International Journal of Neuropsychopharmacology (2012), **15**, 289–296. © CINP 2011
doi:10.1017/S1461145711000666

ARTICLE

CINP

Does the frequency of follow-up assessments affect clinical trial outcome? A meta-analysis and meta-regression of placebo-controlled randomized trials

Nadia Iovieno[1], Enrico Tedeschini[1,2], Yeciel Levkovitz[1,3], Victoria E. Ameral[1] and George I. Papakostas[1]

[1] *Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA*
[2] *University of Modena and Reggio Emilia, Italy*
[3] *The Emotion-Cognition Research Center, Shalvata Mental Health Center, Hod and the Sackler Faculty of Medicine, Tel-Aviv University, Israel*

Abstract

The number and temporal distribution of follow-up assessments during a clinical trial is a critical factor which may influence the outcome as well as the overall cost of a trial. Therefore, we aimed to examine whether the overall and differential frequency of study observations during the course of a clinical trial affects the risk ratio (RR) of responding to antidepressants *vs.* placebo, specifically in trials for major depressive disorder (MDD). Medline/Pubmed publication databases were searched for randomized, double-blind, placebo-controlled trials of antidepressants for adults with MDD (1 January 1980–11 May 2010). A total of 142 manuscripts involving 256 drug–placebo comparison were pooled ($n=38\,860$). We found that higher overall frequency (OF, frequency of assessments during the entire trial) and higher late frequency (LF, frequency of assessments after the first 3 wk of the trial), but not higher early frequency (EF, frequency of assessments during the first 3 wk of the trial), of follow-up visits predicted a significantly greater RR of responding to antidepressant *vs.* placebo (coefficient = 0.213, $p=0.014$; coefficient = 0.238, $p=0.003$; and coefficient = 0.021, $p=0.755$, respectively, for OF, LF and EF). None of the measures of frequency examined (OF, EF, LF) significantly predicted the RR of discontinuing antidepressant *vs.* placebo. These findings suggest that increasing the number of follow-up visits, specifically after the third week rather than within the first 3 wk of the trial, may be an effective approach to improve the likelihood of success in placebo-controlled clinical trials for MDD.

Received 9 July 2010; Reviewed 15 December 2010; Revised 16 February 2011; Accepted 28 March 2011; First published online 4 May 2011

Key words: Antidepressants, assessments, clinical trials, major depressive disorder.

Introduction

Major depressive disorder (MDD) is a highly prevalent illness which is often associated with increased morbidity and mortality. Antidepressant medications, along with certain forms of psychotherapy, represent the mainstay of treatment for MDD. Double-blind, randomized, placebo-controlled clinical trials are considered the 'gold standard' for the development of novel antidepressant therapies. However, even for compounds that have been repeatedly proven to be efficacious in treating this illness, statistically significant differences in efficacy *vs.* placebo are not always apparent throughout all clinical trials conducted. Such 'failed' or 'negative' trials may, in turn, lead to delays in bringing new treatments to the clinic, as well as increased costs for the development of new treatments (Fava *et al.* 2003).

For this reason, over the last two decades, a number of researchers have investigated the relationship between various elements of clinical trial design and the likelihood of obtaining a 'positive' result in MDD studies, including the severity of depression at baseline (Papakostas & Fava, 2009; Stein *et al.* 2006), the choice of primary outcome measure (Carmody *et al.*

Address for correspondence: E. Tedeschini, M.D., Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, 1 Bowdoin Square – 6th Floor, Boston, Massachusetts 02114.
Tel.: (617) 724-3678 *Fax*: (617) 724-3028
Email: etedeschini@partners.org

2006; Faries *et al.* 2000), the presence and duration of the placebo lead-in period (Faries *et al.* 2001; Lee *et al.* 2004; Trivedi & Rush, 1994), the effect of concomitant medications administered during the study (Wernicke *et al.* 1997), the relationship between the probability of receiving placebo and the attrition rates (Tedeschini *et al.* 2010), and the duration of the trial (Tedeschini *et al.* 2011). However, the impact of the number and frequency of follow-up assessments on the clinical trial outcome has not yet been sufficiently studied. This is, indeed, a critical issue which might affect the chance of success as well as the cost of the clinical trial (Gelenberg *et al.* 2008). In a recent study, Posternak *et al.* (2007) evaluated the therapeutic impact of the number of follow-up assessments on placebo and antidepressant response rates in MDD trials. The authors found that more frequent assessments during the trial were associated with a greater reduction in depression severity scores, both in antidepressant- and placebo-treated patients.

