Psychological Medicine, 2005, **35**, 743–749. © 2004 Cambridge University Press
doi:10.1017/S0033291704003873 Printed in the United Kingdom

Magnitude of placebo response and drug–placebo differences across psychiatric disorders

ARIF KHAN[1,2]*, RUSSELL L. KOLTS[3], MARK H. RAPAPORT[4], K. RANGA RAMA KRISHNAN[2], AMY E. BRODHEAD[1] AND WALTER A. BROWN[5]

[1] *Northwest Clinical Research Center, Bellevue, WA, USA*; [2] *Department of Psychiatry and Behavioral Sciences, Duke University Medical Center, Durham, NC, USA*; [3] *Department of Psychology, Eastern Washington University, Cheney, WA, USA*; [4] *Department of Psychiatry, Cedars–Sinai Medical Center and the David Geffen School of Medicine at UCLA, Los Angeles, CA, USA*; [5] *Department of Psychiatry, Brown Medical School, Providence, RI and Tufts University School of Medicine, Boston, MA, USA*

ABSTRACT

Background. Placebo response, drug response, and drug–placebo differences appear to vary among psychiatric conditions.

Method. We evaluated the Food and Drug Administration (FDA) Summary Basis of Approval (SBA) reports to compare the magnitude of placebo response, magnitude of psychotopic drug response, and drug–placebo differences among various diagnostic groups such as depression, anxiety, and psychotic disorders.

Results. Six diagnostic groups (psychosis, obsessive–compulsive disorder (OCD), generalized anxiety disorder (GAD), depression, post-traumatic stress disorder, panic) varied in response to both placebo and active drug treatments. Response to placebo was high among patients participating in GAD, depression, and panic disorder clinical trials. Conversely, patients participating in psychotic disorder and OCD trials experienced low response to placebo.

Conclusion. Our findings indicate that the magnitude of placebo response and drug response were heterogeneous and were statistically significantly different among various psychiatric disorders. Although a noticeable degree of heterogeneity was detected in the drug–placebo ratio among various disorders, the differences did not reach statistical significance. This finding suggests that placebo use should be continued for newer agents being tested for all of the psychiatric disorders. These findings may help in the development of psychopharmacology trial designs and in the deliberations of ethics committees.

INTRODUCTION

In a seminal and widely quoted paper Beecher (1955) concluded that 30–40% of patients afflicted with a variety of conditions improve with placebo. Subsequent research shows that although the placebo response is ubiquitous, some medical conditions are more responsive to placebo than others.

Placebo response, drug response and drug–placebo differences appear to vary among psychiatric disorders, but the data bearing on this matter are not entirely consistent. In outpatients treated for depression or anxiety, antidepressants and anxiolytics fail to demonstrate an advantage over placebo 50% of the time, while 30–50% of patients improve with placebo treatment (Khan *et al.* 2002). Alternatively, patients with psychosis or obsessive–compulsive disorder (OCD) show a much lower response to placebo. Nearly 70% of trials have a lower placebo response, with less than 25% of patients improving with placebo (Srisurapanont & Maneeton, 1999; Khan *et al.* 2001; Woods *et al.* 2001).

* Address for correspondence: Dr Arif Khan, 1900 116th Ave NE, Bellevue, WA 98004, USA.
(Email: akhan@nwcrc.net)

In order to obtain more evidence regarding the comparative magnitude of placebo response among psychiatric conditions, we evaluated the Food and Drug Administration (FDA) Summary Basis of Approval (SBA) reports. Using FDA SBA reports is particularly advantageous for two reasons. First, these reports include results from both published and unpublished reports, thus reducing publication bias. Publication bias often occurs in psychopharmacology research because the majority of published studies are ones that obtained positive results. Secondly, by using FDA SBA reports we can be more confident that these trials were tightly controlled and conducted with specific guidelines, thus diminishing potential confounds.

We consider that magnitude of placebo response as well as drug response may have a bearing in the understanding of the pathophysiology of various psychiatric disorders. Further, magnitude of placebo response as well as drug response may help in the design of future clinical trials. Specifically, it may help define the need for placebo controls for some of the psychiatric disorders compared to others. For example, the need for placebo controls is greater in neurotic disorder trials compared to psychotic disorder trials (Srisurapanont & Maneeton, 1999; Woods *et al*. 2001; Khan *et al*. 2002; Walsh *et al*. 2002).

