The Nature of Placebo Response in Clinical Studies of Major Depressive Disorder

George I. Papakostas, MD; Søren D. Østergaard, MD; and Nadia Iovieno, MD, PhD

ABSTRACT

Objective: To review factors influencing placebo response and clinical trial outcome in depression, and suggest ways to optimize trial success in mood disorders.

Data Sources: PubMed searches were conducted by cross-referencing the terms *depression*, *depressive* with *placebo*, *clinical trial*, and *clinical trials* for studies published in English between 1970 and September 2013.

Study Selection: Relevant abstracts were identified in PubMed, including clinical trials, quantitative studies, and qualitative research. We obtained and reviewed relevant articles and utilized their information to synthesize the present review.

Data Extraction: Included articles were grouped in the following areas of relevance: (1) biological validity of illness, (2) baseline severity of illness, (3) chronicity of the index episode of depression, (4) age of participants, (5) medical and psychiatric comorbidity, (6) probability of receiving placebo, (7) use of prospective treatment phases (lead-in) (8) dosing schedule, (9) trial duration, (10) frequency of follow-up assessments, and (11) study outcome measure.

Results: Several key elements emerge as critical to the ultimate success of a clinical trial, including the probability of receiving placebo, study duration, dosing schedule, visit frequency, the use of blinded lead-in phases, the use of centralized raters, illness severity and duration, and comorbid anxiety.

Conclusions: Our increasing understanding of the placebo response in clinical trials of major depressive disorder lends to a, gradually, more predictable phenomenon and, hopefully, to one that becomes lesser in magnitude and variability. Several elements have emerged that seem to play a critical role in trial success, gradually reshaping the design of clinical, translational, as well as mechanistic studies in depression.

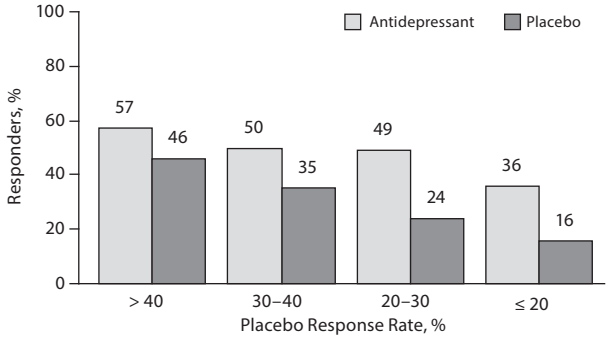
J Clin Psychiatry 2015;76(4):456–466

Submitted: *June 8, 2014; accepted November 7, 2014.*
Online ahead of print: *February 3, 2015 (doi:10.4088/JCP.14r09297).*
Corresponding author: *George I. Papakostas, MD, Massachusetts General Hospital, Clinical Trials Network and Institute, Harvard Medical School, One Bowdoin Sq, 6th Floor, Boston, MA 02114 (gpapakostas@partners.org).*

Major depressive disorder (MDD) is a highly prevalent medical illness,[1] often associated with significant morbidity, mortality, and functional impairment for those affected, resulting in profound economic burden worldwide.[2] Major depressive disorder is also a treatable illness, with pharmacologic agents along with various forms of psychotherapy representing the cornerstone of treatment as recommended by organizations such as the American Psychiatric Association,[3] the Canadian Network for Mood and Anxiety Treatments,[4] and the UK National Institute for Health and Clinical Excellence.[5] However, for many patients with MDD, treatments delivered do not always have the desired effect. Therefore, it remains crucial for the field to aid in the development of new and more efficacious antidepressants.

The double-blind, randomized, placebo-controlled trial (RCT) remains the gold-standard test of efficacy for any new pharmacotherapy that is being developed for MDD. In this type of study, the effect of an investigational medication is compared to that of placebo—a pill containing ingredients with no known specific antidepressant activity—for patients identified according to several a priori criteria. The typical design of an RCT involves the parallel comparison of 1 or more potential treatments with placebo, with sample sizes considered adequate to detect a therapeutic signal, given the expected placebo response rates in the specific population. In this setup, the *placebo response* is defined as the degree of reduction in depressive symptoms demonstrated among patients randomly assigned to receive a placebo pill as opposed to an experimental agent or established agent. Thus, a drug is not considered to be efficacious in the treatment of MDD unless it has been shown to "beat" placebo, in other words, to demonstrate a clinically significant and statistically greater reduction in depressive symptoms during the course of the study in at least 2 (according to the US Food and Drug Administration) clinical trials. Conversely, a pharmacologic agent may not beat placebo for 1 of 2 broad reasons: either it is truly not effective in treating the illness (at least at doses delivered and in the population selected) (negative study) or the study could not demonstrate whether a treatment effect exists or not (failed or uninformative study). The latter situation is particularly problematic, since it can deprive patients of a potentially useful treatment despite significant investments made aimed at demonstrating efficacy, or it can serve to continue investment in a futile way. Unfortunately, over the past 30 years, failed or uninformative RCTs in MDD have become the rule more than the exception.[6] The presence of a sizeable and difficult-to-predict placebo response rate seems to be a key culprit in relation to the many uninformative trials in MDD. Indeed, in a landmark study-level meta-analysis by Walsh et al[7] assessing 75 double-blind trials conducted among patients with MDD, it was shown that the response to placebo was highly variable and often of a substantial size (average 30%, range 12%–52%). This problem becomes even more serious when considering that quite a few negative or failed trials remain unpublished. As a result of the sizeable, highly variable placebo response rate in MDD, a significantly larger number of patients need to be randomized in order to demonstrate efficacy, while the variability and unpredictability of the placebo response rate makes it difficult to make

Figure 1. Placebo Response Rates and Clinical Trial Outcome[a]



[a]Data from Iovieno and Papakostas.[8]

accurate calculations regarding power and effect size when designing studies. Recently, in a study-level meta-analysis, Iovieno and Papakostas[8] demonstrated that, as placebo response rates increase, the difference in efficacy between antidepressants and placebo does not remain constant as one would expect, but progressively shrinks, which would suggest that placebo response rates in antidepressant clinical trials also represent a proxy of study quality, with higher placebo response rates associated with higher chances of having a failed study (Figure 1).

As a result, it becomes clear that identifying factors that influence the placebo response as well as the relative likelihood of responding to antidepressants versus placebo becomes of paramount importance as a challenge to pharmacologic drug development. Fortunately, a growing number of studies investigating such factors have been conducted within recent years. The aim of this article is to give a focused review of this literature and to suggest potential ways to optimize clinical outcome in future placebo-controlled RCTs in MDD.

DATA SOURCES AND STUDY SELECTION

PubMed searches were conducted by cross-referencing the terms *depression, depressive* with *placebo, clinical trial,* and *clinical trials*. Studies searched were published in English between 1970 and September 2013. Reference lists of relevant articles were also searched for the inclusion of works not identified by PubMed. The authors obtained and reviewed all relevant articles and utilized their information to synthesize the present review.

DATA EXTRACTION

Numerous relevant articles were identified. Our review will examine the results of these articles grouped in the following areas of relevance to placebo response and study design in MDD: (1) biological validity of illness, (2) baseline severity of illness, (3) chronicity of the index episode of depression, (4) age of participants, (5) medical and psychiatric comorbidity, (6) probability of receiving placebo, (7) use of prospective treatment phases (lead-in) (8) dosing schedule, (9) trial duration, (10) frequency of follow-up assessments, and (11) study outcome measure.