However, antidepressant and placebo response rates are not uniform during the duration of a clinical trial, but rather follow an inverse U-shaped distribution. Therefore, it would be interesting to also examine whether the positive correlation between the frequency of follow-up assessments and the magnitude of reduction in depression severity previously reported by Posternak & Zimmerman (2007) varies depending on the temporal distribution of assessments during the course of a trial. Specifically, we hypothesize that a greater number of assessments later on in the trial (i.e. after week 3), when both antidepressant and placebo response rates are less pronounced (Posternak & Zimmerman, 2005), would have a grater impact on study effect size than the number of assessments during the first weeks of the trial. Therefore, the purpose of the present work is to examine whether the differential (early *vs.* late) frequency of study observations influences the risk ratio (RR) of responding to antidepressants *vs.* placebo (study effect size) in clinical trials on MDD. Secondary objectives of the study were (1) to replicate the findings by Posternak & Zimmerman (2007), and, (2) to examine the relationship between follow-up visit frequency and study attrition.

Methods

Data sources and search strategy

We sought to identify double-blind, randomized, placebo-controlled trials of antidepressants used as monotherapy for the treatment of MDD for possible inclusion in the meta-analysis. As antidepressants, we defined pharmacological agents which have or had, at one point, received a letter of approval by either the USA, Canadian, Japanese, Australian or EU drug regulatory agencies for the treatment of MDD. According to this definition, the following pharmacological agents met criteria to be considered as 'antidepressants': amitriptyline, nortriptyline, imipramine, desipramine, clomipramine, trimipramine, protriptyline, dothiepin, doxepin, lofepramine, amoxapine, maprotiline, amineptine, nomifensine, bupropion, phenelzine, tranylcypromine, isocarboxazid, moclobemide, brofaromine, fluoxetine, sertraline, paroxetine, citalopram, escitalopram, fluvoxamine, zimelidine, tianeptine, ritanserin, trazodone, nefazodone, agomelatine, venlafaxine, desvenlafaxine, duloxetine, viloxazine, milnacipran, reboxetine, mirtazapine, mianserin.

Elgible studies were first identified using searches of Pubmed/Medline, by cross-referencing the search term 'placebo' with each of the above-mentioned agents. The Pubmed/Medline search was limited to articles that were published between 1 January 1980 and 11 May 2010 (inclusive). 1980 was used as a cut-off in our search in order to decrease diagnostic variability, since DSM-III was introduced in 1980. In order to expand our database, we then reviewed the reference list of all studies identified with Pubmed/Medline. Final inclusion of articles was determined by consensus between the authors.

Study selection

We selected for randomized, double-blind, placebo-controlled trials of antidepressants used as monotherapy for the acute-phase treatment of MDD. We then selected for studies that also met all of the following criteria:

(*a*) Defined MDD according to either DSM-III criteria (APA Task Force on Nomenclature and Statistics, APA Committee on Nomenclature and Statistics, 1980), DSM-III-R criteria (APA Work Group to revise DSM-III, 1987), DSM-IV criteria (APA Task Force on DSM-IV, 1994), reseach diagnostic criteria (Spitzer *et al.* 1978), or Feighner's diagnostic criteria (Feighner *et al.* 1972).
(*b*) Were of, at least, 4 wk in duration.
(*c*) Were of, at most, 12 wk in duration.
(*d*) Focused on the use of antidepressants in their oral formulation.
(*e*) Presented entirely original (not previously published) data.
(*f*) Focused on the treatment of adult patients.