Based upon previous evidence from psychopharmacology clinical trials, we explored the differences among psychiatric conditions in response to placebo and drugs as well as in the amount of drug–placebo differences. Additionally, we examined the improvement in patients assigned to placebo compared to those patients assigned to drug treatments among the various psychiatric conditions. In order to compare placebo response and drug–placebo differences across various psychiatric disorders, we explored the results of both the primary scales for each disorder as well as results of Clinical Global Impressions of Severity scale (CGI-S). The CGI-S ratings were helpful in comparing our results across the different psychiatric disorders.

METHOD

We obtained FDA clinical trial data under the Freedom of Information Act (U.S. Congress, 1999) for nine antidepressants [fluoxetine hydrochloride (Prozac), sertraline hydrochloride (Zoloft), paroxetine hydrochloride (Paxil), venlafaxine hydrochloride (Effexor), nefazodone hydrochloride (Serzone), mirtazapine (Remeron), bupropion hydrochloride SR (Wellbutrin SR), venlafaxine hydrochloride ER (Effexor XR), and citalopram hydrobromide (Celexa)], three antipsychotics [risperidone (Risperdal), olanzapine (Zyprexa), and quetiapine fumarate (Seroquel)], one agent used for bipolar disorder [divalproex sodium (Depakote)], four agents used for panic disorder [alprazolam (Xanax), paroxetine hydrochloride (Paxil), clonazepam (Klonopin), and sertraline hydrochloride (Zoloft)], two agents used for generalized anxiety disorder (GAD) [buspirone (Buspar) and venlafaxine hydrochloride ER (Effexor XR)], one agent used for social anxiety disorder [paroxetine hydrochloride (Paxil)], five agents used for OCD [clomipramine (Anafranil), paroxetine hydrochloride (Paxil), fluvoxamine maleate (Luvox), fluoxetine hydrochloride (Prozac), and sertraline hydrochloride (Zoloft)], and one agent used for post-traumatic stress disorder (PTSD) [sertraline hydrochloride (Zoloft)] approved in the United States between January 1985 and January 2000. The data were sent on microfiche or a hard paper copy for a small fee by a specific request to the FDA (Freedom of Information Staff, 5600 Fishers Lane, HFI-35 Rockville, MD 20857, USA). Some of the more recent clinical trial data were accessed via the Internet.

Of the 26 research programs investigating psychotropic medications, the FDA considered 105 trials to be pivotal (i.e. randomized, double-blind, placebo-controlled with established criteria for the sample under study and defined criteria for response). Of these 105 pivotal trials, we excluded four antidepressant trials and two PTSD treatment trials from our analysis. Three antidepressant trials were excluded due to insufficient data, one antidepressant trial because it focused on relapse prevention rather than acute treatment response, and two anti-PTSD trials because they did not contain baseline data.

Among the eight diagnostic groups 99 trials were analyzed (Table 1); 57 were two-armed (investigational drug *versus* placebo), and 42 were three-armed – investigational drug *versus* active comparator (imipramine, amitriptyline, trazodone, paroxetine, venlafaxine, haloperidol,

Table 1. *Summary of pivotal trials for FDA-approved psychotropics*

Indication	No. of pivotal trials	No. of trials with active comparator	No. of randomized patients			
			Investigational drug	Active comparator	Placebo	Total
Depression	52	24[a]	5682	1577	3411	10 670
Psychosis	7	4[b]	1203	261	462	1926
Obsessive compulsive	13	3[c]	1843	266	1126	3235
Panic	12	3[d]	1681	256	1157	3094
Generalized anxiety	8	7[e]	708	328	435	1471
Social anxiety	3	0	522	0	339	861
Post-traumatic stress	2	0	145	0	144	289
Bipolar	2	1[f]	89	36	97	222
Pooled data	**99**	**42**	**11 873**	**2724**	**7171**	**21 768**

[a] Imipramine hydrochloride, amitriptyline hydrochloride, trazodone hydrochloride, paroxetine hydrochloride or venlafaxine hydrochloride; [b] haloperidol; [c] clomipramine; [d] imipramine hydrochloride, phenylzine, clomipramine, or alprazolam; [e] diazepam, clobazam, or buspirone; [f] lithium.

clomipramine, phenylzine, alprazolam, diazepam, clonazepam, buspirone, or lithium) *versus* placebo. For the purpose of this study, six of the eight diagnostic groups were examined, excluding bipolar and social anxiety. Bipolar was excluded for two reasons; one, due to a lack of generalization with only one drug studied (Depakote) and the other due to the small number of trials. Social anxiety was also excluded due to the small number of trials, as well as lack of information on CGI-S scores.

