A Meta-analysis of Factors Impacting Detection of Antidepressant Efficacy in Clinical Trials: The Importance of Academic Sites

Boadie W Dunlop*[1], Michael E Thase[2], Chuan-Chuan Wun[3], Rana Fayyad[3], Christine J Guico-Pabia[3], Jeff Musgnung[3] and Philip T Ninan[4]

[1]Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA, USA; [2]Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; [3]Pfizer Inc, Collegeville, PA, USA; [4](Formerly of) Pfizer Inc, Collegeville, PA, USA

Variability in placebo response greatly complicates the design, conduct, and interpretation of clinical trials of antidepressant medications. To identify factors that impact detection of antidepressant–placebo differences, we conducted a meta-analysis of all relevant phase II–IV clinical trials for major depressive disorder conducted by the manufacturer of venlafaxine and desvenlafaxine completed by March 2011. We examined 15 factors potentially relevant to trial outcomes, using the standardized mean difference on the Hamilton Rating Scale for Depression (HAM-D$_{17}$) score as the primary outcome. Thirty trials comprising 8933 patients were included. In univariate analyses, antidepressant efficacy (ie, drug *vs* placebo difference) was predicted most strongly ($\beta = 3.74$, $p = 0.0002$) by the proportion of patients in the trial enrolled from academic sites. Other factors predicting larger drug–placebo differences included lower participant completion rate, fewer post-baseline study visits, earlier year of study, and study drug (venlafaxine > desvenlafaxine). In multivariate meta-regression modeling, only the proportion of patients from academic sites maintained statistical significance as a predictor of drug–placebo separation for both HAM-D$_{17}$ continuous score change ($\beta = 2.24$, $p = 0.034$) and response rate ($\beta = 2.26$, $p = 0.035$). Including a higher proportion of academic sites may increase the ability to detect differences between active drug and placebo in clinical trials of major depressive disorder.
Neuropsychopharmacology (2012) **37,** 2830–2836; doi:10.1038/npp.2012.153; published online 22 August 2012

Keywords: placebo; academic medical centers; signal detection; antidepressants; venlafaxine; desvenlafaxine

INTRODUCTION

Regulatory approval for new treatments of medical conditions requires superior performance in clinical trials *vs* a comparator, often a placebo. However, patients receiving placebo can experience symptomatic improvement, particularly for subjective rather than objective outcomes (Hrobjartsson and Gotzsche 2001). Understanding the factors contributing to placebo response is therefore important for clinical trials using relatively subjective measures as primary outcomes, such as pain, fatigue, sexual function, sleep, and psychiatric disorders. Major depressive disorder (MDD) is a common psychiatric illness for which psychotherapy and antidepressant medications are considered first-line treatments (American Psychiatric Work Group on Major Depressive Disorder, 2010). However, a growing crisis of confidence exists regarding the therapeutic potential of antidepressants (Turner *et al,* 2008). Approxi-

mately half of trials of newer marketed antidepressants in the US Food and Drug Administration (FDA) database failed to demonstrate superiority over placebo (Khan *et al,* 2002). Recent meta-analyses of antidepressant study data submitted to the FDA have raised doubts of their effectiveness in relation to placebo, except for the most severely depressed patients (Fournier *et al,* 2010; Kirsch *et al,* 2008). Based on data for six antidepressants submitted to the FDA and approved between 1987 and 1999, Kirsch *et al* (2008) determined that approximately 80% of the response to medication was attributable to placebo effects. The mean drug–placebo difference in this analysis was approximately 1.8 on the Hamilton Rating Scale for Depression (HAM-D), which is below the three-point difference suggested by the National Institute for Clinical Excellence to denote clinical significance (National Collaborating Centre for Mental Health, 2004).

A commonly cited reason for poor signal detection in MDD trials is the magnitude of placebo response, which has been increasing over the last three decades (Walsh *et al,* 2002; Rief *et al,* 2009). Placebo response has been attributed to a variety of factors, such as regression to the mean, spontaneous recovery, expectation bias (Rutherford *et al,* 2010; Kraemer *et al,* 2002), clinical attention (Frank and

Frank 1991), unreliable measurement, and inclusion of inappropriate patients (Fava et al, 2003). Greater severity and longer duration of illness have been associated with greater drug–placebo separation in some (Fournier et al, 2010; Kirsch et al, 2008; Khan et al, 1991; Kobak et al, 2009), but not all (Khan et al, 2007) studies. Among trial design factors, greater placebo effects have been associated with greater number of treatment arms (Sinyor et al, 2010), more frequent trial visits (Posternak and Zimmerman 2007), and flexible rather than fixed dosing (Khan et al, 2007).

