Therapeutic effect of follow-up assessments on antidepressant and placebo response rates in antidepressant efficacy trials

Meta-analysis

MICHAEL A. POSTERNAK and MARK ZIMMERMAN

Background It remains unclear how much various factors contribute to the placebo response.

Aims To estimate the therapeutic impact of follow-up assessments on placebo response in antidepressant trials.

Method Double-blind, placebo-controlled antidepressant trials that reported weekly changes in Hamilton Rating Scale for Depression (HRSD) scores over 6 weeks were selected. Included studies (n=41) were divided into those that conducted four, five or six follow-up assessments. Reductions in HRSD scores as a function of the different follow-up schedules were compared.

Results An extra follow-up visit at week 3 was associated with a 0.86 further reduction in HRSD score; an extra visit at week 5 was associated with a 0.67 further reduction. These effects represented approximately 34–44% of the placebo response that occurred over these time frames. Two additional visits were associated with twice the reduction in HRSD score than one, suggesting that the therapeutic impact of assessment visits is cumulative and proportional. A comparable therapeutic effect was also found in participants receiving active medication.

Conclusions Follow-up assessments in antidepressant treatment trials incur a significant therapeutic effect for participants on placebo, and this represents about 40% of the placebo response.

Declaration of interest None.

Reports in both scientific journals and the media have questioned whether the true benefits of antidepressant medications have been exaggerated (Goleman, 1995; Fisher & Greenberg, 1997; Horgan, 1998; Kirsch & Sapirstein, 1999), and a recent review of the Food and Drug Administration (FDA) database found that that as many as half of antidepressant trials yield negative results (Khan et al, 2002). A major hindrance to establishing antidepressant efficacy is the remarkably high rates of improvement among participants receiving placebo, which have been increasing over the past two decades (Walsh et al, 2002). Factors that have been implicated in the placebo response include the instillation of hope, response expectancies (Kirsch, 1985), motivation to please investigators (Orne, 1969), the therapeutic impact of assessment contact, rater bias and spontaneous improvement (Harrington, 1999). A better understanding of how much each contributes would allow a more accurate gauge of the true antidepressant effect and could lead to improved trial designs.

In the present study, we sought to evaluate the therapeutic impact of frequent follow-up assessments. In standard antidepressant trials, participants are usually seen on a weekly basis to assess depression severity, level of functioning and side-effects. Such visits typically last 30 min or more and are conducted by trained research assistants over the course of 6 weeks. The impact of so much contact with a healthcare provider is unknown but could be substantial. Furthermore, this amount of contact is much greater than in routine clinical practice where two to three 15-min visits for management of medication are the norm (Posternak et al, 2002a). To evaluate the impact of these follow-up assessments, we conducted a meta-analysis of 41 double-blind, placebo-controlled antidepressant trials published over the past two decades. We primarily focused on the impact that follow-up assessments had on the placebo response but also examined their effect on participants receiving active medication.

METHOD

Sources of data and criteria for review

The collection of studies used here is the same as in our previous meta-analysis which evaluated the time course of improvement on antidepressant medication and placebo (Posternak & Zimmerman, 2005). These studies were compiled by reviewing the bibliography of the meta-analysis evaluating placebo response rates in antidepressant trials published over the past two decades (Walsh et al, 2002). To augment this database, we also systematically reviewed each article published from January 1992 through December 2001 in six psychiatric journals (American Journal of Psychiatry, Archives of General Psychiatry, British Journal of Psychiatry, Journal of Clinical Psychiatry, Journal of Clinical Psychopharmacology and Psychopharmacology Bulletin).