Initially, we attempted to assess placebo and drug response rates (% satisfactorily relieved of their illness/symptoms). However, FDA SBA reports do not allow for such an evaluation. Also, data such as standard deviations and standard error of the mean were not available. As a related method, we assessed magnitude of symptom reduction for the eight diagnostic groups (Table 1) for patients assigned to placebo, an active comparator, or an investigational drug. Symptom reduction was measured as the mean total percent change from double-blind randomization (baseline) to end-point [last observation carried forward (LOCF)], where patients prematurely terminating from a trial are assumed to experience no further improvement, thus the last measured scores are considered end-of-trial). While results of actual change were also calculated, we believe percentage change is more useful because it accounts for variability in baseline severity.

For purposes of analysis, we combined the results from both investigational drug (now approved and marketed) and active controls into one group called the psychotropic drug group.

The measurements of symptom reduction in the clinical trials were the Brief Psychiatric Rating Scale (BPRS), Mania Rating Scale (MRS), Yale–Brown Obsessive Compulsive Scale (YBOCS), Leibowitz Social Anxiety Scale (LSAS), Hamilton Anxiety Scale (HAMA), Hamilton Depression Scale (HAMD), Clinician-Administered PTSD Scale Part 2 (CAPS-2), and the number of panic attacks that occurred while taking an antipsychotic, bipolar treatment agent, OCD treatment agent, social anxiety treatment agent, GAD treatment agent, antidepressant, PTSD treatment agent, or panic treatment agent. Based on the primary rating scales, the mean percentage of symptom improvement for placebo and active drug treatments among the individual diagnostic groups are presented in Fig. 1 *a*.

However, we realized that comparing data from various primary rating scales is less than ideal, since construction of these scales is geared towards individual illnesses. To overcome this deficit, we considered several options including data transformation, obtaining additional normative data, or using published literature for comparisons. None of these options appeared to satisfactorily remedy this deficit. However, an alternative offered itself in the use of the Clinical Global Impressions of Severity scale (CGI-S). CGI-S is an extensively used clinical rating scale. An experienced clinician conducts the rating to assess a patient's overall pathological state. We used CGI-S as a surrogate to measure

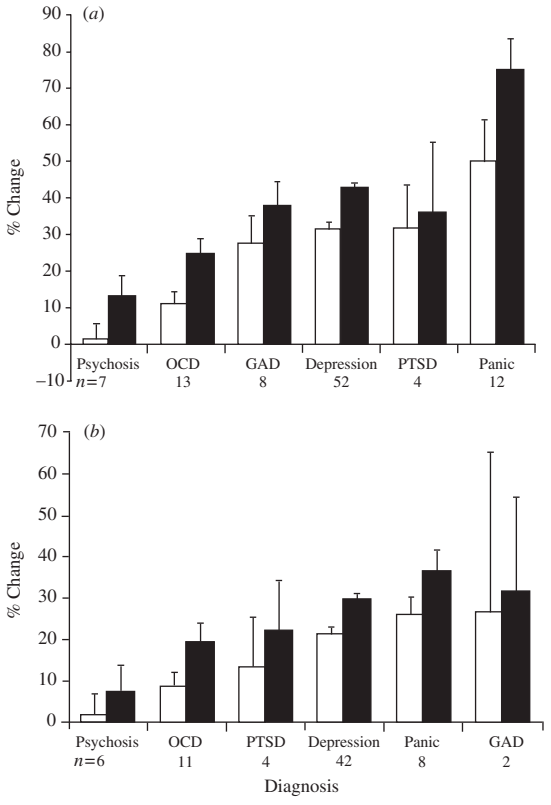

FIG. 1. (*a*) Mean percentage symptom improvement for placebo (□) and psychotropics (■) based on primary rating scale. (*b*) Mean percentage symptom improvement for placebo (□) and psychotropics (■) based on CGI-S.

the level of psychopathology, across the different disorders, as this is the primary function of the scale. Based on the CGI-S, Fig. 1*b* displays the mean percentage of symptom improvement for placebo and active drug treatment. It is noted that for further analyses with CGI-S, PTSD and GAD were eliminated due to the lack of CGI-S data.