Understanding the temporal increase of placebo response requires consideration of other factors that have changed over time. Increased expectations of improvement by participants and greater reliance on advertising for recruitment have been discussed, but there are no published data to support these inferences. Additionally, over the past two decades, the number of for-profit private clinical research sites has dramatically risen, although the impact of this trend also remains unexamined (Rettig 2000).

We performed a meta-analysis to examine factors that may be associated with placebo response and detection of antidepressant efficacy in short-term trials of two antidepressants, venlafaxine and desvenlafaxine, in a series of studies conducted across nearly 20 years. We examined the impact of variables related to three aspects of clinical trial design and conduct: patient characteristics, study design, and research environment characteristics. Although the ideal approach to analyzing the effects of antidepressants vs placebo would be a patient-level meta-regression, outcome data at the individual patient level were not available for all studies, so we evaluated mean scores and predictors at the trial level for this meta-analysis.

PATIENTS AND METHODS

Trial Selection

We identified all phase II–IV placebo-controlled trials of venlafaxine or desvenlafaxine in the Pfizer database as of March 1, 2011. Both compounds were developed and brought to market by Wyeth, which was bought by Pfizer in 2009. In total, 30 trials were identified, 5 of which are unpublished. Study details are listed in Supplementary Table S1.

Inclusion criteria for trials included those that were conducted in adult outpatients aged 18 to 65 years with a primary diagnosis of MDD. Trials were required to: be 6 to 12 weeks in duration, use double-blind drug administration, and use the 17-item HAM-D (HAM-D$_{17}$) (Hamilton, 1960).

All trials excluded patients with a lifetime diagnosis of bipolar disorder or a psychotic disorder, current substance abuse or dependence, or a primary diagnosis of any other psychiatric disorder. Use of other psychoactive medications during the trials was prohibited; however, several trials allowed the use of $\leqslant 6$ doses of hypnotic medications during the first 2 weeks of treatment. Trial arms that used doses below the FDA-approved dosing ranges were excluded. Two trials (D309 and D317) included both desvenlafaxine and venlafaxine treatment arms. To avoid overweighting the effects of the predictor variables from these trials, only data from the investigational arm (desvenlafaxine) were evaluated, as recommended by the Cochrane Handbook (Higgins and Green, 2011). Because this approach can be critiqued as being vulnerable to 'results-related choices', we separately examined the effect size of venlafaxine vs placebo in the two trials. In both cases, the effect sizes were larger for venlafaxine than for desvenlafaxine. Therefore, our approach of including only the desvenlafaxine arms from these trials is a conservative one.

Factors Examined

Based on previous research, 15 variables for each trial were examined as potential predictors of drug–placebo separation and placebo response. Patient characteristics included mean age, gender, race (white vs others), proportion of patients enrolled in an academic vs a nonacademic site, and mean and median baseline depression severity. Study design characteristics included the number of treatment arms, assessments per visit, post-baseline study visits, the duration of the trial in weeks, fixed vs flexible dosing, and study drug (venlafaxine or desvenlafaxine). Research environment characteristics included year of study initiation, the percentages of patients from US sites, and trial completion rates. For some studies, the source of patients enrolled from US sites ($N = 2$ studies) or academic sites ($N = 5$ studies) was insufficiently coded to allow for confident identification, so we imputed the values for the proportion of patients recruited from United States or academic sites in these trials from the percentage of US sites or academic sites (respectively) in these trials.

Outcome Measures

The primary outcome was the effect size for change from baseline HAM-D$_{17}$, calculated as the mean difference between drug and placebo divided by the corresponding pooled SD for the difference.

Secondary outcomes were mean change from baseline effect size in placebo subjects, HAM-D$_{17}$ response rate in placebo subjects (defined as $\geqslant 50\%$ reduction in HAM-D$_{17}$ score from baseline to endpoint), between-group difference in the response rate (odds ratio), and the probability of positive study outcome (defined as statistically significant drug–placebo separation at endpoint).