RESULTS

Subject-Related Factors

Biological validity of illness. The presentation of MDD in its current definition is, clinically, highly heterogeneous. For instance, Ostergaard et al[9] have shown that *DSM* symptoms of MDD can be combined in 1,497 different ways in order to constitute a depressive episode. Similarly, the same authors have also shown that 2 patients may fulfill the diagnostic criteria for MDD without sharing a single symptom.[9] Findings like these have led many experts in the field to argue that, while MDD may be associated with some diagnostic reliability, from a biologic standpoint the disorder is probably highly multifactorial, with no final common biological pathway leading to disease in all patients.[10] Furthermore, the relatively low heritability of MDD (37%) compared to other psychiatric syndromes such as alcohol dependence (56%), anorexia nervosa (60%), schizophrenia (81%), and bipolar disorder (85%)[11] further raises the possibility that some subtypes of MDD may be much more environmentally than biologically driven. Desseilles et al[12] have proposed this dichotomy may have potential consequences with respect to the relative likelihood of a patient responding to drug versus placebo in an RCT and, therefore, have proposed criteria aimed at enhancing signal detection in antidepressant RCTs by enrolling patients with a more biologically driven illness. Critical aspects of defining such patients according to their criteria include that, in order for inclusion, symptoms should be established as *pervasive*, *persistent*, and *pathological*. Thus far, at least 2 published clinical trials[13,14] have employed such inclusion/exclusion criteria as part of their methodology. Alternatively, efforts are also made in order to determine whether biomarkers covering various different biologic processes associated with the pathophysiology of depression in the literature (thereby lessening the need of conceptually identifying a single, final common disease pathway) can identify MDD from nondepressed control subjects with adequate specificity and sensitivity. In 1 such study,[15] a multianalyte panel demonstrated greater than 90% sensitivity and greater than 80% specificity in discriminating between these 2 groups, although potential implications of test score on treatment outcome in antidepressant RCTs has yet to be examined. A more interesting approach would be to examine the impact of combining tailored interviews such as those described by Desseilles et al[12] along with biomarker tests such as the one tested by Papakostas et al[15] on RCT outcome. Alternatively, a number of treatments with novel, nonmonoaminergic mechanisms of action have recently been studied,[16–19] offering novel opportunities for the conduct of biomarker analyses aimed at identifying subpopulations particularly responsive to treatment with these agents (moderators). Investigations are currently underway examining the potential role of various markers as moderators of drug-placebo outcome with these agents. Ultimately, if successful, such approaches may yield the basis for a tailored design of

more intelligent clinical trials in MDD and, in turn, for the growth of personalized medicine in mood disorders.

Baseline severity of illness. The observation that patient as well as clinical trial outcomes can vary as a function of illness severity has been repeatedly made throughout the course of antidepressant RCT history.[20,21] What is important to keep in mind when reviewing this literature is that the degree or probability of improvement as a function of illness severity varies according to whether improvement is defined a continuous measure (whereby patients with more severe symptoms can demonstrate, numerically, a greater reduction in scores than patients with milder symptoms) or a dichotomous one such as response or remission (whereby patients with milder symptoms require smaller degrees of symptom reduction in order to be considered improved). What is of interest for the present review, however, is how severity at baseline impacts the relative efficacy of antidepressants versus placebo—however that is defined. For this reason, we chose to cite exactly how outcome was defined in each analysis.

In the past decade, Khan et al[22–26] have conducted a number of analyses investigating the significance of baseline severity of MDD on outcome defined in a variety of ways. The first, a study-level meta-analysis[22] (58 trials) demonstrated a statistically significant positive correlation between the mean total baseline 17-item Hamilton Depression Rating Scale ($HDRS_{17}$) score and the change in mean $HDRS_{17}$ score during treatment in the antidepressant arm, and a statistically significant negative correlation was demonstrated in the placebo arm. A positive correlation between depression severity at baseline and change in symptom scores during the trial in the antidepressant-treated group and not in the placebo group was also confirmed in a subsequent patient-level analysis by Khan et al[24] based on patients enrolled in 1 of 15 different RCTs at that site. In a more recent, patient-level meta-analysis[26] based on patients enrolled in 1 of 16 RCTs at 1 site, the baseline severity according to the $HDRS_{17}$ was again significantly associated with a greater reduction in $HDRS_{17}$ in the antidepressant group but not in the placebo-treated group. In parallel, in a separate study-level report,[23] the authors had also demonstrated that baseline $HDRS_{17}$ score was significantly higher in trials where the antidepressant-placebo difference in reduction in symptom scores during treatment was ≥ 3.07 points on the $HDRS_{17}$ (the median difference) in favor of antidepressants compared to trials where the difference was < 3.07. Finally, in an analysis based on study-level data from 51 RCTs, Khan et al[25] showed that the mean prerandomization $HDRS_{17}$ total score was associated with higher rates of trial success (operationalized as the degree of reduction of symptoms according to the $HDRS_{17}$ in the antidepressant treatment group subtracted by reduction in symptoms in the placebo group). Consequently, the results of these studies all point in the same direction, namely, that higher baseline severity is associated with greater difference in efficacy between antidepressants and placebo, in favor of the former group, ie, with greater chances of successful trial outcome. This has been replicated by other groups as well.[6,27,28] Notably, a meta-analysis by Papakostas and Fava[28] based on study-level data from 182 RCTs and more than 36,000 patients confirmed that greater depressive severity at baseline predicts a greater risk ratio of responding to antidepressants than placebo (defined as a 50% or greater reduction in symptom scores during treatment) and further showed that this tendency was driven by a progressively decreased response to placebo with increased depressive severity. This was true when controlling for possible confounders such as year of publication, probability of receiving placebo, dosing strategy (fixed versus flexible), age, and trial duration.

The findings outlined above suggest that setting a higher severity threshold for inclusion into clinical trials may reduce placebo response rates and increase the probability to detect an effect of investigational medication over placebo. Ironically, however, this was not found to be the case in a large, landmark meta-analysis by Khan et al.[25] Specifically, although, as discussed in the previous paragraph, baseline severity was found to positively correlate with a more favorable trial outcome for antidepressants versus placebo, the same outcome was not found to be the case when severity threshold for inclusion (as a function of the severity inclusion criterion for each trial) in the study was analyzed. A likely explanation offered by the authors for this rather counterintuitive finding is that pressure on clinicians to enroll patients (or for patients to be enrolled) may have led to baseline score inflation, ie, a consistent tendency to selectively inflate scores of patients during the entry (eligibility) assessment visit, resulting in an artificial improvement seen during subsequent visits.[29] When postbaseline ratings return to scores reflecting actual clinical severity, a pseudoimprovement becomes apparent, which serves to decrease signal-to-noise ratio in clinical trials and, thereby, increase the risk of a failed or uninformative study.[30] In fact, this tendency for severity of site investigator ratings to be higher than those of centralized raters during eligibility assessments for study entry but not thereafter has been demonstrated at least once.[31] As a result, the use of remote assessments of severity conducted by unbiased investigators is now widely seen as a standard in the field, although such methods do not necessarily remove patient bias to inflate severity for inclusion. Another issue raised in the relevant section of this review that can crucially influence RCT outcome involves the actual symptom profile comprising severe MDD.[32]

Chronicity of the index episode of depression. Greater chronicity of a depressive episode has also been associated with poor response to antidepressants in non–placebo-controlled RCT study settings.[33,34] However, a recent study-level meta-analysis[35] showed that response rates to antidepressants were not significantly different in patients with dysthymia, a depressive condition defined by a mild and chronic course, persistent symptoms, and an insidious onset, compared to patients with MDD. Interestingly, in this study, patients with dysthymia showed significantly lower placebo response rates than MDD patients, and the probability to detect an antidepressant-placebo separation was found to

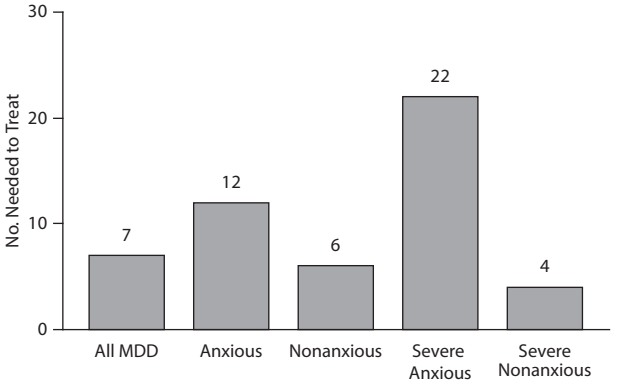
be significantly higher in clinical trials for patients with dysthymia than in trials for patients with MDD.

It should be noted, however, that modern RCTs typically employ crude measures to assess onset of the current episode. It remains unclear whether lifetime burden of illness (ie, whether the percentage of one's lifetime one has spent being euthymic versus depressed) can serve as a better predictor of RCT outcome. Similarly, illness severity often varies throughout an episode, with longer episodes, on average, having greater chances of demonstrating higher intraepisode variability of severity than shorter ones. Perhaps, then, a measure incorporating cumulative severity over time can serve as a more useful tool for assessing patient outcome in RCTs or RCT outcome itself. The aforementioned criteria outlined by Desseilles et al[12] do attempt to integrate the temporal pattern of severity of the index episode (requiring symptoms be persistent as well as pervasive and pathological). It would be interesting to examine whether any of these 3 parameters impact RCT outcome more so than the others.

