Related to Placebo Effect?" *Psychopharmacology Bulletin*, 24, 9–13 (1988).
 16. A. Khan, S.R. Dager, S. Cohen, D.H. Avery, B. Scherzo, and D.L. Dunner, "Chronicity of Depressive Episode in Relation to Antidepressant-Placebo Response," *Neuropsychopharmacology*, 4, 125–130 (1991).
 17. J.W. Stewart, F.M. Quitkin, P.J. McGrath, J. Amsterdam, M. Fava, J. Fawcett, F. Reimherr, J. Rosenbaum, C. Beasley, and P. Roback, "Use of Pattern Analysis to Predict Differential Relapse of Remitted Patients with Major Depression during 1 Year of Treatment with Fluoxetine or Placebo," *Archives of General Psychiatry*, 55, 334–343 (1998).
 18. R.P. Greenberg, S. Fisher, and J.A. Riter, "Placebo Washout Is Not a Meaningful Part of Antidepressant Drug Trials," *Perceptual and Motor Skills*, 81, 688–690 (1995).
 19. R.J. Bialik, A.V. Ravindran, D. Bakish, and Y.D. Lapierre, "A Comparison of Placebo Responders and Nonresponders in Subgroups of Depressive Disorder," *Journal of Psychiatry and Neuroscience*, 20, 265–270 (1995).
 20. E. Schweizer and K. Rickels, "Placebo Response in Generalized Anxiety: Its Effect on the Outcome of Clinical Trials," *Journal of Clinical Psychiatry*, 58 (supplement 11) 30–38 (1997).
 21. J. Fawcett and R.L. Barkin, "Efficacy Issues with Antidepressants," *Journal of Clinical Psychiatry*, 58 (supplement 6) 32–39 (1997).
 22. S. Lurie, "Placebo Effects and Antidepressant Medication: Implications for Research and Clinical Care," *Bulletin of the Menninger Clinic*, 60 (1) 94–101 (1996).
 23. M. Versiani, M. Amin, and G. Chouinard, "Double-Blind, Placebo-Controlled Study with Reboxetine in Inpatients with Severe Major Depressive Disorder," *Journal of Clinical Psychopharmacology*, 20, 28–34 (2000).
 24. I.A. Niklson, P.E. Reimitz, and C. Sennel, "Factors that Influence the Outcome of Placebo-Controlled Antidepressant Clinical Trials," *Psychopharmacology Bulletin*, 33 (1) 41–51 (1997).
 25. N.J. Voudouris, C.L. Peck, and G. Coleman, "Conditioned Response Models of Placebo Phenomena: Further Support," *Pain*, 38, 109–116 (1989).
 26. N.J. Voudouris, C.L. Peck, and G. Coleman, "The Role of Conditioning and Verbal Expectancy in the Placebo Response," *Pain*, 43, 121–128 (1990).
 27. I. Kirsch and G. Sapirstein, "Listening to Prozac but Hearing Placebo: A Meta-Analysis of Antidepressant Medication," *Prevention and Treatment*, 1, article 0002a, 26 June 1998; available online at <http://journals.apa.org/prevention/volume1/pre0010002a.html>.
 28. American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed. (APA, Washington, DC, 1994).
 29. M. Hamilton, "A Rating Scale for Depression," *Journal of Neurology, Neurosurgery and Psychiatry*, 23, 56–62 (1960).
 30. S.A. Montgomery and M. Åsberg, "A New Depression Scale Designed to Be Sensitive to Change," *British Journal of Psychiatry*, 134, 382–389 (1979).
 31. I. McDowell and C. Newell, *Measuring Health: A Guide to Rating Scales and Questionnaires*, 2nd ed. (Oxford University Press, New York, 1996), pp. 238–286.
 32. D. Faries, J. Heilgenstein, D.J. De Brota, and W.Z. Potter, "The Double Blind Variable Placebo Lead-In Period: Results from Two Antidepressant Studies," poster presented at New Clinical Drug Evaluation Unit (NCDEU) meeting, Boca Raton, FL, 1–4 June 1999.

Tony Womack is a medical writer for PPD Development, Austin, TX. **John Potthoff, PhD,*** is president and CEO of *Tanistry, Inc.*, 2499 Capital of Texas Highway S., Building A, Suite 201, Austin, TX 78746, (512) 327-7333, fax (512) 327-6996, e-mail: john.potthoff@tanistry.com. **Carol Udell** is a biostatistician at PPD Development.

*To whom correspondence should be addressed.



INC Research, a full service contract research organization, manages all aspects of clinical drug development programs for pharmaceutical and biotechnology companies developing drug therapies and devices used for disorders of the central nervous system. Its therapeutic expertise in neurology, psychiatry and pain management, and specialized experience in providing clinical data services across all therapeutic areas, offers a unique resource to customers. INC Research has offices in San Diego, California; Bridgewater, New Jersey; Raleigh, North Carolina; Austin, Texas; and Charlottesville, Virginia. For more information visit www.incresearch.com.



John G. Potthoff, PhD

Vice President, CNS Therapeutics

Dr. Potthoff currently leads the CNS Therapeutics business unit for INC Research. Prior to joining INC Research, Dr. Potthoff was President & CEO of Tanistry, Inc., a niche-CRO focused on CNS projects, which was acquired by INC Research in February 2001. Prior to founding Tanistry, Dr. Potthoff held several positions with PPD Development from 1993–2000, most recently as Executive Director of Clinical Operations. Dr. Potthoff is a graduate of the The University of Texas with a PhD in Psychology.

INC Research

Corporate Office: 6131 Falls of Neuse Road, Suite 200

Raleigh, NC 27609

Phone: 919-876-9300 • Fax: 919-876-9360