Statistical Analysis

All analyses were performed using the intent-to-treat population, defined as all randomized participants who took $\geqslant 1$ dose of study medication and completed $\geqslant 1$ post-randomization HAM-D$_{17}$ evaluation. All statistical tests were two-sided with a significance level of $\alpha = 0.05$. As these analyses were exploratory in nature, no adjustments were made for multiple comparisons.

For both placebo treatment effect and drug–placebo difference, the following outcomes were calculated: mean HAM-D$_{17}$ effect size, mean HAM-D$_{17}$ change from baseline to endpoint in placebo subjects, and response rate in placebo subjects and difference in response rate between drug and placebo. All analyses were based on the last-observation-carried-forward approach, as this conservative method for accounting for the impact of study attrition was used in all of the trials. Meta-analyses and univariate meta-regression models were performed using

the Comprehensive Meta-Analysis Software System Software, Version 2 (Biostat, Englewood, NJ). Pooled placebo effect and drug–placebo differences were assessed by meta-analysis based on both fixed- and random-effect models. To account for variability among studies, the primary analyses were based on the random effects model.

The primary objective of this study was to assess the predictability of each of the 15 pre-specified factors on drug–placebo difference in the standardized mean change from baseline to end of study HAM-D$_{17}$ score via univariate analyses. A similar approach was applied to the following endpoints and was considered as secondary analyses: drug–placebo difference in the response rate, mean change from baseline to end of study HAM-D$_{17}$, and the response rate within the placebo treatment group.

Univariate meta-regression models assessed the association between predictor variables and each defined outcome. Continuous outcomes were analyzed using a linear meta-regression model, and a logistic meta-regression model was used for binary outcomes. The regression coefficients of the intercept and slope were estimated based on the random effects model. The null hypothesis that the slope equals zero (ie, no association), was tested using a standard normal Z-test. Correlations among predictor variables were assessed with weighted Pearson correlations, in which weights were obtained from the random effects model. To adjust for confounding effects among predictor variables, multivariate meta-regression models were also performed using 'metareg' command of STATA statistical software, Version 10 (Stata Corp, College Station, TX). Predictor variables found to be significant ($p < 0.05$) from univariate analysis were entered into the multivariate meta-regression model. Because of the small number of studies relative to the number of variables assessed, p-values for multivariate meta-regression analyses were based on t-tests using Monte–Carlo permutation with 10 000 replications. The relationship between the likelihood of positive study outcome and the predictor variables were assessed by weighted logistic regression analysis, where the weights were obtained from the random effects model. Furthermore, to adjust for the possible confounding effect of study year on other predictor variables, sensitivity analyses were performed by stratifying the studies by median year of study (2000) as the cut point.

Heterogeneity was assessed using the Cochran's Q and Higgin's I^2 statistics (Higgins and Thompson, 2002). Cochran's Q statistics were calculated based on the weighted sum of squared deviations distributed as χ^2 with degrees

of freedom equal to $k - 1$, where k is the number of studies. For the Q test, statistical significance was set at $p \leqslant 0.05$. Higgin's I^2 statistics describes the percentage of variability in effect size due to heterogeneity rather than chance.

RESULTS

Studies

Although the sponsor's database contained 131 phase II, III or IV studies of either venlafaxine or desvenlafaxine, 57 studies did not include a placebo and 26 did not study MDD. Among the remainder, 30 were eligible (Figure 1). The major reason for exclusion was duration (ie, < 6 or > 12 weeks; $n = 12$). A total of 5463 antidepressant-treated and 3470 placebo-treated patients were included. The effect sizes from the random effects models for the primary outcome for placebo and antidepressant–placebo differences are presented in Supplementary Figure S1.

Placebo Effects

The overall standardized mean difference (SMD) in HAM-D$_{17}$ change scores among placebo-treated patients was 1.15 ± 0.04 ($p < 0.001$), equating to a 9-point change. Significant heterogeneity was identified ($Q = 84.76$, $p < 0.0001$). After excluding the six studies that contributed to the heterogeneity, the SMD was 1.22 ± 0.03, reflecting no meaningful difference from the overall SMD estimate, so the full sample of 30 studies was used in all analyses.