Age of participants. Age of patients included in clinical trials is another factor that has been thought to potentially affect clinical trial outcome. When age has been included as a covariate in meta-analyses of MDD RCTs, it has generally not been found to be significantly associated with trial outcome.[23,26,28] While this finding is informative with respect to the existence of a linear relationship between age and clinical trial outcome extending across ages represented by the bulk of these datasets (ie, adults, on average, aged 40 years), it does not rule out differences in specific age groups relatively underrepresented (ie, those subjects older than 65 years of age, since these patients are typically excluded from phase 3 RCTs, or patients younger than 25 years, since they tend to be underrepresented). Consequently, if an actual effect of age on treatment outcome mainly applies to populations older than 65 years, it is likely to have gone undetected using these methods. Two meta-analyses[36,37] have confronted this limitation by also including RCTs with elderly subjects among the participants, while stratifying their analyses on age as a categorical rather than continuous measure. First, a study-level meta-analysis by Nelson et al[36] assessed the efficacy of second-generation antidepressants from 10 RCTs, pooling data for patients above the age of 60 years, and concluded that these agents were more effective than placebo in this age group, albeit with significant statistical variability between studies and an overall modest difference in efficacy (less than 10% difference in response rate, which translated to a number needed to treat for response of 10 versus 6[28] in reports examining all adults with MDD regardless of age). Subsequently, a study-level meta-analysis by Tedeschini et al[37] included trials of both first- and second-generation antidepressants involving patients 55 years of age or older and found that, while antidepressants did beat placebo overall, this was not the case for the subset of patients aged 65 years and above. Interestingly, the results showed that, when adjusting for confounders such as year of publication, baseline severity, study duration, and probability of receiving placebo, placebo response rates were similar in the trials comparing elderly patients (older than 55 years) versus nonelderly adults, while drug response rates were significantly lower among patients older than 65 years versus nonelderly adults. The latter finding indicates that the lack of efficacy of antidepressants over placebo among those aged ≥ 65 years is due to lower antidepressant response rather than increased placebo effect.[37] Several elements common in the elderly, such as executive dysfunction,[38] white matter hyperintensities,[39,40] Axis III comorbidity,[41] and chronic depression,[34] could potentially moderate lower antidepressant response rates in this population. Studies designed to take such factors into account would be helpful to tease out whether old age per se is indeed associated with poor antidepressant response or whether this is accounted for by other confounding factors. More recently, Nelson et al[42] conducted the first patient-level meta-analysis of placebo-controlled RCTs of antidepressants in the elderly and found greater illness duration and severity to predict greater differences in efficacy between antidepressants and placebo, echoing respective findings reviewed in those previous sections of this review, although age was not found to be a predictor after controlling for the duration of the index episode of depression. Finally, there is emerging albeit preliminary data suggesting that the effect of antidepressants versus placebo may also be smaller among young adults (ie, ages 25 years or younger) versus those between 25 and 65 years of age.[43]

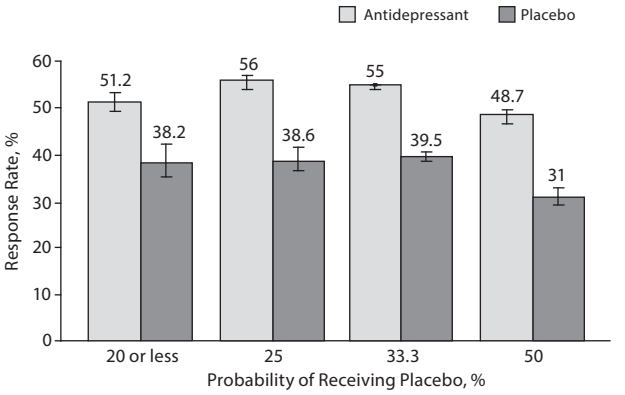
Medical and psychiatric comorbidity. The exclusion of subjects with comorbid medical and psychiatric disorders, albeit at the expense of limited generalizability of study findings to such populations, is a widespread practice that has been thought to optimize clinical trial outcome by creating a more homogeneous population in terms of potential underlying MDD etiology. However, there is little evidence to suggest that this practice achieves its goal with respect to Axis III comorbidity. Clearly, numerous non–placebo-controlled studies that examine the impact of medical and psychiatric comorbidity on antidepressant outcome suggest lower levels of symptom improvement among patients with greater comorbidity burden than those without.[44–46] However, careful analyses of placebo-controlled RCTs in populations selected for the presence of certain comorbidities do not support the notion that exclusion of all types of comorbidity necessarily results in enhanced antidepressant-placebo treatment differences. For instance, in a study-level meta-analysis by Iovieno et al,[47] the margin of efficacy (risk ratio of response) of antidepressants versus placebo was statistically similar between studies which focused on patients with MDD and comorbid Axis III disorders and general MDD studies, with a trend toward statistical significance for higher placebo response rates and significantly higher antidepressant response rates among patients with comorbid Axis III disorder compared to patients with MDD only. This study would suggest that restricting entry criteria based on the presence of medical illness for the sake of better chances of signal detection rather than patient safety does not appear to be an effective

Figure 2. Numbers Needed to Treat for Remission of Major Depressive Disorder (MDD) Following 8 Weeks of Monotherapy With a Selective Serotonin Reuptake Inhibitor Versus Placebo for Various Populations[a]



[a]Reprinted with permission from Papakostas et al.[32]

Figure 3. Probability of Placebo and Response Rates (N = 36,385; 262 drug-placebo pairwise comparisons)[a]



[a]Reprinted with permission from Papakostas and Fava.[28]

approach to warrant offsets with respect to a study finding generalizability.[48]

With respect to certain Axis I comorbid disorders, however, there is some evidence that treatment effect sizes may differ across MDD populations. For example, in a study-level meta-analysis[49] of 4 RCTs among patients with comorbid opiate use disorder on methadone maintenance therapy, the effect of antidepressants was not greater than that of placebo. Similarly, results from the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) indicated that the presence of a greater burden of anxiety symptoms in depression are associated with a poor treatment outcome in citalopram nonresponders.[45] While a number of patient-[50,51] and study-level[52] meta-analyses suggest that the presence or absence of a greater burden of anxiety symptoms in MDD does not influence the relative efficacy of antidepressants versus placebo in MDD trials, the limitation of these analyses has been that they do not control for illness severity (which influences drug-placebo differences in MDD and is also more common in anxious than nonanxious MDD). Interestingly, in a patient-level meta-analysis of 1,690 subjects from RCTs of selective serotonin reuptake inhibitor antidepressants, Papakostas et al[32] recently demonstrated that severe depression with a high burden of anxiety symptom features was associated with a very high number needed to treat (to obtain remission) compared to the nonanxious subtype of severe depression (Figure 2), an important finding that has yet to be replicated.

Not all studies support this trend, however. For instance, a study-level meta-analysis[53] showed that MDD patients with comorbid alcohol use disorders had higher, although not significantly higher, antidepressant and placebo response rates compared with pure MDD patients, and found no difference in the probability to detect a drug-placebo separation between studies of MDD patients with or without alcohol use disorders. In general, with respect to the issue of comorbidity and study outcome, it is also worthwhile pointing out that many studies have yet to be replicated. Doing so, preferably with patient-level data, should be a priority.

Role of Study-Related Factors

Probability of receiving placebo. It seems intuitive that the chances of receiving active treatment versus placebo would also influence placebo response rates since a greater chance of receiving placebo might dampen patient and rater expectation of improvement (anticipation), thereby limiting the placebo response rate.[54] Both Khan et al[23] and Stein et al[55] (the former a study-level and the latter a patient-level meta-analysis) have detected an inverse relationship between the number of treatment arms, a proxy of the probability of receiving placebo, and antidepressant-placebo separation in clinical trials for depression. However, in those studies potential confounding from other sources was not accounted for (ie, year of publication, baseline severity of illness, etc). Subsequently, in a study-level meta-analysis of 182 RCTs of antidepressants, Papakostas and Fava[28] reported that a greater likelihood of receiving placebo predicted greater antidepressant-placebo separation in clinical trials for MDD, which was primarily derived from lower placebo response rates as the probability of placebo increased. This was true when other potential confounders were controlled, including year of publication, baseline severity, age of subjects, trial duration, dosing strategy (fixed or flexible), and sample size (Figure 3).