Studies were included if they: (a) were in English; (b) were published from January 1981 through December 2001; (c) were primarily composed of out-patients with major depressive disorder according to Research Diagnostic Criteria (RDC; Spitzer et al, 1978); (d) had at least 20 participants in the placebo group; (e) randomly assigned participants to receive a putative antidepressant drug or drugs and placebo; (f) reported the total number of participants assigned to placebo and medication group(s); (g) assessed participants under double-blind conditions; and (h) utilised the Hamilton Rating Scale for Depression (HRSD; Hamilton, 1960) to assess improvement. We excluded studies that did not report mean baseline HRSD scores, did not present weekly or biweekly (every other week) changes in HRSD scores, evaluated agents with unproven antidepressant properties or evaluated accepted antidepressant agents that were used at subtherapeutic doses, or focused on specific subpopulations of patients such as the elderly. Forty-seven trials that met these inclusion criteria were included in our original meta-analysis. Of these, we included six studies (Claghorn et al, 1983; Dominguez et al, 1985; Hormazabal et al, 1985; Amsterdam et al, 1986; Ferguson et al, 1994; Khan, 1995) for the present
met-analysis because they did not conduct outcome assessments at week 6.

**Follow-up schedules**

For the 41 studies included in the present meta-analysis, three types of follow-up schedules were used: 15 studies (Cohn & Wilcox, 1985; Byerley et al, 1988; Cohn et al, 1989; Lineberry et al, 1990; Reinherr et al, 1990; Smith et al, 1990; Fontaine et al, 1994; Heiligenstein et al, 1994; Wilcox et al, 1994; Brenner, 1995; Claghorn & Lesem, 1995; Fabre et al, 1995; Mendels et al, 1995; Claghorn et al, 1996; Schatzberg, 2000) conducted weekly follow-up assessments over the course of 6 weeks (weekly cohort); 19 studies (Feighner & Boyer, 1989; Versiani et al, 1989; Gelenberg et al, 1990; Claghorn et al, 1992; Cohn & Wilcox, 1992; Fabre, 1992; Kie, 1992; Rickels et al, 1992; Shrivastava et al, 1992; Smith & Glaudin, 1992; Mendels et al, 1993; Cunningham et al, 1994; Cunningham, 1997; Thase, 1997; Khan et al, 1998; Rudolph et al, 1998; Rudolph & Feiger, 1999; Silverstone & Ravindran, 1999; Stahl, 2000) conducted assessments at weeks 1, 2, 3, 4 and 6 without an assessment at week 5 (skip week 5 cohort); 7 studies (Feighner et al, 1983; Merideth & Feighner, 1983; Rickels et al, 1985; Mendels & Schless, 1986; Rickels et al, 1991; Anonymous, 1994; Laakman et al, 1995) conducted assessments at weeks 1, 2, 4 and 6 without assessments at weeks 3 and 5 (skip weeks 3 and 5 cohort). We utilised these differences in follow-up schedules as a way to focus on the specific therapeutic effects of follow-up assessments.

**Establishing reduction in HRSD scores**

The method for establishing mean baseline scores and weekly improvement in HRSD scores is the same as in our previous meta-analysis (Posternak & Zimmerman, 2005). Baseline HRSD scores and weekly reductions in HRSD scores were established for each study, and all analyses accounted for differences in sample size between studies. Some studies depicted changes in HRSD scores graphically. In these instances, weekly changes in HRSD scores were obtained by measuring each data-point with rounding to the nearest 0.5. A research assistant who was unaware of the purposes of the study remeasured each data-point. Of the 476 data-points extracted from graphs, 456 (95.8%) were remeasured by the research assistant within 0.5 points, suggesting that data extraction was performed reliably and without bias.

**Hypotheses**

We hypothesised that follow-up assessments would have a discernible therapeutic effect on placebo response rates. Differences in follow-up schedules allowed us to compare reductions in HRSD scores in cohorts that met on a weekly basis with those that by design skipped 1 or 2 weeks. Our specific hypotheses were: (a) reductions in HRSD scores from week 4 to week 6 will be greater for the weekly cohort compared with the skip week 5 and skip weeks 3 and 5 cohort; (b) reductions in HRSD scores from week 2 to week 4 will be greater for the weekly cohort and the skip week 5 cohort compared with the skip weeks 3 and 5 cohort; (c) there will be a proportional and cumulative therapeutic effect of having multiple extra assessments; to examine this question, we compared reductions in HRSD scores from week 2 to week 6 in the skip weeks 3 and 5 cohort, skip week 5 cohort, and the weekly cohort; (d) to confirm that placebo effects do not differ between cohorts, we predicted that reductions in HRSD scores would be comparable between cohorts from baseline through week 2; because we considered this the most direct method to confirm that there are no random differences in placebo response rates, we deemed it unnecessary to control for potential confounding variables such as fixed vs. flexible dose design, year of publication, etc.; (e) if follow-up assessments are found to convey a therapeutic effect for participants receiving placebo, we would predict that all of the above findings would be replicated in cohorts receiving antidepressant medication.