The placebo response rate was 38.4% (95% CI 35.9–41.1%), with significant heterogeneity across trials ($Q = 71.19$, $p < 0.0001$); excluding the six studies that contributed to the heterogeneity did not affect the mean placebo response rate (38.1% (95% CI 36.1–40.2%)).

Predictors of Placebo Response

The analyses of factors predicting placebo response are presented in Table 1. Only completion rate predicted both higher HAM-D$_{17}$ change score and percentage of responders. Year of study and desvenlafaxine treatment, which are confounded due to the timeframe that these drugs were developed, were predictive of HAM-D$_{17}$ change score but not response rate. Number of assessments per visit also predicted HAM-D$_{17}$ change score but not response rate. Both the proportion of patients from academic and US sites
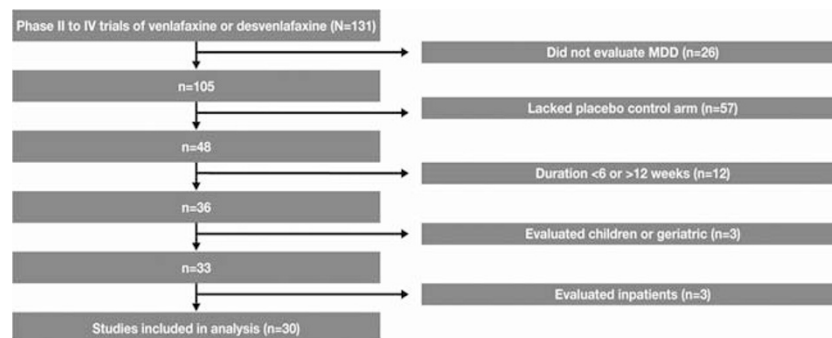
Figure 1 Selection of trials for inclusion in analysis.

Table 1 Significant Predictors of Change in HAM-D$_{17}$ Score and Response Rate Among Patients Treated with Placebo

	HAM-D$_{17}$ score change					HAM-D$_{17}$ response rate				
Predictor	Univariate meta-regression		Multivariate meta-regression		Predictor	Univariate meta-regression		Multivariate meta-regression		
	Standardized β	p-value	Standardized β	p-value		Standardized β	p-value	Standardized β	p-value	
Completion rate, %	3.67	0.0002	1.96	0.062	US sites, %	− 3.09	0.0020	− 1.31	0.207	
Year of study	3.57	0.0004	1.52	0.141	Completion rate, %	3.07	0.0021	0.83	0.423	
Study drug (DVS vs VEN)	2.53	0.0113	− 1.22	0.240	Median baseline HAM-D$_{17}$ score	2.85	0.0043	0.66	0.513	
Assessments per visit	2.47	0.0134	1.27	0.219	Academic sites, %	− 2.06	0.0390	− 1.63	0.119	

Abbreviations: DVS, desvenlafaxine; HAM-D$_{17}$, Hamilton Depression Rating Scale; VEN, venlafaxine.

in a trial inversely predicted the percentage of placebo responders. Greater median baseline HAM-D$_{17}$ score predicted higher rates of placebo response. However, no predictors remained statistically significant in the multivariate meta-regression analyses.
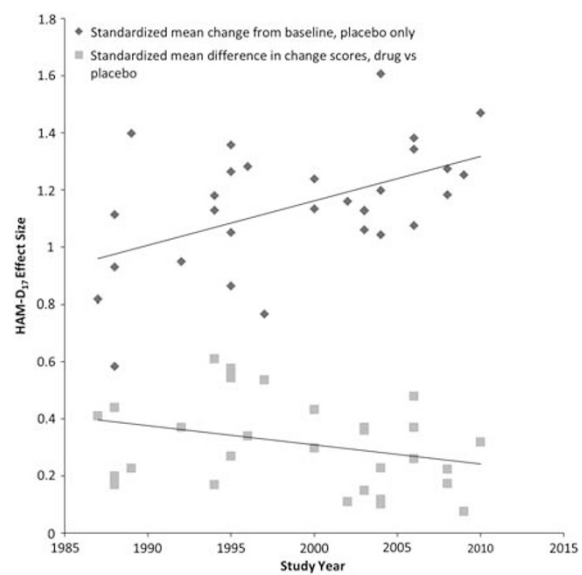
Drug–Placebo Differences

The SMD between antidepressant and placebo on the HAM-D$_{17}$ was 0.31 ± 0.03 ($p < 0.001$), reflecting a HAM-D$_{17}$ score difference of 2.42. Significant heterogeneity was observed again ($Q = 44.04$, $p = 0.0363$). After excluding the study (D3362) responsible, the SMD was 0.33 ± 0.03. Reduction in HAM-D$_{17}$ score with placebo was inversely correlated with the difference in reduction in HAM-D$_{17}$ between antidepressant and placebo ($r = -0.45$, $p = 0.0119$). Figure 2 presents the relationship between the SMD of effect size for placebo and drug–placebo difference over time.