Similar results were found in a recent study-level meta-analysis by Sinyor et al,[56] where response rates for placebo were significantly lower in studies comparing 1 rather than 2 or more active medications (34.3% vs 44.6%, $P = .003$).

Use of prospective treatment phases (lead-in). A method used in earlier trials to try to manipulate anticipation of improvement, and thereby attempting to minimize the placebo response rates in clinical trials, involved the use of a single-blind placebo lead-in phase of short duration, usually between 3–14 days, just preceding randomization. In this design, patients are unknowingly treated with placebo, but

Figure 4. Sequential Parallel Comparison Design Trial Design Variant

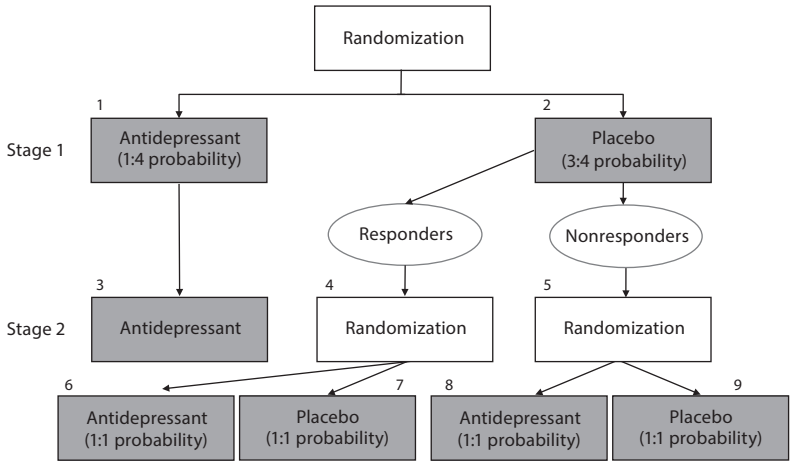
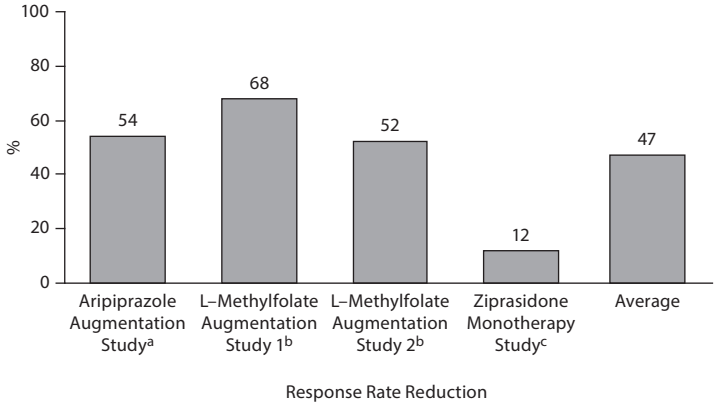


Figure 5. Percentage Reduction in Placebo Response Rates in Phase 2 (study primary outcome measure) From Sequential Parallel Comparison Design Trials



[a]Data from Fava et al.[13]
[b]Data from Papakostas et al.[16]
[c]Data from Papakostas et al.[61]

their clinicians are aware of the placebo treatment. Thus, the rationale is that those responding to placebo can be excluded prior to the actual double-blind portion of the RCT.[57,58] However, a study-level meta-analysis of 101 RCTs revealed that a placebo run-in phase did not "(1) lower the placebo response rate, (2) increase the drug-placebo difference, or (3) affect the drug response rate post-randomization in either inpatients or outpatients for any antidepressant drug group."[58(p33)] Trivedi and Rush attributed the lacking effect of the placebo lead-in phase to the fact that "the process of consenting, being educated, being screened for eligibility, and the subsequent post-screen visits even without a pill placebo are just as effective as a pill placebo in identifying placebo run-in 'responders.'" That is, the pill placebo may add little except cost in identifying run-in phase responders.[58(p37)] An alternative explanation is that, given the time course of placebo response in MDD RCTs, the majority of which extend over 4 weeks, the lead-in phase has been too short to identify most placebo responders and that (1) raters know that patients are on placebo and, therefore, their expectation is according, and (2) the issue of score inflation remains in light of imminent randomization at the end of the lead-in period.

A relatively new design for RCTs, the sequential parallel comparison design (SPCD), developed by Fava et al,[59,60] aims at reducing the placebo-response rate, thereby reducing the required sample size (Figure 4). One key aspect of the SPCD is that the overall likelihood of receiving placebo at some point during the study is relatively high. In brief, the SPCD consists of 2 stages of treatment: the first stage involves an unbalanced randomization between placebo and active treatment, with the majority of patients randomized to placebo (eg, experimental treatment–placebo ratio of 1 to 3 as outlined in the figure above). In the second stage, placebo-treated patients continue to either experimental treatment or placebo. Since these patients have already failed placebo during phase 1 (ie, they are placebo nonresponders), their placebo response in stage 2 is likely to be reduced. Furthermore, those not responding to experimental treatment in stage 1 will be treated with placebo in stage 2. The final analysis comparing experimental treatment versus placebo pools the data from both stages (ie, groups 1 and 8 for drug, and 2 and 9 for placebo).[59] There are 4 central aspects of the SPCD, which will manipulate both clinician and patient expectation of improvement, thereby reducing the anticipation effect and thus the placebo response: (1) there is a long lead-in period (stage 1) allowing sufficient time for a potential placebo response to manifest; (2) patients and clinicians remain blinded during the course of stage 1 as to treatment assignment; (3) all patients continue in stage 2 regardless of improvement or lack thereof in stage 1, thereby removing any bias for overscoring or underscoring improvement; and (4) the cumulative probability of receiving placebo at some point during the study is greater than 50%, thereby further reducing placebo response rates. A number of depression studies utilizing versions of this design have been completed (references 13, 16, 61, and www.clinicaltrials.gov trial NCT01500200, NCT01500200, and NCT01318434) or are in progress (www.clinicaltrials.gov trials NCT01204918, NCT01654796, NCT01665950, NCT01784666, and, NCT01913535), 3 of which have been funded by the National Institute of Mental Health (NCT01204918, NCT01654796, and NCT01913535). Thus far, this design has been proved to be very successful in reducing placebo response rates (Figure 5).

Dosing schedule. Antidepressant dosing schedule (flexible dose versus fixed dose) is another factor of clinical trial design that could potentially influence trial outcome. In a fixed dosing schedule, patients are assigned to a specific dose that cannot be changed during the trial. This is opposed to a flexible-dose design, where the clinician adjusts the dose in order to maximize therapeutic response and minimize side effects. On one hand, fixed-dose trials may reduce placebo response rates that may be precipitated during the course of a study as patient and raters anticipate potential improvement with each dose increase. On the other hand, flexible-dose trials may enhance signal detection by (1) balancing efficacy with tolerability and, thereby, adherence, and (2) allowing for a dosing that can be more flexibly adapted from patient to patient (ie, metabolism or body mass index differences). A study-level meta-analysis by Khan et al[62] attempting to examine these hypotheses showed that studies employing the flexible-dose schedule were associated with a significantly lower placebo response compared to those with a fixed-dose schedule and also with a higher likelihood of statistically significant separation between experimental treatment and placebo. Symptom reduction was similar in the antidepressant arms of flexible- and fixed-dose trials. These findings were, however, not supported by the aforementioned meta-analysis from Papakostas and Fava,[28] which indicated that dosing schedule does not impact clinical trial outcome but that flexible dosing is associated with increased antidepressant response rates. One factor to keep in mind when exploring the results of these meta-analyses is that, often, in phase 3 programs, fixed-dose studies may actually be designed with the intention of some treatment arms not to show efficacy (so that a dose-response trend is demonstrated). Ultimately, the decision whether to choose a fixed- versus flexible-dose design with success in mind should be driven by results of different studies examining target engagement at different doses of the drug (ie, receptor or transporter occupancy, change in measurable parameters such as vital signs, electroencephalographic findings, or neuroimaging parameters), as balanced by tolerability across doses rather than meta-analyses. For instance, it may be unwise to employ a dose range for an agent associated with good tolerability at near-full receptor or transporter occupancy. On the other hand, a dose range may make sense for an agent where better efficacy can be seen at dose ranges not well tolerated by all patients.