Finally, if follow-up assessments convey a non-specific therapeutic effect, we hypothesised that treatment effect sizes would be greater in trials with fewer follow-up assessments. However, only a handful of studies published weekly or end-point standard deviations. Therefore, we were unable to establish effect sizes or confidence intervals.

**RESULTS**

**Cohorts**

For participants randomised to placebo, the weekly cohort comprised 941 people from 15 separate studies; the skip week 5 cohort comprised 1449 people drawn from 19 studies and the skip weeks 3 and 5 cohort comprised 673 participants drawn from 7 studies. The baseline mean HRSD scores for these three groups were 25.6 (s.d.=1.78), 25.9 (s.d.=1.47) and 24.3 (s.d.=2.53) respectively.

For participants randomised to active medication, the weekly cohort comprised 1507 people from 25 cohorts (some studies included more than one active medication group); the skip week 5 cohort comprised 2284 people from 31 cohorts and the skip weeks 3 and 5 cohort comprised 820 participants from 9 cohorts. The baseline HRSD scores for these three groups were 25.6 (s.d.=1.82), 25.9 (s.d.=1.49) and 25.0 (s.d.=2.42) respectively.

**Week 5 assessment**

From week 4 to week 6, the mean decrease in HRSD scores for cohorts receiving placebo that met at week 5 (the weekly cohort) was 1.52 points. For cohorts that did not meet at week 5 (the skip week 5 and the skip weeks 3 and 5 cohorts), the mean decrease in HRSD scores from week 4 to week 6 was 0.85 points. Thus, participants who returned for an extra follow-up visit at week 5 experienced a 0.67 greater reduction in HRSD scores over this 2-week period than those who did not have a week 5 visit. This difference represents 44% of the decrease in HRSD scores over this period.

**Week 3 assessment**

From week 2 to week 4, the mean decrease in HRSD scores for cohorts receiving placebo that met at week 3 (the weekly cohort and skip week 5 cohort) was 2.56 points. For cohorts that did not have a scheduled follow-up assessment at week 3 (the skip weeks 3 and 5 cohort), the mean decrease in HRSD scores from week 2 to week 4 was 1.70 points. Thus, participants who returned for an extra follow-up visit at week 3 experienced a 0.86 greater reduction in HRSD scores over this 2-week period than those who did not have a week 3 follow-up visit. This represents 34% of the decrease in HRSD scores over this period.

**Therapeutic impact of multiple extra assessments**

To examine whether there is a cumulative and proportional therapeutic impact of multiple extra assessments, we compared
reductions in HRSD scores from week 2 to week 6 in the weekly cohort with reductions in the skip week 5 and skip weeks 3 and 5 cohorts. The first group had four scheduled follow-up assessments, the second group had three and the third group had two. Reductions in HRSD scores were 4.24, 3.33 and 2.49 points respectively. Thus, the reduction with one extra assessment (skip weeks 3 and 5 cohort) was 0.84 HRSD points whereas with two extra assessments (skip weeks 3 and 5 cohort v. weekly cohort) was 1.73 HRSD points. This suggests that the therapeutic impact of follow-up assessments is cumulative and proportional.

Control analysis
To evaluate whether placebo effects are otherwise comparable between the cohorts of interest, we compared reductions in HRSD scores from baseline to week 2 between the weekly cohort and the skip week 5 and skip weeks 3 and 5 cohorts. Because all three cohorts received weekly follow-up assessments through week 2, we predicted that reductions in HRSD scores would be similar. The reduction in HRSD scores from baseline to week 2 in the weekly cohort was 5.35 points. In the two cohorts that subsequently skipped one or two follow-up assessments, the reduction in HRSD scores was 5.41 points. Thus, placebo effects were comparable between the cohorts when the frequency of follow-up visits was the same.