We utilized Kruskal–Wallis tests to assess differences between diagnostic groups with regard to placebo response, drug response, and drug–placebo response ratios. In order to account for differing numbers of subjects in the various included trials, the above indices were weighted by sample size prior to testing, done by multiplying the index (for example, percentage improvement in placebo patients from trial A) by the sample size of the trial condition from which it was drawn (for example, all patients in the placebo condition for trial A). To calculate the weighted

Table 2. *Proportion of response and percentage of change from baseline to final LOCF visit seen with placebo compared to psychotropics among psychiatric disorders in 72 clinical trials*

	Depression	Psychosis	PTSD	OCD	GAD	Panic
No. of trials	41	6	4	11	2	8
% of Change from baseline to final visit	2·43	4·0	3·85	6·86	1·22	2·09

LOCF, Last observation carried forward; PTSD, post-traumatic stress disorder; OCD, obsessive–compulsive disorder; GAD, generalized anxiety disorder.

average drug–placebo response ratios for the various diagnoses presented in Table 2, we first calculated individual response ratios for each trial by dividing the percentage improvement seen in the drug condition by the improvement seen in the placebo condition. We then calculated a weighted average ratio for each diagnosis by utilizing the formula $\Sigma x_i n_i / N$, in that the drug–placebo response ratios for each diagnosis were multiplied by the total sample size for the specific trials from which they were drawn, summed across all trials for the diagnosis, and then divided by the combined sample size for all trials within the given diagnostic category.

RESULTS

Patients in each of the six diagnostic groups (psychosis, OCD, GAD, depression, PTSD, panic) varied in response to both placebo and active drug treatments. This is evident by examining the mean percentage symptom reduction scores on diagnosis-specific primary rating scales that are presented in Fig. 1*a*. Based on the CGI-S, mean percentage improvement scores are presented in Fig. 1*b*. Discrepancies between the number of trials presented in Fig. 1*b* and Fig. 1*a* are due to the fact that not all trials utilized the CGI-S and, therefore, were not accounted for in Fig. 1*b*.

In order to assess whether these differences using CGI-S data were significant, Kruskal–Wallis tests were conducted to compare patient response to placebo in depression trials ($n=42$), psychotic disorder trials ($n=6$), OCD ($n=11$), and panic disorder trials ($n=8$). Trials for PTSD ($n=4$) and GAD ($n=2$) were excluded from these analyses due to their small numbers

https://doi.org/10.1017/S0033291704003873 Published online by Cambridge University Press

and inadequate data on CGI-S. Response scores in these analyses were weighted by sample size to account for differing numbers of patients in the various trials. The Kruskal–Wallis test examining percentage improvement on the CGI-S in patients receiving placebo revealed significant differences across diagnostic groups ($\chi^2 = 24{\cdot}76$, df = 3, $p < 0{\cdot}001$). Patient groups receiving active drug treatments also differed significantly by diagnosis with regard to their improvement on the CGI-S ($\chi^2 = 13{\cdot}85$, df = 3, $p = 0{\cdot}003$).

Finally, we examined improvement in patients receiving placebo compared to patients receiving active drug treatments for the six diagnostic groups. To do this, we calculated ratio scores by dividing CGI-S percentage improvement in patients receiving active drug treatments by CGI-S percentage improvement in patients receiving placebo for each study weighted these scores by sample size, and then calculated the average ratio across all of the studies for each disorder (Table 2). Because two of the trials with psychotic patients revealed a worsening of symptoms (which would produce a negative ratio), ratios for these trials were calculated by including a value of 1 in the denominator. Similarly, because one trial revealed a worsening of symptoms in the drug group, a value of 0, indicating no increased efficacy of drug over placebo, was assigned. This ratio index gives an approximate measure of relative efficacy for drug and placebo. For example, depressed patients receiving drug treatments improved 2·43 times as much as depressed patients receiving placebo treatments, while OCD patients receiving active drug treatments improved 6·86 times as much as their counterparts receiving placebo. The differences in this ratio in depressive, psychotic, OCD, and panic disorder patients did not reach statistical significance ($\chi^2 = 5{\cdot}55$, df = 3, $p = 0{\cdot}136$). Again, trials for PTSD ($n = 4$) and GAD ($n = 2$) were excluded due to their small numbers and lack of CGI-S data.