The risk ratio for response between antidepressant and placebo was 1.36 (95% CI 1.28–1.44); the risk difference was 14.5% (95% CI 11.8%–17.1%, $p < 0.0001$). Tests of heterogeneity were not significant for this outcome ($Q = 39.90$, $p = 0.0855$).

Predictors of Antidepressant–Placebo Separation

Results of the predictor analyses of differences in HAM-D$_{17}$ change score and response rate are presented in Table 2. The strongest predictor for both outcomes was the percentage of patients enrolled from academic sites. Other significant positive predictors of both outcomes were study drug (venlafaxine) and number of post-baseline visits. Year of study and completion rate were negative predictors of both outcomes. Median baseline HAM-D$_{17}$ score negatively predicted separation in response rate.

In the multivariate meta-regression analysis of HAM-D$_{17}$ change score, only the percentage of patients enrolled from academic sites maintained statistical significance. For the multivariate meta-regression analysis of differences in response rate, both percentage of patients from academic sites and median baseline HAM-D$_{17}$ score remained significant predictors. Again, higher baseline severity predicted poorer antidepressant–placebo separation on response rates.

Figure 2 Effect sizes of placebo treatment and drug–placebo differences over time.

Predictors of Positive Study Outcomes

Of the 30 studies, 20 demonstrated statistically significant drug–placebo differences on the HAM-D$_{17}$. Four predictors were significantly associated with positive study outcomes: greater percentage of patients from academic sites ($\chi^2 = 9.27$, $p = 0.0023$), lower completion rate ($\chi^2 = 7.00$, $p = 0.0082$), longer trial duration ($\chi^2 = 6.30$, $p = 0.0121$), and lower median baseline HAM-D$_{17}$ ($\chi^2 = 6.26$, $p = 0.0124$).

Sensitivity Analyses

Studies started before 2001 ($N = 16$) all used venlafaxine, whereas desvenlafaxine was the investigational drug in all subsequent studies ($N = 14$). Among the venlafaxine studies, greater reduction in HAM-D$_{17}$ score with placebo was associated with lower age and a greater number of assessments per visit. Higher placebo response rates were significantly associated with lower percentage of patients from US sites and greater completion rates. No factors predicted placebo response among the desvenlafaxine trials.

Table 2 Significant Predictors of Differences in HAM-D$_{17}$ Score and Response Rates in Patients Treated with Venlafaxine or Desvenlafaxine *vs* Placebo

	HAM-D$_{17}$ score change					HAM-D$_{17}$ response rate				
Predictor	Univariate meta-regression		Multivariate meta-regression		Predictor		Univariate meta-regression		Multivariate meta-regression	
	Standardized β	p-value	Standardized β	p-value			Standardized β	p-value	Standardized β	p-value
Academic sites, %	3.74	0.0002	2.24	**0.034**	Academic sites, %		3.71	0.0002	2.26	**0.035**
Study drug (DVS vs VEN)	−3.00	0.0027	−0.89	0.387	Completion rate, %		−3.55	0.0004	0.29	0.781
Completion rate, %	−2.94	0.0033	−0.15	0.884	Study drug (DVS vs VEN)		−2.77	0.0056	0.55	0.591
No. of post-baseline visits	2.72	0.0066	1.32	0.201	Year of study		−2.64	0.0083	−1.01	0.331
Year of study	−2.14	0.0323	0.31	0.760	Median baseline HAM-D$_{17}$ score		−2.22	0.0266	−2.30	**0.030**
White race, %	2.03	0.0429	—	—	No. of post-baseline visits		2.09	0.0363	1.71	0.103

Abbreviations: DVS, desvenlafaxine; HAM-D$_{17}$, 17-item Hamilton Rating Scale for Depression; VEN, venlafaxine.
Bolded values represent statistically significant results.