Trial duration. The decision on treatment duration in clinical trials of depression is a critical one. Shorter durations may risk not detecting a treatment effect for an agent that has a slower than anticipated mechanism of action, as well as for an agent that requires slow titration to the minimally effective dose. On the other hand, longer durations can result in an increase in placebo response rates; a potential for higher attrition rates resulting in increased reliance on imperfect imputation methods (ie, last observation carried forward, mixed-effect model repeated measures [MMRM]) in order to utilize data from patients who prematurely discontinued treatment; and an unnecessary prolongation of patients' exposure to placebo or an ineffective drug. Walsh et al[7] and others[63] have noted that, in antidepressant, placebo-controlled clinical trials, there is often a statistical difference in mean depressive symptoms score reduction between antidepressant- and placebo-treated patients by the third week, and almost always by the fourth week after randomization. Baldwin et al[64] found that the chance of responding beyond week 4 in MDD trials was 20% or less if no effect had occurred by week 2. In their study-level meta-analysis of clinical trials of at least 4 weeks' duration, Papakostas and Fava[28] found that treatment duration past 4 weeks did not influence the separation between experimental treatment and placebo. This finding indicated that, titration schedule permitting, restricting the trial duration could represent a way of minimizing the burden on the patients (short duration of exposure to placebo and experimental treatment), thereby reducing patient attrition rates, and curtail the overall cost of the trial (eg, less assessments). A subsequent study-level meta-analysis[65] by our group showed that 4 weeks should be considered the minimum trial duration to reliably detect or rule out differences between experimental treatments and placebo, primarily because of the increased risk of type II errors in shorter trials, ie, erroneously concluding that an effective treatment is ineffective. Replicating these findings with patient-level data can help further our understanding of what constitutes optimal trial outcome in MDD in general as well as a function of an agent's titration schedule to the minimally effective dose.

Frequency of follow-up assessments. The number and temporal distribution of follow-up assessments during a clinical trial is also a factor that may influence the outcome of a trial. On one hand, more frequent assessments may allow for more precise measurement and finer medication titration. On the other hand, more frequent measurements may unnecessarily expose patients to increased study burden as well as inflate placebo response rates due to increased exposure to nonspecific treatment effects associated with the trial. Posternak et al[66] (study-level data) showed that more frequent follow-up assessments during the trial was associated with a greater reduction in depression severity scores in both antidepressant- and placebo-treated patients. Effect on study outcome was not examined in that study. In a more recent study-level meta-analysis,[67] we detected that the effect of assessment frequency was disproportional in the antidepressant and the placebo arm, such that a greater antidepressant-placebo separation was seen in studies with a greater number of assessments. Moreover, we found that it was the frequency of assessments after week 3 that influenced trial outcome, rather than the frequency of assessments in the first 3 weeks of the trial, with less frequent assessments associated with poorer study outcome. This led us to believe that clinical trials could be optimized by (1) conducting the necessary number of assessments during the first 3 weeks of the trial serving, in addition to measurement of outcome, to optimize the dose of antidepressant therapy and (2) retaining the same frequency of those assessments after week 3 rather

than tapering their frequency.[67] It has also been suggested to simplify (shorten) study visits and assessments in order to limit the nonspecific therapeutic effects associated with study participation, thereby reducing the placebo response.[59] Since outcome measurement requires most of the time allotted to clinical visits in studies, it is therefore wise to choose a measure that is both brief and sensitive. Finally, it is also worth mentioning that the issue of visit frequency in studies of elderly persons with MDD may require separate attention, since at least 1 study-level meta-analysis[68] has shown an inverse relationship between visit frequency and antidepressant-placebo differences in efficacy, specifically in elderly patients.

Study outcome measure. In psychiatric clinical trials, as in any branch of medicine, choosing a valid outcome measure is of utmost importance. In major depression, where surrogate biologic markers of illness severity are not yet available, formalized clinical assessment must necessarily form the basis for determining efficacy. As a consequence, clinical assessment has been operationalized in a number of rating scales that allow quantification of depressive symptomatology.[69,70] An ideal rating scale for depression should meet 2 essential criteria: it must be clinically valid (capturing the severity of depression from its absence to the most severe cases) and psychometrically unidimensional (the symptoms represented by the scale items appear orderly as the severity of depression increases). When these criteria are met, each individual item adds unique information regarding severity, and the item scores can therefore be added to a clinically and mathematically meaningful total score.[69]

In depression trials, at least from a regulatory perspective (ie, US Food and Drug Administration), the $HDRS_{17}$[71] and the Montgomery-Asberg Depression Rating Scale (MADRS)[72] represent the 2 gold standard instruments for severity measurement and thus for establishing and comparing the efficacy of treatments. However, both of these scales have several inherent weaknesses as quantitative measures of major depression. For instance, in a landmark article[73] from 1975, it was shown that the total score of the $HDRS_{17}$ did not reflect the global severity of depression as perceived by experienced clinicians. In addition, it has been demonstrated in several studies that neither the $HDRS_{17}$[74–76] nor the MADRS[77,78] are unidimensional, which implies that using the total score of the scale is problematic from a mathematical and statistical perspective. Another concern is that a rating scales' sensitivity to change may not be equal across the depression severity spectrum. This entails that apparent differences in efficacy at different severity levels may be a result of the assessment instrument and not treatment itself.[79,80] Santor and Coyne[81] showed that many items from the $HDRS_{17}$ were insensitive to change in the lower end of the illness severity spectrum. Isacsson and Adler[79] reanalyzed the data from the meta-analysis conducted by Fournier et al[27] by means of item-response theory and showed that the precision of ratings for the $HDRS_{17}$ and its sensitivity to change decreased at decreasing levels of depressive severity. Therefore, many failed trials of antidepressants as well as the low effect size found in the positive trials may be consequences of the shortcomings of the $HDRS_{17}$ and MADRS and not lack of efficacy of the treatment.[79] In order to solve this important problem, it has been suggested that shorter, but yet clinically valid, and unidimensional versions of the $HDRS_{17}$ and MADRS should be employed instead of the full scales.[69,70] For instance, several short versions of the $HDRS_{17}$ have been developed (see Helmreich et al[82] for a review), and it appears that these short scales may indeed be superior to the full scales in terms of drug-placebo separation.[82–84]

Finally, it must be stressed that even the most sophisticated scale may not be reliable in detecting changes in depressive severity if it is not administered in a meticulous manner by certified and experienced raters.[85] In terms of interrater reliability and test-retest reliability, the $HDRS_{17}$ generally performs relatively well,[86] although there is some evidence to support that centralized ratings may be more reliable (less prone to baseline inflation) compared to site-based ratings.[31] Thus, the ideal approach in depression trials may be to have centralized ratings on short unidimensional rating scales. Such an approach could and should be tested empirically.

DISCUSSION

Decades of accumulating data from RCTs in MDD are beginning to reveal insights into factors that influence placebo response rates as well as clinical trial outcomes. As a result, our increasing understanding of the placebo response in MDD RCTs lends to a gradually more predictable phenomenon and, hopefully, to one that is lesser in magnitude and variability. Several key elements are emerging as critical to the ultimate success of a clinical trial and should be considered. Clearly, the probability of receiving placebo versus a comparator perceived as potentially active is a critical element. Treatment arms should also be kept at a necessary minimum in studies. Dosing schedule and study duration should reflect what is known regarding the pharmacokinetic and pharmacodynamic properties of agents tested, initiating at or reaching minimally effective doses as quickly as possible, while unnecessarily long studies should be avoided for several reasons. Visit frequency should be adequate in order to ensure proper titration of agents employed as well as measurement, particularly toward the end of the study, rather than being minimized or tapered once titration has been completed.