(*g*) Did not exclusively focus on the treatment of elderly patients, patients with treatment-resistant depression, or patients with other depressive disorders, including bipolar disorder, depression with psychotic features, dysthymic disorder, neurotic depression, or minor depression.
(*h*) Did not exclusively focus on the treatment of MDD in patients with comorbid alcohol or substance use disorders, or patients with a specific comorbid medical illness.
(*i*) Involved the use of either the Hamilton Depression Rating Scale (HAMD; Hamilton, 1960), the Montgomery–Asberg Depression Rating Scale (MADRS; Montgomery & Asberg, 1979), or the Clinical Global Impression – Improvement Scale (CGI-I; Guy *et al.* 1976) as one of their outcome measures.
(*j*) Reported the following outcome of interest: number and schedule of follow-up visits.

Definitions

Clinical response was defined as a $\geqslant 50\%$ reduction in HAMD or MADRS scores, baseline to endpoint, or a CGI-I score of <3 at the final visit. For consistency, the HAMD was chosen over the MADRS or CGI when response rates from multiple scales were reported. For studies that only reported CGI-based response rates, the HAMD-based response rates were either obtained from the sponsor or imputed using the method of Walsh *et al.* (2002). Discontinuation rate was defined as per each protocol. For consistency, we used intent-to-treat (ITT)-based response rates in the present analysis. Whenever ITT-based response rates were not available in the publication, the sponsor was contacted to obtain ITT-based response rates. In cases where the sponsor could not retrieve ITT-based response rates, we utilized response rates based on completers. The probability of receiving placebo was computed from the number of treatment arms and the randomization schedule (i.e. 1:1:1) of each trial. For example, a two-arm trial with a 2:1 randomization favouring antidepressant treatment yields a 1 in 3 chance of receiving placebo.

Quantitative data synthesis

The following measures were calculated utilizing the number and timing of follow-up visits and the duration (number of weeks) of each trial:

- Overall frequency (OF) defined as the number of follow-up visits during the trial divided by the duration of the trial in weeks.
- Early frequency (EF) defined as the number of follow-up visits in the first 3 wk divided by 3.
- Late frequency (LF) defined as the number of follow-up visits after the first 3 wk divided by the duration of the trial in weeks minus 3.

The study analyses were conducted as follows:

- First, a random-effects meta-analysis was utilized to estimate the pooled RR of responding to antidepressants *vs.* placebo in all trials.
- Second, three meta-regressions were performed with the RR of responding to antidepressants *vs.* placebo as a dependent variable and OF, EF and LF as the independent variables. For each meta-regression, year of publication, severity at baseline and the probability of being randomized to placebo were also entered as covariates since they had also previously been found to influence the RR of clinical response to antidepressant *vs.* placebo therapy (Papakostas & Fava, 2009).
- Third, three meta-regressions were performed with the RR of discontinuing antidepressants *vs.* placebo as a dependent variable and OF, EF and LF as the independent variables. For each meta-regression only study duration was entered as covariate since no other variable had previously been found to influence the RR of discontinuing antidepressants *vs.* placebo (Tedeschini *et al.* 2010).
- Finally, the analyses were repeated in trials of 6-wk or 8-wk duration.

All tests conducted were two-tailed, with alpha set at the 0.05 level.

Results

Initially 7337 abstracts were identified in Pubmed/Medline. Of these, 6907 were excluded (they were either reports that addressed other topics, reviews, or not RCTs of antidepressants). Abstracts for the remaining 430 clinical trials of antidepressants in MDD were obtained, and reviewed thoroughly. Fifteen additional articles were identified after reviewing the reference lists of these 430 manuscripts as well as two large meta-analyses. Of the 445 potential trials, 303 were excluded for the reasons listed (Fig. 1).