In the venlafaxine studies, drug–placebo separation on continuous outcomes was positively predicted by a greater number of post-baseline visits and percentage of patients from academic sites. Significant differences in response rates were predicted by percentage of patients from academic sites and negatively correlated with completion rate. These findings were not observed in the desvenlafaxine studies. Here, greater mean age, greater percentage of women, and number of assessments per visit predicted differences in HAM-D$_{17}$ change score. Year of study was not significant in any of the sensitivity analyses.

Relationships Between Predictors

Factors with a moderate positive correlation with study year included completion rate ($r = 0.66$, $p < 0.0001$) and assessments per visit ($r = 0.55$, $p < 0.002$). Moderate negative correlations with study year included mean baseline HAM-D$_{17}$ ($r = -0.57$, $p = 0.001$), percent of patients from academic sites ($r = -0.42$, $p = 0.024$, see Figure 3), and white race ($r = -0.42$, $p = 0.043$). Completion rate was also inversely correlated with percent of patients from academic ($r = -0.53$, $p = 0.003$) and US sites ($r = -0.37$, $p < 0.05$), and number of post-baseline visits ($r = -0.36$, $p < 0.05$).

DISCUSSION

In this meta-analysis of placebo-controlled trials of venlafaxine and desvenlafaxine for treatment of MDD, the most consistent predictor of statistically significant drug–placebo separation was the percentage of patients enrolled from trial sites based in academic institutions. Specifically, a higher proportion of patients from academic sites predicted a lower placebo response rate, greater drug–placebo separation, and a greater likelihood of positive study outcome. We believe this is the first published analysis to document the potentially negative impact of the decreasing role of academic sites in industry-sponsored clinical trials.

We found that the participation of academic sites has declined over the past 20 years. Today, in a research climate
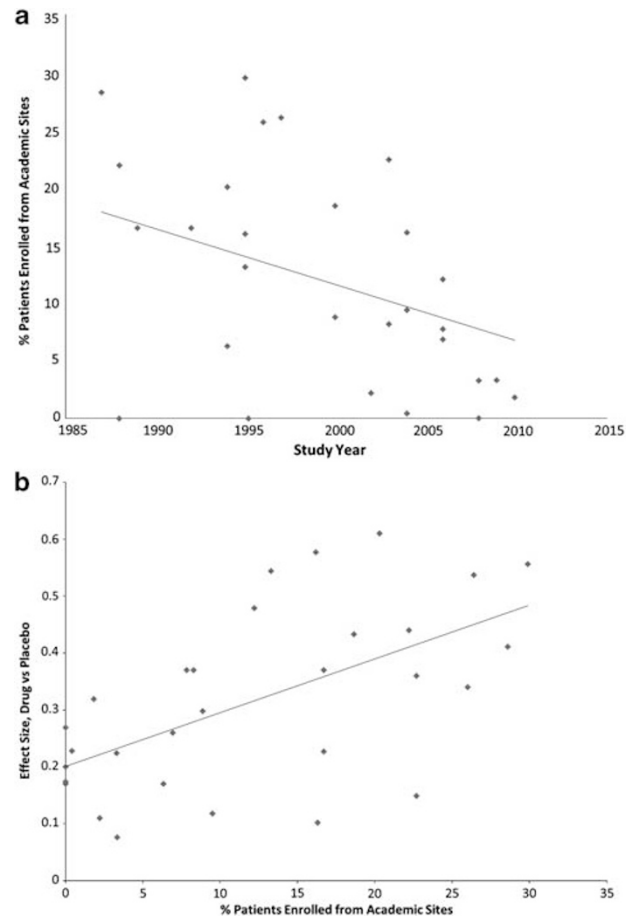
Figure 3 (a) Relationship between the percentage of patients enrolled in a trial from academic sites and the year of study initiation. (b) Relationship between the percentage of patients in a trial enrolled from academic sites and the effect size of drug–placebo difference.

emphasizing rapid recruitment, academic sites contribute only a small fraction of the total enrolled patients. Although our results pertain to studies of MDD conducted by one

sponsor, the strength of these findings suggests that similar analyses should be performed in other therapeutic indications and areas of medicine.