DISCUSSION

The aim of our study was to assess the magnitude of symptom reduction with placebo, with psychotropic drugs, and to assess the magnitude of drug–placebo differences in psychopharmacology clinical trials. The results support our hypothesis that there is considerable heterogeneity in placebo response as well as in drug differences among different psychiatric conditions. This result is based not only on the measures assessing primary symptoms of these psychiatric disorders, but also based on the clinician's overall impression of the severity of a patient's clinical status (CGI-S). This indicates that patients in psychopharmacology clinical trials will respond differently to placebo, as well as active drug treatments, dependent on the diagnosis. Thus, assessing the use of placebo should take into account the disorder under study, rather than assume erroneously that placebo response, as well as drug response, is similar among all psychiatric illnesses.

Although the differences in the drug–placebo ratio (magnitude of drug response change/magnitude of placebo response change) for depressive, psychotic, OCD, and panic disorders were not statistically significant, a noticeable degree of heterogeneity exists. This suggests caution in abandoning placebo use among various psychiatric disorders. Another pattern worth noting relates to the fact that drug–placebo differences were larger among trials for OCD and psychosis. Paradoxically, the absolute effect with drug and placebo were small such that definitions used to define response in those clinical trials are more modest compared to the other disorders.

Most noteworthy are our findings of a lower response to placebo in patients treated for psychosis. These results are consistent with previous reports. Woods *et al.* (2001) conducted a meta-analysis of nine placebo-controlled trials of risperidone, olanzapine, and quetiapine, all treatments for psychotic disorders. They found an antipsychotic-treatment effect size of 0·67 for categorical response rates and >0·82 using the continuous measure, the BPRS. Similarly, Srisurapanont & Maneeton (1999) conducted a meta-analysis of published, randomized, placebo-controlled trials of olanazapine, quetiapine, and risperidone. They found even greater effect sizes of 1·75 for olanzapine, 1·71 for quetiapine, and 3·28 for risperidone. In terms of patients treated for OCD, Hollander *et al.* (2003) reported that those taking fluvoxamine CR had a 31·7% decrease on the YBOCS compared to a 21·2% decrease for patients taking placebo.

https://doi.org/10.1017/S0033291704003873 Published online by Cambridge University Press

Our findings suggest that placebo response may intrinsically be related to subjective distress among the patients. In other words, patients with high levels of psychological distress coupled with a high level of insight into their mental illness such as traditionally termed neurotic disorders (depression, anxiety, panic and PTSD), may have a predilection for non-specific therapeutic response, in other words placebo response. Interestingly, other similar disorders such as bronchial asthma, pain, and arthritis also tend to respond well to placebo. Frank & Frank (1991) make a compelling case that the treatment situation itself is an active ingredient in placebo response. For example, placebo-treated patients receive all the components of the treatment situation common to any treatment (i.e. a thorough evaluation; an explanation for distress; an expert healer; a plausible treatment; expectation of improvement; a healer's commitment, enthusiasm, and positive regard; and an opportunity to verbalize their distress). Additionally, the capsule a patient receives is pharmacologically inert, but hardly inert with respect to its symbolic value and its power as a conditioned stimulus.

On the other hand, patients with relatively low psychological distress as well as relative lack of insight into their mental illness such as psychosis and OCD may lack a predilection for a non-specific therapeutic effect such as placebo. Other disorders such as dementias, eating disorders, sexual identity disorders and personality disorders follow the pattern of a small magnitude of change with placebo (Rinne *et al*. 2002; Hedges *et al*. 2003; Lanctot *et al*. 2003).

Although our analyses support our hypotheses, there are several limitations with this study. First, we had to assess magnitude of symptom reduction for patients assigned to placebo, active comparator, or an investigational drug; the FDA SBA simply does not allow for an evaluation of placebo and drug response rates (% satisfactorily relieved of their illness/symptoms). In addition, data on standard deviations and standard error of the means was lacking or not available, limiting the extent of statistical analysis used in this study.