Participants receiving active medication
We repeated all the analyses described above for participants receiving active medication. Reduction in HRSD score from week 4 to week 6 for the weekly cohort was 2.35 points compared with 1.38 for cohorts who did not have a week 5 visit (a difference of 0.97 points). Reduction in HRSD score from week 2 to week 4 for cohorts that met at week 3 (the weekly cohort and the skip week 5 cohort) was 3.69 points compared with 2.57 for cohorts that did not have a week 3 visit (a difference of 1.12 points). Reductions in HRSD scores from week 2 to week 6 for the weekly cohort, skip week 5 cohort and skip weeks 3 and 5 cohort were 5.87, 5.05 and 4.29 respectively. One extra assessment visit therefore accounted for a reduction of 0.76 HRSD points whereas a second extra assessment accounted for an additional 0.82 points. For the control analysis, we again compared reductions in HRSD scores from baseline to week 2 in the weekly cohort with the two cohorts that skipped at least one follow-up assessment. Reductions in HRSD scores were 7.78 and 7.61 HRSD points respectively, again suggesting comparable treatment effects except when there were differences in follow-up schedules.

DISCUSSION
The ubiquitous and robust placebo response has for years both intrigued and frustrated mood disorder researchers. Although there is general consensus as to which factors are responsible for the placebo response, it remains unclear how much each particular component contributes to the overall effect. One exception to this is the role that spontaneous improvement may play. In a meta-analysis comparing treatment effect sizes for people with depression randomised to placebo with those randomised to no treatment, spontaneous improvement was estimated to constitute about one-third of the placebo response (Kirsch & Sapirstein, 1999). Other investigators have provided independent confirmation of this estimate (Posternak & Zimmerman, 2001; Posternak et al, 2006).

Main results
In the present study, we isolated one of the remaining components – the therapeutic impact of follow-up assessments – to determine the importance of this factor to the remaining two-thirds of the placebo response. We found that scheduling an extra follow-up visit at week 3 was associated with an additional 0.86-point reduction in HRSD scores, whereas scheduling an additional week 5 visit was associated with an additional 0.67 reduction in HRSD scores. These reductions represent approximately 40% of the placebo response that occurred over their respective time frames. When we examined the cumulative effect of scheduling two additional follow-up visits, we found that the therapeutic impact of each visit was cumulative and proportional. Thus, one extra visit was associated with a 0.84 greater reduction in the HRSD score whereas a second extra visit was associated with a 0.91 further reduction in the HRSD score. As further illustration of the impact of follow-up assessments on the placebo response, participants who were assessed on a weekly basis experienced an overall drop in HRSD scores of 9.6 points over the course of 6 weeks. By comparison, participants receiving placebo who were assessed only four times experienced only a 7.3-point drop in HRSD score.

Since follow-up assessments had a discernible therapeutic effect for participants receiving placebo, we expected they would also have a discernible and comparable effect for those receiving active medication. Indeed, each of our analyses from the placebo cohorts was replicated for cohorts receiving active medication, as each additional follow-up visit was associated with a further reduction of 0.97–1.12 in HRSD scores.

Design of meta-analysis
The ideal method for evaluating the therapeutic impact of follow-up assessments on the placebo response would be to randomise participants with depression receiving placebo to different follow-up schedules. Such a study has not been performed to date and most likely never will. In the present meta-analysis, we have in effect randomised cohorts rather than individuals. Since the methodology of efficacy trials of antidepressants has remained largely unchanged over the years (Thase, 1999), heterogeneity between studies is likely to be minimal: all studies involved out-patients with moderate-to-severe depression who received identical treatment (placebo) over the course of 6 weeks using the same outcome measure (the HRSD). Where an extra follow-up assessment was conducted, a clear therapeutic effect was associated with that visit as hypothesised. Although it is possible that this could be attributable to random differences between studies, we would argue that this is extremely unlikely. The present meta-analysis included the majority of acute-phase, placebo-controlled antidepressant trials published over the past two decades, and our analyses were therefore based on large sample sizes. Second, improvement on placebo was comparable between all three cohorts during the first 2 weeks of treatment when follow-up assessment schedules were identical. As this is the most direct method for evaluating random differences in placebo response rates, it would be superfluous to attempt to control for other potential confounding variables such as year of publication, episode duration,
comorbidity, etc. Furthermore, all of our findings that supported a clear, therapeutic effect from assessment contact were replicated in cohorts receiving active medication.