Thus, a total of 142 manuscripts were found eligible for inclusion in our pooled analysis (list available upon request). While 138 of these manuscripts reported the results of a single trial, four reported results of several (a total of eight) trials. Therefore, a total of 256 antidepressant *vs.* placebo contrasts from 146 clinical trials were pooled [$n=38\,860$ patients randomized to an antidepressant ($n=24\,911$) *vs.* placebo

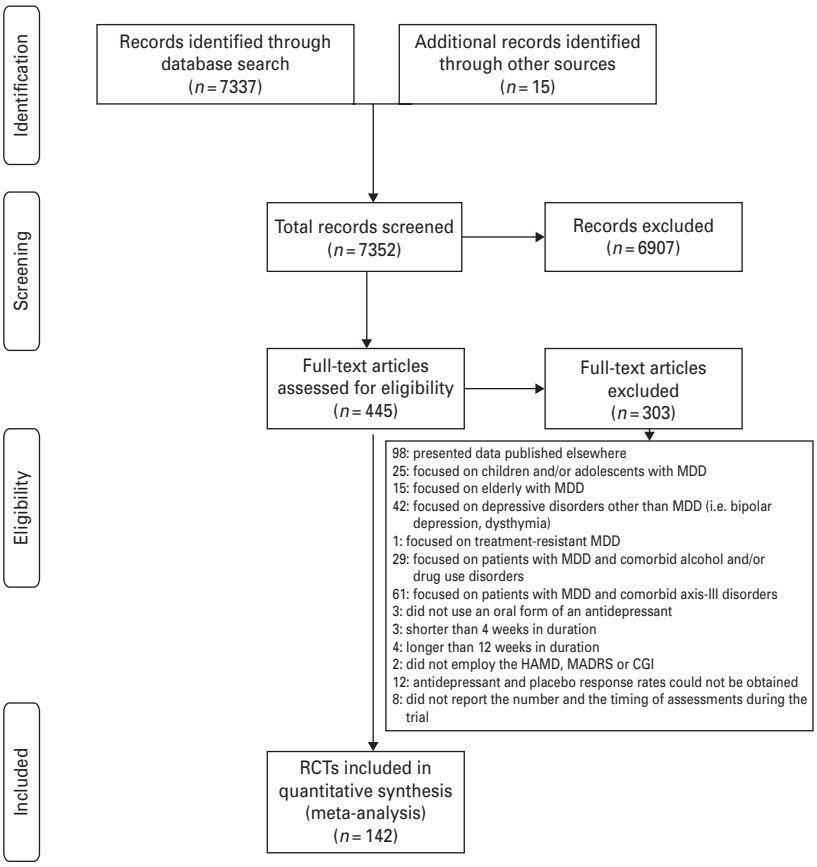

Fig. 1. Flow diagram: trial identification and selection process.

($n = 13\,949$)], 202 of which were derived from the 110 clinical trials lasting either 6 wk or 8 wk (75.5%) [$n = 31\,798$ patients randomized to treatment with an antidepressant ($n = 20\,623$) *vs.* placebo ($n = 11\,175$)]. Specific description of the trials is reported in Table 1.

Meta-analysis and meta-regression results (*studies of 4- to 12-wk duration*)

The result of the random-effects meta-analysis indicated that antidepressant therapy resulted in statistically significant higher response rates than placebo (RR 1.387, 95% CI 1.348–1.420, $p < 0.0001$), and there was no evidence for statistically significant heterogeneity across the studies ($Q = 290.971$, d.f. = 255, $p = 0.060$).

Meta-regression analysis suggested that both a higher OF and a higher LF of follow-up visits during the trial predicted a significantly greater RR of responding to antidepressant *vs.* placebo (coefficient = 0.213, $p = 0.014$; coefficient = 0.238, $p = 0.003$, respectively). The frequency of assessments during the first 3 wk (EF), however, did not predict a significant difference in the RR of responding to antidepressant *vs.* placebo (coefficient = 0.021, $p = 0.755$). Finally, neither OF (coefficient = −0.003, $p = 0.982$), EF (coefficient = −0.032, $p = 0.740$) nor LF (coefficient = 0.056, $p = 0.637$) significantly predicted the RR of discontinuing antidepressant *vs.* placebo.