Substantial differences exist between how academic and private sites function and are incentivized that warrant consideration. Contracts with academic sites are made with universities, not investigators. Consequently, academic investigators do not experience personal financial gain from greater enrollment; however, there are motivations to enroll sufficiently to cover costs and generate reserves to support other academic endeavors. Academic sites also may have lower turnover of personnel and more stringent training procedures, producing more experienced trial staff. It is also possible that academic medical centers enroll a greater proportion of patients who are referred by other physicians due to poor treatment response, thereby reducing the likelihood of placebo response. We were unable to determine whether including more academic sites produced effect size estimates that are closer to 'true' treatment effects. Therefore, further examination is warranted of the underlying differences between private and academic sites and the generalizability of the impact of study sites on trial outcomes for other disorders.

Our findings replicate and extend the phenomenon of increasing placebo response over time, which was first identified by Walsh et al (2002). The SMD of 1.15 for placebo treatment in our analysis was larger than the 0.92 SMD reported by Kirsch et al (2008), who analyzed 35 trials of six antidepressants (including venlafaxine) submitted to the FDA. This is not surprising in that, unlike the Kirsch study, about one-third of the studies included in this meta-analysis were conducted after 2000. Nevertheless, the drug–placebo SMD of 0.31 in this study is comparable to that reported by Kirsch et al (2008), suggesting that the results reported here likely extend beyond the two medications studied.

Median baseline depression severity emerged as a second, relatively consistent predictor of placebo response and positive study outcome. In contrast to previous studies (Fournier et al, 2010; Kirsch et al, 2008), we found lower median severity predicted better drug–placebo separation. When coupled with the temporal trend in placebo response, our findings suggest that efforts to reduce placebo response by increasing minimum baseline severity criteria have been counterproductive, perhaps due to inflation of baseline scores by raters (Landin et al, 2000). Investigators planning studies of novel antidepressants in the 21st century would be wise to adopt methods to ensure that pretreatment severity scores are accurate and do not distort or inflate these critical assessments.

We also found that higher completion rates predicted smaller drug–placebo differences and lower likelihood of positive study outcomes. Given that the primary reason for early termination from a trial for patients receiving placebo is inadequate response, higher completion rates likely reflect a more highly placebo-responsive study sample.

Based on our analysis of drug vs placebo differences, one might conclude that venlafaxine appears as a more effective antidepressant than desvenlafaxine. However, we believe this is unlikely because this comparison is completely confounded by year of study and the accompanying decline in participation of academic centers. Specifically,

the studies of venlafaxine were conducted before those of desvenlafaxine, when placebo response rates were lower and the contribution of academic sites was larger. If this confound is replicable in other data sets, caution is warranted in meta-analytic comparisons of older and newer compounds that do not take era of study into account.

These data also suggest that emphasis on the speed of recruitment into clinical trials has been counterproductive. As companies have increasingly turned to commercial sites for trial recruitment, drug–placebo differences have grown smaller, resulting in a ballooning of sample sizes required to find the smaller effect sizes, which in turn leads companies to pursue yet more nonacademic sites to meet recruitment goals. A larger number of sites also is likely to increase problems in quality control, including lowering of the reliability of key clinician-administered outcome measures. Current clinical trial designs suggest companies recognize this problem, as witnessed by the growth of third-party companies employed to oversee the quality of trial sites' work in selecting and evaluating potential clinical trial participants.

Strengths of this analysis are the inclusion of all trials conducted for this indication with these medications, thus eliminating the possibility of publication bias that can reduce the validity of meta-analyses. We were also able to evaluate a large number of variables, and thereby allow for more comprehensive multivariate analyses.

Limitations include the restriction to only two antidepressants studied by a single company. Unique characteristics of these medications, or aspects of trial design and administration by the sponsor, may limit generalizability of our findings. We were also limited by the absence of patient-level outcomes for some studies in this data set, which would have permitted a subject-level meta-analysis and evaluation of moderator effects between the predictors and treatment response.

If replicable, our results suggest that industry should re-engage with academic sites. Several forces interfere with academic sites engaging with industry-sponsored trials, including (1) slow start-up times owing to local Institutional Review Board requirements and the complexities of contract language; (2) growing concerns among academic investigators about apparent conflicts of interest associated with industry involvement; (3) lack of institutional support or recognition of investigators who conduct industry-sponsored trials; and (4) historically slow recruitment rates. Our analysis suggests that improved confidence in study outcomes may well be worth the costs of nurturing greater involvement of academia in clinical trials.