The use of blinded lead-in phases of sufficient duration should be strongly considered. The SPCD is one such design approach involving the use of a blinded lead-in phase that has been proven to be successful in reducing placebo response rates. One question regarding the use of SPCD-type variants is whether alternate treatment assignment probabilities can lead to further gains in signal detection. Specifically, in light of the inverse relationship between the probability of placebo and placebo response rates, it is quite possible than assigning patients to a 2:1 probability of drug versus placebo in stage 1 in favor of placebo may result in a more efficient blinded

lead-in phase whereby the lower than 50% probability of receiving placebo results in higher placebo response rates and thereby the generation of a population with inherently even lower placebo response rates in stage 2. In stage 2, all patients would then be rerandomized in a 1:1 fashion to either drug or placebo, the increase of the probability of placebo in stage 2 to 50% serving to further aid in signal detection. It would be worthwhile examining whether this SPCD variant can achieve better results than the traditional one and, if so, whether there would be any trade-offs in terms of overall sample size. Such an approach, generating a population with an inherently low probability of placebo in stage 2, may be even better suited in the study of biomarkers in MDD or in the design of translational pilot studies. From a measurement perspective, remote independent assessments of illness validity, illness severity, and treatment refractoriness, whether for the determination of participant inclusion or assessment of outcome, are becoming standard. This is particularly important given strong incentives for including patients in clinical trials (patient access to care, meeting enrollment deadlines, financial reimbursement of patient and site), especially since access to care has become more difficult over time and financial realities and pressure of running a clinical trial program, whether academic based or not, more challenging.

From a patient perspective, the confluence of information suggests that trial success increases progressively as illness duration and severity increases, while there appears to be a negative correlation between trial success and the severity of comorbid anxiety as the severity of depression increases. Given existing patient and local assessment biases toward subject inclusion, remote assessment of eligibility with respect to these parameters as well (not only baseline severity) should also serve as a standard. Finally, for studies conducted in subjects older than 65 years, specific measures should be employed to assess for population-related factors that may influence antidepressant- versus placebo-response rates as well as clinical trial outcome (eg, cognitive measures). While each of these elements may contribute a modest amount to overall trial success, their use in parallel can make a critical difference.

Ultimately, from a drug development perspective, the use of biomarkers complementary to the mechanism of action of the tested agent can further aid in RCT signal detection. In parallel, the placebo response phenomenon in RCTs itself can be conceptually divided into 2 broad phenomena: one encompassing a putative underlying biologic process manifesting in symptom improvement (whether induced by participation in the study or simply a temporal coincidence) and an artificial one driven by biases and measurement error. In order to better understand the pathophysiology of the former (which could aid in the design of clinical trials as well as, potentially, uncover further targets for drug development), minimization of the latter is of paramount importance. Therefore, translational studies aimed at discovering the underlying mechanism of action of existing agents, putative agents, and placebo should be designed with equal rigor as RCTs designed to detect the presence of a treatment effect. Unfortunately, to date this appears to be the exception rather than the case, resulting in the hampering of efforts of translational and mechanistic studies. Therefore, raising the standard of mechanistic and translational studies should become a top priority for the field.

Limitations and Conclusion

Several main limitations of this work should be noted. First, as with any review article, important contributions to the literature may have been inadvertently omitted from inclusion in the present review. This may particularly be the case with new and unpublished works. Similarly, other experts in the field may have reviewed the same articles and synthesized the review in a different way and with different conclusions. Clinicians and researchers interested in furthering their knowledge in this topic should seek additional sources of information, whether in the literature or at educational events. In addition, it also important to point out that, while some findings discussed in this review have been confirmed in both study-level as well as patient-level meta-analyses, others have not. Since patient-level meta-analyses can enhance statistical power and also allow the investigators to control for variables not routinely reported in clinical trials, it would be important to develop for the field a more formal framework so that investigators can test their hypotheses on large, existing patient-level datasets. Furthermore, the purpose of this article was to provide insights that would lead to more modest and easily predictable placebo response rates in clinical trials of MDD. Therefore, it is not best suited to answer the important question of how can we actually enhance placebo response rates in clinical practice. Finally, the overwhelming bulk of information presented here derives from adult patients with nonpsychotic MDD. Whether these findings are generalizable to children and adolescents with depression or to subjects with bipolar depression or psychotic MDD remains unclear. Separate efforts should be made to determine factors that influence placebo response rates and trial outcome in these and other psychiatric illnesses as well, since substantial and variable placebo response rates have also been reported in other Axis I disorders, including bipolar disorder,[87] where the magnitude of the placebo response rate seems similar to that of MDD[88] as well as schizophrenia[89] and generalized anxiety disorder.[55]

In conclusion, our increasing understanding of the placebo response in MDD RCTs lends to a gradually more predictable phenomenon and, hopefully, to one that becomes lesser in magnitude and variability. Several elements have emerged that seem to play a critical role in trial success, gradually reshaping the design of clinical, translational, as well as mechanistic studies in depression.

Drug names: citalopram (Celexa and others), methadone (Methadose and others), ziprasidone (Geodon).

Author affiliations: Massachusetts General Hospital, Harvard Medical School, Boston (Drs Papakostas and Iovieno); and Research Department P, Aarhus University Hospital–Risskov, Risskov, and The Lundbeck Foundation Initiative for Integrative Psychiatric Research (*i*PSYCH), Copenhagen (Dr Østergaard), Denmark.

Potential conflicts of interest: **Dr Papakostas** has served as a consultant for Abbott Laboratories, AstraZeneca PLC, Avanir, Brainsway, Bristol-Myers Squibb, Cephalon, Dey Pharma LP, Eli Lilly, Genentech, GlaxoSmithKline, Evotec AG, H. Lundbeck A/S, Inflabloc, Jazz, Novartis AG, Otsuka, PAMLAB, Pfizer, Pierre Fabre, Ridge Diagnostics (formerly known as Precision Human Biolaboratories), Shire, Sunovion, Takeda, Theracos, and Wyeth; has received honoraria from Abbott Laboratories, AstraZeneca PLC, Avanir, Bristol-Myers Squibb, Brainsway, Cephalon, Dey Pharma LP, Eli Lilly, Evotec AG, GlaxoSmithKline, Inflabloc, Jazz, H. Lundbeck A/S, Meiji Seika, Novartis Pharma AG, Otsuka, PAMLAB, Pfizer, Pierre Fabre, Ridge Diagnostics, Shire, Sunovion, Takeda, Theracos, Titan, and Wyeth; has received research support from AstraZeneca PLC, Bristol-Myers Squibb, Forest, National Institute of Mental Health, PAMLAB, Ridge Diagnostics (formerly known as Precision Human Biolaboratories), Sunovion, and Theracos; and has served (not currently) on the speaker's bureau for Bristol-Myers Squibb and Pfizer. **Drs Iovieno** and **Østergaard** report no financial or other relationship relevant to the subject of this article.
Funding/support: None reported.