Meta-analysis and meta-regression results (*studies of 6-wk or 8-wk duration*)

The result of the random-effects meta-analysis indicated that antidepressant therapy had statistically significant higher response rates than placebo (RR 1.390, 95% CI 1.347–1.435, $p < 0.0001$) and there was no evidence for statistically significant heterogeneity ($Q = 231.419$, d.f. = 201, $p = 0.069$).

Meta-regression analysis suggested that both a higher OF and a higher LF of follow-up visits during the trial predicted a significantly greater RR of responding to antidepressant *vs.* placebo (coefficient = 0.207, $p = 0.047$; coefficient = 0.258, $p = 0.008$, respectively). The frequency of assessments during the first 3 wk (EF), however, did not predict a significant difference in the RR of responding to antidepressant *vs.* placebo (coefficient = 0.042, $p = 0.571$).

Table 1. Characteristics of the sample ($N = 146$ trials)

Variable	
Year of publication, mean ± S.D.	1996 ± 8.3
Probability of receiving placebo, mean % ± S.D.	35.0 % ± 9.2
Severity (HAMD-17 score), mean ± S.D.	21.8 ± 3.8
Sample size per treatment arm, mean ± S.D.	93.4 ± 55.6
Duration, mean ± S.D.	6.7 ± 1.8
4 wk, *n* (%)	22 (15.07)
5 wk, *n* (%)	2 (1.37)
6 wk, *n* (%)	60 (41.1)
7 wk, *n* (%)	1 (0.7)
8 wk, *n* (%)	50 (34.24)
9 wk, *n* (%)	3 (2.01)
10 wk, *n* (%)	2 (1.4)
11 wk, *n* (%)	0 (0)
12 wk, *n* (%)	6 (4.11)
Assessments in all trials	
Overall frequency[a], mean ± S.D. (range)	0.82 ± 0.19 (0.38–1.5)
Early frequency[b], mean ± S.D. (range)	0.90 ± 0.23 (0.33–1.67)
Late frequency[c], mean ± S.D. (range)	0.78 ± 0.2 (0.4–1.0)
Assessments in trials lasting either 6 wk or 8 wk ($N = 110$ trials)	
Overall frequency[a], mean ± S.D. (range)	0.81 ± 0.17 (0.38–1.33)
Early frequency[b], mean ± S.D. (range)	0.88 ± 0.22 (0.33–1.67)
Late frequency[c], mean ± S.D. (range)	0.76 ± 0.18 (0.4–1.0)

[a] Overall frequency: number of follow-up visits/duration of the trial in weeks.

[b] Early frequency: number of follow-up visits in the first 3 wk/3.

[c] Late frequency: number of follow-up visits after the first 3 wk/duration of the trial in weeks minus 3.

In order to allow for a graphic depiction of these relationships, we conducted a meta-analysis of the RR of responding to antidepressants *vs.* placebo in studies of 6-wk duration (Fig. 2), and 8-wk duration (Fig. 3) for common variations in LF study design encountered across trials pooled.

Finally, neither OF (coefficient = 0.127, $p = 0.355$), EF (coefficient = 0.091, $p = 0.375$) nor LF (coefficient = 0.105, $p = 0.367$) significantly predicted the RR of discontinuing antidepressant *vs.* placebo.

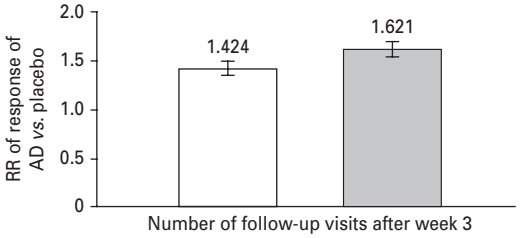

Fig. 2. Risk ratio (RR) of response to antidepressants (AD) *vs.* placebo in studies of 6-wk duration as a function of late frequency. 2 follow-up visits (□) (95% CI 1.29–1.57, $p < 0.0001$, 52 AD *vs.* placebo controls); 3 follow-up visits (▨) (95% CI 1.51–1.73, $p < 0.0007$, 59 AD *vs.* placebo controls).