We would argue that our results are not undermined by relying solely on published studies. Publication bias is a concern for many meta-analyses because negative trials often go unpublished, and attempts to establish effect sizes may consequently overestimate treatment benefits. The goal of the present study, however, was to estimate the therapeutic impact of follow-up assessments. The lack of inclusion of unpublished studies would only undermine our results if unpublished studies were found to systematically have less therapeutic impact of their assessment visits (for example, if raters in unpublished studies were consistently less empathic). Unpublished studies, however, by virtue of having failed to separate drug from placebo, would be expected to have more rather than less robust placebo response rates, and the therapeutic impact of follow-up assessments might, if anything, be more pronounced.

**Design of trials**

Considering the relatively modest effect size of FDA-approved antidepressants over placebo, that side-effects may unmask raters in favour of eliciting drug-placebo differences (Greenberg et al., 1992) and that most negative trials never get published, several investigators have suggested that the benefits of antidepressant medications have been exaggerated over the years (Fisher & Greenberg, 1997; Kirsch & Sapirstein, 1999). Although these arguments are persuasive, we believe an alternative explanation also exists – that the methodology used to elicit and establish antidepressant efficacy is inefficient. As reviewed elsewhere (Posternak et al., 2002b), the methodology used in antidepressant trials evolved largely from traditions established over three decades ago and has never undergone empirical testing. Our results suggest that the frequent and extensive monitoring that occurs in clinical trials confers a significant therapeutic effect for participants receiving placebo (and active medication). High placebo response rates reduce treatment effect sizes and increase the risk that an efficacious agent will be deemed ineffective. Although a comparable therapeutic effect from follow-up visits was found in participants randomised to active medication, reducing an equivalent amount of ‘noise’ in both cohorts would have the effect of increasing the power to detect differences between the active medication and control group (Cohen, 1988).

Knowing the impact that follow-up assessments have on placebo response rates, the design of antidepressant trials could be modified either by reducing the amount of time devoted to assessing participants in follow-up, reducing the frequency of follow-up assessments, or relying more on off-site raters or interactive computer assessment. Of course, consideration of these changes must be balanced against ethical concerns of having insufficient monitoring over the course of a clinical trial. This would apply both to participants randomised to placebo and to those receiving a putative antidepressant agent, especially if there are concerns regarding the potential for increased suicidal ideation following the initiation of an antidepressant.

**Explaining the placebo response**

Our results suggest that the follow-up assessment schedules of standard antidepressant efficacy trials convey a significant therapeutic effect for participants receiving placebo, and that these assessment visits account for an estimated 40% of the placebo response. This does not take into account the therapeutic effect of the initial evaluation, which is typically much more extensive than follow-up assessments and would be expected to convey a larger therapeutic effect. For years, there has been much speculation as to which ingredients comprise the powerful and seemingly magical placebo pill, with some investigators even suggesting that different coloured pills may be associated with different placebo response rates (Jacobs & Nordan, 1979; Buckalew & Coffield, 1982). Our findings suggest that, after accounting for spontaneous improvement, the placebo response in trials of antidepressants stems largely from the attention and care received during the course of the clinical trial.

**REFERENCES**


A fixed-dose (300 mg) efficacy study of mirtazapine in major depression. (suppl. 2), S119–S124.


Therapeutic effect of follow-up assessments on antidepressant and placebo response rates in antidepressant efficacy trials: Meta-analysis
MICHAEL A. POSTERNAK and MARK ZIMMERMAN
Access the most recent version at DOI: 10.1192/bjp.bp.106.028555

References
This article cites 61 articles, 2 of which you can access for free at:
http://bjp.rcpsych.org/content/190/4/287#BIBL

Reprints/permissions
To obtain reprints or permission to reproduce material from this paper, please write to permissions@rcpsych.ac.uk

You can respond to this article at
/letters/submit/bjprcpsych;190/4/287

Downloaded from
http://bjp.rcpsych.org/ on June 17, 2017
Published by The Royal College of Psychiatrists