REFERENCES

1. Kessler RC, Bromet EJ. The epidemiology of depression across cultures. *Annu Rev Public Health*. 2013;34(1):119–138.
2. Murray CJ, Vos T, Lozano R, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380(9859):2197–2223.
3. American Psychiatric Association (APA). Practice Guideline for the Treatment of Patients with Major Depressive Disorder, 3rd ed. 2010. http://psychiatryonline.org/pb/assets/raw/sitewide/practice_guidelines/guidelines/mdd.pdf. Accessed November 13, 2014.
4. Lam RW, Kennedy SH, Grigoriadis S, et al; Canadian Network for Mood and Anxiety Treatments (CANMAT). Canadian Network for Mood and Anxiety Treatments (CANMAT) clinical guidelines for the management of major depressive disorder in adults, 3: pharmacotherapy. *J Affect Disord*. 2009;117(suppl 1):S26–S43.
5. National Institute for Health and Clinical Excellence (NICE). 2010. The treatment and management of depression in adults: updated edition. National Clinical Practice Guideline 90. http://www.nice.org.uk/guidance/cg90/resources/cg90-depression-in-adults-full-guidance2. Accessed December 16, 2014.
6. Khin NA, Chen YF, Yang Y, et al. Exploratory analyses of efficacy data from major depressive disorder trials submitted to the US Food and Drug Administration in support of new drug applications. *J Clin Psychiatry*. 2011;72(4):464–472.
7. Walsh BT, Seidman SN, Sysko R, et al. Placebo response in studies of major depression: variable, substantial, and growing. *JAMA*. 2002;287(14):1840–1847.
8. Iovieno N, Papakostas GI. Correlation between different levels of placebo response rate and clinical trial outcome in major depressive disorder: a meta-analysis. *J Clin Psychiatry*. 2012;73(10):1300–1306.
9. Ostergaard SD, Jensen SO, Bech P. The heterogeneity of the depressive syndrome: when numbers get serious. *Acta Psychiatr Scand*. 2011;124(6):495–496.
10. Ripke S, Wray NR, Lewis CM, et al; Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol Psychiatry*. 2013;18(4):497–511.
11. Bienvenu OJ, Davydow DS, Kendler KS. Psychiatric 'diseases' versus behavioral disorders and degree of genetic influence. *Psychol Med*. 2011;41(1):33–40.
12. Desseilles M, Witte J, Chang TE, et al. Massachusetts General Hospital SAFER criteria for clinical trials and research. *Harv Rev Psychiatry*. 2013;21(5):269–274.
13. Fava M, Mischoulon D, Iosifescu D, et al. A double-blind, placebo-controlled study of aripiprazole adjunctive to antidepressant therapy among depressed outpatients with inadequate response to prior antidepressant therapy (ADAPT-A Study). *Psychother Psychosom*. 2012;81(2):87–97.
14. Ratti E, Bettica P, Alexander R, et al. Full central neurokinin-1 receptor blockade is required for efficacy in depression: evidence from orvepitant clinical studies. *J Psychopharmacol*. 2013;27(5):424–434.
15. Papakostas GI, Shelton RC, Kinrys G, et al. Assessment of a multi-assay, serum-based biological diagnostic test for major depressive disorder: a pilot and replication study. *Mol Psychiatry*. 2013;18(3):332–339.
16. Papakostas GI, Shelton RC, Zajecka JM, et al. L-methylfolate as adjunctive therapy for SSRI-resistant major depression: results of two randomized, double-blind, parallel-sequential trials. *Am J Psychiatry*. 2012;169(12):1267–1274.
17. Raison CL, Rutherford RE, Woolwine BJ, et al. A randomized controlled trial of the tumor necrosis factor antagonist infliximab for treatment-resistant depression: the role of baseline inflammatory biomarkers. *JAMA Psychiatry*. 2013;70(1):31–41.
18. Murrough JW, Iosifescu DV, Chang LC, et al. Antidepressant efficacy of ketamine in treatment-resistant major depression: a two-site randomized controlled trial. *Am J Psychiatry*. 2013;170(10):1134–1142.
19. Furey ML, Zarate CA Jr. Pulsed intravenous administration of scopolamine produces rapid antidepressant effects and modest side effects. *J Clin Psychiatry*. 2013;74(8):850–851.
20. Klerman GL, Cole JO. Clinical pharmacology of imipramine and related compounds. *Pharmacol Rev*. 1965;17:101–141.
21. Brown WA, Johnson MF, Chen MG. Clinical features of depressed patients who do and do not improve with placebo. *Psychiatry Res*. 1992;41(3):203–214.
22. Khan A, Leventhal RM, Khan SR, et al. Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. *J Clin Psychopharmacol*. 2002;22(1):40–45.
23. Khan A, Kolts RL, Thase ME, et al. Research design features and patient characteristics associated with the outcome of antidepressant clinical trials. *Am J Psychiatry*. 2004;161(11):2045–2049.
24. Khan A, Brodhead AE, Kolts RL, et al. Severity of depressive symptoms and response to antidepressants and placebo in antidepressant trials. *J Psychiatr Res*. 2005;39(2):145–150.
25. Khan A, Schwartz K, Kolts RL, et al. Relationship between depression severity entry criteria and antidepressant clinical trial outcomes. *Biol Psychiatry*. 2007;62(1):65–71.
26. Khan A, Bhat A, Faucett J, et al. Antidepressant-placebo differences in 16 clinical trials over 10 years at a single site: role of baseline severity. *Psychopharmacology (Berl)*. 2011;214(4):961–965.
27. Fournier JC, DeRubeis RJ, Hollon SD, et al. Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA*. 2010;303(1):47–53.
28. Papakostas GI, Fava M. Does the probability of receiving placebo influence clinical trial outcome? a meta-regression of double-blind, randomized clinical trials in MDD. *Eur Neuropsychopharmacol*. 2009;19(1):34–40.
29. Mundt JC, Greist JH, Jefferson JW, et al. Is it easier to find what you are looking for if you think you know what it looks like? *J Clin Psychopharmacol*. 2007;27(2):121–125.
30. Robinson DS, Rickels K. Concerns about clinical drug trials. *J Clin Psychopharmacol*. 2000;20(6):593–596.
31. Kobak KA, Leuchter A, DeBrota D, et al. Site versus centralized raters in a clinical depression trial: impact on patient selection and placebo response. *J Clin Psychopharmacol*. 2010;30(2):193–197.
32. Papakostas GI, Fan H, Tedeschini E. Severe and anxious depression: combining definitions of clinical sub-types to identify patients differentially responsive to selective serotonin reuptake inhibitors. *Eur Neuropsychopharmacol*. 2012;22(5):347–355.
33. Rush AJ, Wisniewski SR, Zisook S, et al. Is prior course of illness relevant to acute or longer-term outcomes in depressed out-patients? a STAR*D report. *Psychol Med*. 2012;42(6):1131–1149.
34. Pettit JW, Lewinsohn PM, Roberts RE, et al. The long-term course of depression: development of an empirical index and identification of early adult outcomes. *Psychol Med*. 2009;39(3):403–412.
35. Levkovitz Y, Tedeschini E, Papakostas GI. Efficacy of antidepressants for dysthymia: a meta-analysis of placebo-controlled randomized trials. *J Clin Psychiatry*. 2011;72(4):509–514.
36. Nelson JC, Delucchi K, Schneider LS. Efficacy of second generation antidepressants in late-life depression: a meta-analysis of the evidence. *Am J Geriatr Psychiatry*. 2008;16(7):558–567.
37. Tedeschini E, Levkovitz Y, Iovieno N, et al. Efficacy of antidepressants for late-life depression: a meta-analysis and meta-regression of placebo-controlled randomized trials. *J Clin Psychiatry*. 2011;72(12):1660–1668.
38. Alexopoulos GS, Kiosses DN, Heo M, et al. Executive dysfunction and the course of geriatric depression. *Biol Psychiatry*. 2005;58(3):204–210.
39. Papakostas GI, Iosifescu DV, Renshaw PF, et al. Brain MRI white matter hyperintensities and one-carbon cycle metabolism in non-geriatric outpatients with major depressive disorder (part 2). *Psychiatry Res*. 2005;140(3):301–307.
40. Iosifescu DV, Renshaw PF, Lyoo IK, et al. Brain white-matter hyperintensities and treatment outcome in major depressive disorder. *Br J Psychiatry*. 2006;188(2):180–185.
41. Iosifescu DV, Nierenberg AA, Alpert JE, et al. The impact of medical comorbidity on acute treatment in major depressive disorder. *Am J Psychiatry*. 2003;160(12):2122–2127.
42. Nelson JC, Delucchi KL, Schneider LS. Moderators of outcome in late-life depression: a patient-level meta-analysis. *Am J Psychiatry*. 2013;170(6):651–659.
43. Nelson JC, Zhang Q, Deberdt W, et al. Predictors of remission with placebo