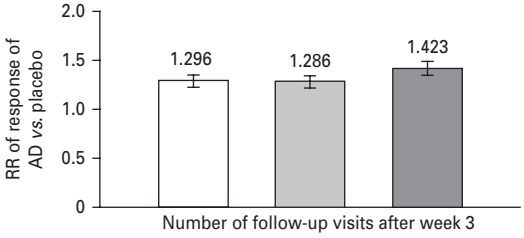

Fig. 3. Risk ratio (RR) of response to antidepressants (AD) *vs.* placebo in studies of 8-wk duration as a function of late frequency. 2–3 follow-up visits (□) (95% CI 1.24–1.35, $p < 0.0001$, 71 AD *vs.* placebo controls); 4 follow-up visits (▨) (95% CI 1.14–1.45, $p < 0.0001$, 6 AD *vs.* placebo controls); 5 follow-up visits (■) (95% CI 1.28–1.58, $p < 0.0001$, 14 AD *vs.* placebo controls).

Discussion

In an earlier meta-analysis, Posternak & Zimmerman (2007) reported a positive relationship between the number of follow-up assessments and symptom reduction following antidepressant and placebo therapy in clinical trials for MDD (greater number of assessments resulting in greater symptom reduction for patients treated with both antidepressants and placebo). However, since antidepressant and placebo response rates are not uniformly distributed throughout the course of a clinical trial, it is reasonable to also examine whether a relationship exists between the temporal distribution of follow-up assessments and clinical trial outcome. In addition, whether the decrease in antidepressant and placebo response rates seen in studies with a greater number of follow-up assessments is proportional or not (in which case it would influence the RR of responding to antidepressants *vs.* placebo), remains unknown. To our knowledge, this meta-analysis is the first systematic attempt to investigate the relationship between the temporal distribution of the frequency of follow-up

assessments and clinical trial outcome expressed as the RR of responding to antidepressants *vs.* placebo in randomized, placebo-controlled, double-blind clinical trials for MDD in adults. Our work suggests that an increased number of follow-up assessments predict a greater antidepressant–placebo separation in MDD trials. This would suggest that while a greater number of assessments could 'inflate' both antidepressant and placebo response rates as clearly demonstrated by Posternak & Zimmerman (2007), this effect is disproportional, such that greater antidepressant–placebo 'separation' is seen in studies with a greater number of assessments. Moreover, we found that, in the same sample, it was the frequency of assessments after week 3 (late assessment frequency, LF) rather than the frequency of assessments before week 3 (early assessment frequency, EF) that influenced trial outcome. In fact, we replicated the same exact results when specifically examining studies of either 6-wk or 8-wk duration, which represented a more homogeneous sample in terms of follow-up assessment schedules. Finally, discontinuation rates did not differ among trials with different number of follow-up visits.

There are several potential study-design implications stemming from our main study finding, namely that visit frequency in the early stages does not influence trial results while visit frequency after week 3 does. Since increased EF was not found to predict a smaller RR of responding to antidepressants *vs.* placebo in MDD clinical trials, future studies should attempt to incorporate a sufficient number of assessments during the first 3 wk of treatment in order for study investigators to be able to: (1) increase the dose of antidepressant therapy to minimally effective or optimal levels, (2) properly assess and, if possible, target treatment-emergent adverse events and, (3) assess for the worsening of symptoms or the emergence and/or worsening of suicidal ideation, which can occur following the initiation of antidepressant therapy. In parallel, since increased LF was found to predict a greater RR of responding to antidepressants *vs.* placebo, future studies should try to optimize the number of these assessments in order to optimize the chances of detecting a difference in antidepressant efficacy between an experimental treatment and placebo. Of interest, the relationship between a greater number of late assessments and a greater antidepressant–placebo 'separation' does not appear to be mediated by differential attrition across studies, since assessment frequency (early, late or overall), was not found to influence premature treatment discontinuation. Of note, the finding that no statistically significant relationship was found between follow-up frequency and study attrition suggests that limiting the number of visits before week 3 would not result in poorer patient retention, which could have undermined any increase in the RR of response by decreasing the statistical power of the study. Although the present study was not designed to explore underlying factors that may mediate the relationship between greater late assessment frequency and greater antidepressant–placebo 'separation' in clinical trials, likely contributing factors may include (1) a better opportunity to optimize antidepressant therapy in treatment non-responders later-on in the study, and, (2) a more accurate measurement of endpoint severity for patients who prematurely discontinue treatment in trials with more frequent assessments after week 3.