Secondly, the generalizability of the randomized controlled trials to the general clinical populations is often questioned. Parker (2004) argues that randomized clinical trials have several deficiencies, limiting their generalizability to the clinical population. Participants in clinical trials undergo an extensive screening process wherein participants with co-morbid disorders or medical conditions are excluded from the studies. Zimmerman and colleagues (2002) suggest only 14% of a depressed clinical population would be eligible to complete a clinical trial.

In summary, our findings suggest that placebo response and drug response are heterogeneous among psychopharmacological clinical trials. Although certain heterogeneity exists for drug–placebo differences, the magnitude of this is small and warrants caution before excluding placebo from psychopharmacological clinical trials. These findings may help inform the design of psychopharmacology trials and the deliberations of ethics committees.

ACKNOWLEDGMENTS

All financial support was provided by the Northwest Clinical Research Center, Bellevue, WA, USA.

DECLARATION OF INTEREST

None.

REFERENCES

Beecher, H. K. (1955). The powerful placebo. *Journal of the American Medical Association* **159**, 1602–1606.

Frank J. D. & Frank J. B. (1991). *Persuasion and Healing: A Comparative Study of Psychotherapy*. The Johns Hopkins University Press: Baltimore, MD.

Hedges, D. W., Reimherr, F. W., Hoopes, S. P., Rosenthat, N. R., Kamin, M., Karim, R. & Capece, J. A. (2003). Treatment of bulimia nervosa with topiramate in a randomized, double-blind, placebo-controlled trial, part 2: improvement in psychiatric measures. *Journal of Clinical Psychiatry* **64**, 1449–1454.

Hollander, E., Koran, L. M., Goodman, W. K., Greist, J. H., Ninan, P. T., Yang, H., Li, D. & Barbato, L. M. (2003). A double-blind, placebo-controlled study of the efficacy and safety of controlled-release fluvoxamine in patients with obsessive-compulsive disorder. *Journal of Clinical Psychiatry* **64**, 640–647.

Khan, A., Khan, S. & Brown, W. A. (2002). Are placebo controls necessary to test new antidepressants and anxiolytics? *International Journal of Neuropsychopharmacology* **5**, 193–197.

Khan, A., Khan, S. R., Leventhal, R. M. & Brown, W. A. (2001). Symptom reduction and suicide risk in patients treated with placebo in antidepressant clinical trials: a replication of the Food and Drug Administration Database. *International Journal of Neuropsychopharmacology* **4**, 113–118.

Lanctot, K. L., Herrmann, N., Yau, K. K., Khan, L. R., Liu, B. A., LouLou, M. M. & Einarson, T. R. (2003). Efficacy and safety of cholinesterase inhibitors in Alzheimer's disease: a meta-analysis. *Canadian Medical Association Journal* **169**, 557–564.

Parker, G. (2004). Evaluating treatments for the mood disorders: time for the evident to get real. *Australian and New Zealand Journal of Psychiatry* **38**, 408–414.

Rinne, T., van den Brink, W., Wouters, L. & van Dyck, R., (2002). SSRI treatment of borderline personality disorder: a randomized, placebo-controlled clinical trial for female patients with borderline personality disorder. *American Journal of Psychiatry* **159**, 2048–2054.

Srisurapanont, M. & Maneeton, N. (1999). Comparison of the efficacy and acceptability of atypical antipsychotic drugs: a meta-analysis of randomized, placebo-controlled trials. *Journal of the Medical Association Thailand* **82**, 341–346.

U.S. Congress (1999). Freedom of Information Act. 5 US Congress. 552 (1994 & Supp. II 1996) (http://www.usdoj.gov/04foia/). Accessed 21 January 1999.

Versiani, M. (2000). A review of 19 double-blind placebo controlled studies in social anxiety disorder (social phobia). *World Journal of Biological Psychiatry* **1**, 27–33.

Walsh, B. T., Seidman, S. N., Sysko, R. & Gould, M. (2002). Placebo response in studies of major depression. *Journal of the American Medical Association* **287**, 1840–1847.

Woods, S. W., Stolar, M., Sernjak, M. J. & Charney, D. S. (2001). Consistency of atypical antipsychotic superiority to placebo in recent clinical trials. *Biological Psychiatry* **49**, 64–70.

Zimmerman, M., Mattia, J. I. & Posternak, M. A. (2002). Are subjects in pharmacological treatment trials of depression representative of patients in routine clinical practice? *American Journal of Psychiatry* **159**, 469–473.