REFERENCES

American Psychiatric Work Group on Major Depressive Disorder (2010). Practice guideline for the treatment of patients with major depressive disorder, 3rd Edition. Available at http://www.psychiatryonline.org/content.aspx?bookid=28&sectionid=1667485.

Fava M, Evins AE, Dorer DJ, Schoenfeld DA (2003). The problem of the placebo response in clinical trials for psychiatric disorders: culprits, possible remedies, and a novel study design approach. *Psychother Psychosom* 72: 115–127.

Fournier JC, DeRubeis RJ, Hollon SD, Dimidjian S, Amsterdam JD, Shelton RC et al (2010). Antidepressant drug effects and depression severity: a patient-level meta-analysis. *JAMA* 303: 47–53.

Frank JD, Frank JB (1991). *Persuasion and Healing: A Comparative Study of Psychotherapy*. Johns Hopkins Press: Baltimore, MD.

Hamilton M (1960). A rating scale for depression. *J Neurol Neurosurg Psychiatry* 23: 56–62.

Higgins JP, Thompson SG (2002). Quantifying heterogeneity in a meta-analysis. *Stat Med* 21: 1539–1558.

Higgins JPT, Green S. (eds) *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration (2011): Available from www.cochrane-handbook.org.

Hrobjartsson A, Gotzsche PC (2001). Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. *N Engl J Med* 344: 1594–1602.

Khan A, Dager SR, Cohen S, Avery DH, Scherzo B, Dunner DL (1991). Chronicity of depressive episode in relation to antidepressant-placebo response. *Neuropsychopharmacology* 4: 125–130.

Khan A, Leventhal RM, Khan SR, Brown WA (2002). Severity of depression and response to antidepressants and placebo: an analysis of the Food and Drug Administration database. *J Clin Psychopharmacol* 22: 40–45.

Khan A, Schwartz K, Kolts RL, Ridgway D, Lineberry C (2007). Relationship between depression severity entry criteria and antidepressant clinical trial outcomes. *Biol Psychiatry* 62: 65–71.

Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT (2008). Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med* 5: e45.

Kobak KA, Brown B, Sharp I, Levy-Mack H, Wells K, Ockun F et al (2009). Sources of unreliability in depression ratings. *J Clin Psychopharmacol* 29: 82–85.

Kraemer HC, Wilson GT, Fairburn CG, Agras WS (2002). Mediators and moderators of treatment effects in randomized clinical trials. *Arch Gen Psychiatry* 59: 877–883.

Landin R, DeBrota DJ, DeVries TA, Potter WZ, Demitrack MA (2000). The impact of restrictive entry criterion during the placebo lead-in period. *Biometrics* 56: 271–278.

National Collaborating Centre for Mental Health (2004). Depression: Management of depression in primary and secondary care. Clinical guideline 23. Available at http://www.nice.org.uk/CG23.

Posternak MA, Zimmerman M (2007). Therapeutic effect of follow-up assessments on antidepressant and placebo response rates in antidepressant efficacy trials: meta-analysis. *Br J Psychiatry* 190: 287–292.

Rettig RA (2000). The industrialization of clinical research. *Health Aff* 19: 129–146.

Rief W, Nestoriuc Y, Weiss S, Welzel E, Barsky AJ, Hofmann SG (2009). Meta-analysis of the placebo response in antidepressant trials. *J Affect Disord* 118: 1–8.

Rutherford BR, Wager TD, Roose SP (2010). Expectancy and the treatment of depression: a review of experimental methodology and effects on patient outcome. *Curr Psychiatry Rev* 6: 1–10.

Sinyor M, Levitt AJ, Cheung AH, Schaffer A, Kiss A, Dowlati Y et al (2010). 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* 71: 270–279.

Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R (2008). Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 358: 252–260.

Walsh BT, Seidman SN, Sysko R, Gould M (2002). Placebo response in studies of major depression: variable, substantial, and growing. *JAMA* 287: 1840–1847.

Supplementary Information accompanies the paper on the Neuropsychopharmacology website (http://www.nature.com/npp)