Several limitations of our work should be taken into account when interpreting these findings. First, we considered only the 'raw' number of assessments, while it would have been more precise to take into account also their approximate duration, measured by the number of scales administered during each follow-up visit. Therefore, it would be interesting to examine if the duration, combined with the frequency, of the follow-up visits may change our results. However, considering that it was impossible to collect those data in a spread dataset, we tried to limit this 'confounding' factor excluding the first visit (screening), which is the longest one, and considering only the subsequent visits that typically last 30 min. Second, only published studies were included in our dataset. Specifically, it is possible that publication bias in the form of the failure to publish equivocal or negative trials may have distorted our findings (since our study only focused on published clinical trials). Third, the present analysis was based on clinical trial-level data as opposed to individual patient-level data. Having individual patient-level data would have been much more preferable, as it would have afforded us the opportunity to test whether individual patient characteristics influenced the relationship between the temporal distribution of the frequency of follow-up assessments and endpoint outcome. Finally, the clinical trials included in the present study focused only on the treatment of MDD and included a number of exclusion criteria and, thus, the findings of this study may not be generalized to the excluded (i.e. patients with bipolar depression, psychotic MDD, patients actively abusing alcohol or drugs, patients with specific medical comorbidities, or patients with serious suicidal ideation), or to treatment with different modalities (i.e. psychotherapy, somatic therapies).

In summary, a major number of follow-up visits in clinical trials for MDD corresponded to a greater

antidepressant–placebo 'separation' and, specifically, the number of assessments after the third week, rather than within the first 3 wk, influenced the result of clinical trials. Therefore, even if a prospective study is necessary to confirm our results, increasing the number of follow-up visits may be an effective approach to improve the likelihood of success in placebo-controlled clinical trials.

Acknowledgements

None.

Statement of Interest

Dr Levkovitz has served as a speaker for AstraZeneca plc, and has served as a consultant for Neuroderm Ltd and Brainsway Ltd.

Dr Papakostas has served as a consultant for Abbott Laboratories, AstraZeneca plc, Brainsway Ltd, Bristol–Myers Squibb Company, Cephalon Inc., Eli Lilly Co., GlaxoSmithKline, Evotec AG, Inflabloc Pharmaceuticals, Jazz Pharmaceuticals, Otsuka Pharmaceuticals, PAMLAB LLC, Pfizer Inc., Pierre Fabre Laboratories, Ridge Diagnostics (formerly known as Precision Human Biolaboratories), Shire Pharmaceuticals, and Wyeth Inc. He has received honoraria from Abbott Laboratories, Astra Zeneca plc, Bristol–Myers Squibb Company, Brainsway Ltd, Cephalon Inc., Eli Lilly Co., Evotec AG, GlaxoSmithKline, Inflabloc Pharmaceuticals, Jazz Pharmaceuticals, Lundbeck, Otsuka Pharmaceuticals, PAMLAB LLC, Pfizer, Pierre Fabre Laboratories, Ridge Diagnostics, Shire Pharmaceuticals, Titan Pharmaceuticals, and Wyeth Inc. Dr Papakostas has received research support from Bristol–Myers Squibb Company, Forest Pharmaceuticals, the National Institute of Mental Health, PAMLAB LLC, Pfizer Inc., and Ridge Diagnostics, and has served (not currently) on the speakers' bureaux for Bristol–Myers Squibb Co. and Pfizer Inc.

References

APA Task Force on DSM-IV (1994). *Diagnostic and Statistical Manual of Mental Disorders: DSM-IV*. Washington, DC: American Psychiatric Association.

APA Task Force on Nomenclature and Statistics, APA Committee on Nomenclature and Statistics (1980). *Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC: American Psychiatric Association.