using an integrated study database from patients with major depressive disorder. *Curr Med Res Opin*. 2012;28(3):325–334.
44. Trivedi MH, Rush AJ, Wisniewski SR, et al; STAR*D Study Team. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. *Am J Psychiatry*. 2006;163(1):28–40.
45. Fava M, Rush AJ, Alpert JE, et al. Difference in treatment outcome in outpatients with anxious versus nonanxious depression: a STAR*D report. *Am J Psychiatry*. 2008;165(3):342–351.
46. Rush AJ, Wisniewski SR, Warden D, et al. Selecting among second-step antidepressant medication monotherapies: predictive value of clinical, demographic, or first-step treatment features. *Arch Gen Psychiatry*. 2008;65(8):870–880.
47. Iovieno N, Tedeschini E, Ameral VE, et al. Antidepressants for major depressive disorder in patients with a co-morbid Axis-III disorder: a meta-analysis of patient characteristics and placebo response rates in randomized controlled trials. *Int Clin Psychopharmacol*. 2011;26(2):69–74.
48. Posternak MA, Zimmerman M, Keitner GI, et al. A reevaluation of the exclusion criteria used in antidepressant efficacy trials. *Am J Psychiatry*. 2002;159(2):191–200.
49. Pedrelli P, Iovieno N, Vitali M, et al. Treatment of major depressive disorder and dysthymic disorder with antidepressants in patients with comorbid opiate use disorders enrolled in methadone maintenance therapy: a meta-analysis. *J Clin Psychopharmacol*. 2011;31(5):582–586.
50. Papakostas GI, Larsen K. Testing anxious depression as a predictor and moderator of symptom improvement in major depressive disorder during treatment with escitalopram. *Eur Arch Psychiatry Clin Neurosci*. 2011;261(3):147–156.
51. Nelson JC. Anxiety does not predict response to duloxetine in major depression: results of a pooled analysis of individual patient data from 11 placebo-controlled trials. *Depress Anxiety*. 2010;27(1):12–18.
52. Nelson JC, Delucchi K, Schneider LS. Anxiety does not predict response to antidepressant treatment in late life depression: results of a meta-analysis. *Int J Geriatr Psychiatry*. 2009;24(5):539–544.
53. Iovieno N, Tedeschini E, Bentley KH, et al. Antidepressants for major depressive disorder and dysthymic disorder in patients with comorbid alcohol use disorders: a meta-analysis of placebo-controlled randomized trials. *J Clin Psychiatry*. 2011;72(8):1144–1151.
54. Rutherford BR, Roose SP, Sneed J. Mind over medicine: the influence of expectations on antidepressant response. *J Am Psychoanal Assoc*. 2009;57(2):456–460.
55. Stein DJ, Baldwin DS, Dolberg OT, et al. Which factors predict placebo response in anxiety disorders and major depression? an analysis of placebo-controlled studies of escitalopram. *J Clin Psychiatry*. 2006;67(11):1741–1746.
56. Sinyor M, Levitt AJ, Cheung AH, et al. Does inclusion of a placebo arm influence response to active antidepressant treatment in randomized controlled trials? results from pooled and meta-analyses. *J Clin Psychiatry*. 2010;71(3):270–279.
57. Faries DE, Heiligenstein JH, Tollefson GD, et al. The double-blind variable placebo lead-in period: results from two antidepressant clinical trials. *J Clin Psychopharmacol*. 2001;21(6):561–568.
58. Trivedi MH, Rush H. Does a placebo run-in or a placebo treatment cell affect the efficacy of antidepressant medications? *Neuropsychopharmacology*. 1994;11(1):33–43.
59. Fava M, Evins AE, Dorer DJ, et al. 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(3):115–127.
60. Doros G, Pencina M, Rybin D, et al. A repeated measures model for analysis of continuous outcomes in sequential parallel comparison design studies. *Stat Med*. 2013;32(16):2767–2789.
61. Papakostas GI, Vitolo OV, Ishak WW, et al. A 12-week, randomized, double-blind, placebo-controlled, sequential parallel comparison trial of ziprasidone as monotherapy for major depressive disorder. *J Clin Psychiatry*. 2012;73(12):1541–1547.
62. Khan A, Khan SR, Walens G, et al. Frequency of positive studies among fixed and flexible dose antidepressant clinical trials: an analysis of the food and drug administration summary basis of approval reports. *Neuropsychopharmacology*. 2003;28(3):552–557.
63. Stahl SM, Nierenberg AA, Gorman JM. Evidence of early onset of antidepressant effect in randomized controlled trials. *J Clin Psychiatry*. 2001;62(suppl 4):17–23, discussion 37–40.
64. Baldwin DS, Stein DJ, Dolberg OT, et al. How long should a trial of escitalopram treatment be in patients with major depressive disorder, generalised anxiety disorder or social anxiety disorder? an exploration of the randomised controlled trial database. *Hum Psychopharmacol*. 2009;24(4):269–275.
65. Tedeschini E, Fava M, Papakostas GI. Placebo-controlled, antidepressant clinical trials cannot be shortened to less than 4 weeks' duration: a pooled analysis of randomized clinical trials employing a diagnostic odds ratio-based approach. *J Clin Psychiatry*. 2011;72(1):98–113.
66. Posternak MA, Zimmerman M. Therapeutic effect of follow-up assessments on antidepressant and placebo response rates in antidepressant efficacy trials: meta-analysis. *Br J Psychiatry*. 2007;190(4):287–292.
67. Iovieno N, Tedeschini E, Levkovitz Y, et al. Does the frequency of follow-up assessments affect clinical trial outcome? a meta-analysis and meta-regression of placebo-controlled randomized trials. *Int J Neuropsychopharmacol*. 2012;15(3):289–296.
68. Rutherford BR, Tandler J, Brown PJ, et al. Clinic visits in late-life depression trials: effects on signal detection and therapeutic outcome. *Am J Geriatr Psychiatry*. 2014;22(12):1452–1461.
69. Bech P. *Clinical Psychometrics*. Chichester, UK: Wiley-Blackwell, John Wiley & Sons Ltd; 2012.
70. Østergaard SD, Meyers BS, Flint AJ, et al; STOP-PD Study Group. Measuring psychotic depression. *Acta Psychiatr Scand*. 2014;129(3):211–220.
71. Hamilton M. Development of a rating scale for primary depressive illness. *Br J Soc Clin Psychol*. 1967;6(4):278–296.
72. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry*. 1979;134(4):382–389.
73. Bech P, Gram LF, Dein E, et al. Quantitative rating of depressive states. *Acta Psychiatr Scand*. 1975;51(3):161–170.
74. Bech P, Allerup P, Gram LF, et al. The Hamilton depression scale: evaluation of objectivity using logistic models. *Acta Psychiatr Scand*. 1981;63(3):290–299.
75. Licht RW, Qvitzau S, Allerup P, et al. Validation of the Bech-Rafaelsen Melancholia Scale and the Hamilton Depression Scale in patients with major depression; is the total score a valid measure of illness severity? *Acta Psychiatr Scand*. 2005;111(2):144–149.
76. Bech P, Fava M, Trivedi MH, et al. Factor structure and dimensionality of the two depression scales in STAR*D using level 1 datasets. *J Affect Disord*. 2011;132(3):396–400.
77. Maier W, Philipp M. Comparative analysis of observer depression scales. *Acta Psychiatr Scand*. 1985;72(3):239–245.
78. Adler M, Hetta J, Isacsson G, et al. An item response theory evaluation of three depression assessment instruments in a clinical sample. *BMC Med Res Methodol*. 2012;12(1):84.
79. Isacsson G, Adler M. Randomized clinical trials underestimate the efficacy of antidepressants in less severe depression. *Acta Psychiatr Scand*. 2012;125(6):453–459.
80. Reise SP, Haviland MG. Item response theory and the measurement of clinical change. *J Pers Assess*. 2005;84(3):228–238.
81. Santor DA, Coyne JC. Examining symptom expression as a function of symptom severity: item performance on the Hamilton Rating Scale for Depression. *Psychol Assess*. 2001;13(1):127–139.
82. Helmreich I, Wagner S, Mergl R, et al. Sensitivity to changes during antidepressant treatment: a comparison of unidimensional subscales of the Inventory of Depressive Symptomatology (IDS-C) and the Hamilton Depression Rating Scale (HAMD) in patients with mild major, minor or subsyndromal depression. *Eur Arch Psychiatry Clin Neurosci*. 2012;262(4):291–304.
83. Bech P. Is the antidepressive effect of second-generation antidepressants a myth? *Psychol Med*. 2010;40(2):181–186.
84. Bech P, Boyer P, Germain JM, et al. HAM-D17 and HAM-D6 sensitivity to change in relation to desvenlafaxine dose and baseline depression severity in major depressive disorder. *Pharmacopsychiatry*. 2010;43(7):271–276.
85. Engelhardt N, Feiger AD, Cogger KO, et al. Rating the raters: assessing the quality of Hamilton rating scale for depression clinical interviews in two industry-sponsored clinical drug trials. *J Clin Psychopharmacol*. 2006;26(1):71–74.
86. Trajković G, Starčević V, Latas M, et al. Reliability of the Hamilton Rating Scale for Depression: a meta-analysis over a period of 49 years. *Psychiatry Res*. 2011;189(1):1–9.
87. Selle V, Schalkwijk S, Vázquez GH, et al. Treatments for acute bipolar depression: meta-analyses of placebo-controlled, monotherapy trials of anticonvulsants, lithium and antipsychotics. *Pharmacopsychiatry*. 2014;47(2):43–52.
88. Iovieno N, Walker RSD, Papakostas GI. Correlation between different levels of placebo response rate and clinical trial outcome in bipolar depression. Poster presented at the Annual Meeting of the American Society of Clinical Psychopharmacology; Hollywood, FL; June 16–19, 2014.
89. Agid O, Siu CO, Potkin SG, et al. Meta-regression analysis of placebo response in antipsychotic trials, 1970–2010. *Am J Psychiatry*. 2013;170(11):1335–1344.