APA Work Group to revise DSM-III (1987). *Diagnostic and Statistical Manual of Mental Disorders: DSM-III-R*. Washington, DC: American Psychiatric Association.

Carmody TJ, Rush AJ, Bernstein I, Warden D, *et al.* (2006). The Montgomery Asberg and the Hamilton ratings of depression: a comparison of measures. *European Neuropsychopharmacology* **16**, 601–611.

Faries D, Herrera J, Rayamajhi J, DeBrota D, *et al.* (2000). The responsiveness of the Hamilton Depression Rating Scale. *Journal of Psychiatric Research* **34**, 3–10.

Faries DE, Heiligenstein JH, Tollefson GD, Potter WZ (2001). The double-blind variable placebo lead-in period: results from two antidepressant clinical trials. *Journal of Clinical Psychopharmacology* **21**, 561–568.

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. *Psychotherapy and Psychosomatics* **72**, 115–127.

Feighner JP, Robins E, Guze SB, Woodruff RA, *et al.* (1972). Diagnostic criteria for use in psychiatric research. *Archives of General Psychiatry* **26**, 57–63.

Gelenberg AJ, Thase ME, Meyer RE, Goodwin FK, *et al.* (2008). The history and current state of antidepressant clinical trial design: a call to action for proof-of-concept studies. *Journal of Clinical Psychiatry* **69**, 1513–1528.

Guy W (1976). *ECDEU Assessment Manual for Psychopharmacology*. Rockville, MD.: U.S. Dept. of Health, Education, and Welfare, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration, National Institute of Mental Health, Psychopharmacology Research Branch, Division of Extramural Research Programs.

Hamilton M (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery, and Psychiatry* **23**, 56–62.

Lee S, Walker JR, Jakul L, Sexton K (2004). Does elimination of placebo responders in a placebo run-in increase the treatment effect in randomized clinical trials? A meta-analytic evaluation. *Depression and Anxiety* **19**, 10–19.

Montgomery SA, Asberg M (1979). A new depression scale designed to be sensitive to change. *British Journal of Psychiatry* **134**, 382–389.

Papakostas GI, Fava M (2009). Does the probability of receiving placebo influence clinical trial outcome? A meta-regression of double-blind, randomized clinical trials in MDD. *European Neuropsychopharmacology* **19**, 34–40.

Posternak MA, Zimmerman M (2005). Is there a delay in the antidepressant effect? A meta-analysis. *Journal of Clinical Psychiatry* **66**, 148–158.

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

Spitzer RL, Endicott J, Robins E (1978). Research diagnostic criteria: rationale and reliability. *Archives of General Psychiatry* **35**, 773–782.

Stein DJ, Baldwin DS, Dolberg OT, Despiegel N, *et al.* (2006). Which factors predict placebo response in anxiety disorders and major depression? An analysis of placebo-controlled studies of escitalopram. *Journal of Clinical Psychiatry* **67**, 1741–1746.

Tedeschini E, Fava M, Goodness TM, Papakostas GI (2010). Relationship between probability of receiving placebo and probability of prematurely discontinuing treatment in double-blind, randomized clinical trials for MDD: a meta-analysis. *European Neuropsychopharmacology* **20**, 562–567.

Tedeschini E, Fava M, Papakostas GI (2011). Placebo-controlled, antidepressant clinical trials cannot be shortened to less than four weeks duration. a pooled analysis of randomized clinical trials employing a diagnostic odds ratio-based approach. *Journal of Clinical Psychiatry* **72**, 98–118.

Trivedi MH, Rush H (1994). Does a placebo run-in or a placebo treatment cell affect the efficacy of antidepressant medications? *Neuropsychopharmacology* **11**, 33–43.

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

Wernicke JF, Sayler ME, Koke SC, Pearson DK, *et al.* (1997). Fluoxetine and concomitant centrally acting medication use during clinical trials of depression: the absence of an effect related to agitation and suicidal behavior. *Depression and Anxiety* **6**, 31–39.