Publication Bias and the Efficacy of Antidepressants

Two recently published studies (1, 2) significantly challenge widely accepted views regarding the efficacy of antidepressant medications for unipolar major depressive disorder. The first study contends that publication bias of data from U.S. Food and Drug (FDA) registration trials results in an inaccurate characterization of antidepressant efficacy (1), while the second study argues that even when registration trials are positive, antidepressant efficacy is modest and of doubtful clinical significance (2). Although these reports offer a sober perspective on the benefit of our most commonly prescribed antidepressant medications, the trials suffer from poor generalizability to "real-world" patients (3). Important clinical management issues, such as the optimal duration of treatment, the role of psychotherapy, augmentation strategies, etc., are unaddressed in FDA pivotal trials. To address this gap, the landmark NIMH-funded STAR*D trial examined the acute and longer-term effectiveness of antidepressants and augmentation strategies (including cognitive therapy) in a large and broadly representative sample of major depressive disorder patients undergoing one to four successive treatment steps (4). Although the acute and longer-term remission rates were disappointing, patients who completed all phases of the study had an overall cumulative remission rate of 67% (4). This commentary examines the evidence for publication bias for FDA-registration trials of individual antidepressant medications (1) and evaluates a recent meta-analysis of short-term placebo-controlled studies of newer antidepressants (2). Recommendations to enhance transparent reporting of clinical trial results and reinvigorate antidepressant drug discovery are offered.

Publication Bias and the Evidence Base

Selective reporting of scientific research is of course not unique to randomized clinical trials of antidepressants and impedes the evidence base in medicine (5, 6). Turner et al. (1), in a widely discussed article in the *New England Journal of Medicine*, asked the question: "How accurately does the published literature convey data on drug efficacy to the medical community?" The investigators compared data from 74 FDA-registered randomized controlled trials (for 12 antidepressants involving 12,564 patients) submitted for regulatory approval, with the published literature. They found evidence of the "file drawer effect," that is, publication bias in favor of positive studies. Their major findings included the following: 1) Approximately one-third of all studies, comprising 3,449 patients, went unpublished. 2) Publication status was directly associated with study outcome: 37 of 38 studies with positive results were published, whereas a significantly smaller proportion of studies viewed by the FDA as having negative or questionable results were published (or were published in a way that conveyed a positive outcome). 3) Ninety-four percent of antidepressant trials in the published literature were reported as positive, whereas the FDA database considered only 51% of those same trials as positive. 4) Accordingly, there was a 32% overall increase in effect size of antidepressants in the published literature when compared to the effect size derived from the FDA database. The authors noted examples of misleading information in manuscript abstracts, as well as inappropriately characterized secondary or post hoc analyses. Specific examples of selective reporting included 1) presenting only the positive data from single sites within multicenter studies whose overall results were categorized as negative by the FDA; 2) reporting "efficacy subset" analyses rather than protocol-specified intent-to-treat population analyses; and 3) including data from a site with significant protocol deviations, which resulted in statistical significance for the primary efficacy measure (analyses excluding this outlier site showed a non-significant p value).

Turner et al. (1) acknowledged study limitations, including restricting analyses to industry-supported FDA-registration efficacy trials and the inability to ascertain reasons for nonpublication. Therefore, we cannot exclude the (unlikely) possibility that nonpublished manuscripts were submitted for publication but rejected. It is uncertain how publication bias impacts nonpharmacological antidepressant treatments (phototherapy, psychotherapy, etc.) that do not require regulatory approval, as well as FDA pivotal trials in major depressive disorder using brain stimulation approaches (e.g., vagus nerve stimulation, rTMS).

To ensure validity of meta-analytic studies, the investigator must have access to data from all studies performed, regardless of publication status and ultimate classification as positive, negative, "failed," or equivocal. An example of an equivocal trial would be an active control equivalence trial, or non-inferiority trial, in which a new drug is compared to a known effective drug, in the absence of a placebo control (7–9). Failure to detect differences in efficacy between two treatments in an active control equivalence trial cannot indicate efficacy of the new drug unless assay sensitivity is demonstrated with placebo control (10). A new antidepressant claiming lack of a statistically significant difference from fluoxetine, for example, may be marketed as therapeutically equivalent, although affirming the null hypothesis, as Klein has written, is a "far cry from asserting equivalent benefit" (11). In contrast to active control equivalence trials, "failed" studies, in which neither the standard drug nor the investigational drug is superior to placebo—are much less likely to be published. It has been argued that failed studies have limited scientific value, cannot meaningfully be interpreted, and (not unlike a failed laboratory experiment) should therefore not be submitted for publication (12). We disagree. Unpublished trials, in particular large multicenter phase three studies, are scientifically and ethically problematic because clinicians and researchers cannot make accurate estimates of a drug's efficacy and safety, and these trials lack accountability to patient volunteers exposed to risk.

Efficacy of Antidepressants and Severity of Depression

The primary objective of Kirsch et al.'s (2) meta-analysis of complete data sets (unpublished and published) for four antidepressants—fluoxetine, venlafaxine, nefazodone, and paroxetine—submitted to the FDA for regulatory approval was to examine the relationship between baseline severity and antidepressant efficacy. Of the 35 short-term (primarily 6-week duration) double-blind, placebo-controlled, randomized controlled trials analyzed, involving 3,292 patients on drug and 1,841 on placebo, 31 studies showed an efficacy advantage for drug, determined by mean reduction from baseline on the Hamilton Rating Scale for Depression (HAM-D). The overall drug effect size d was equal to 0.32 (signifying a 1.80-point drug-placebo difference in HAM-D scores), which was similar to Turner's larger study of 12 antidepressants (d=0.31) (1). More robust drug-placebo differences in HAM-D scores (d>0.5) were observed only in patients with severe baseline depressive symptoms. For patients treated with antidepressants, there was no linear relationship between baseline symptom severity and response to antidepressant medication; in other words, similar improvements were found in patients with milder symptoms and those with very severe symptoms (HAM-D>28). In contrast, patients with very severe depressive symptoms treated with placebo showed a marked decline in response compared to patients on placebo with milder depressive severity. Because the overall antidepressant effect size fell significantly below the 0.5 threshold for clinical significance (signifying a three-point difference in HAM-D scores) recommended by the U.K.'s National Institute for Health and Clinical Excellence, the authors concluded: "there seems little evidence to support the prescription of antidepressant medication to any but the most severely depressed patients unless alternative treatments have failed to provide benefit" (2).

Several problems with this conclusion are evident. First, analyses based solely on mean differences in HAM-D at study endpoint between drug and placebo (used to cal-

culate d) address only group-level effects and provide no clinically interpretable information (13, 14). For an individual patient, informative outcomes may include the percentage of patients experiencing response (50% reduction from baseline) or remission (HAM-D score ≤7), number needed to treat, and quality of life. A recent comprehensive analysis of new-generation antidepressants (six selective serotonin reuptake inhibitors and two serotonin neuroepinephrine reuptake inhibitors) submitted for European regulatory approval, which included 56 placebo-controlled trials in 7,374 patients, failed to find a relationship between baseline severity and response rates in either the antidepressant or placebo groups, in contrast to analyses using change from baseline in HAM-D scores (2, 15). The European study found a 16% difference in overall response rates (95% CI: 12%–20%) between antidepressant medication (48%) and placebo (32%) (13). This translates to a number needed to treat of 6.25; that is, approximately six patients would require treatment with an antidepressant medication to produce one response that would not have occurred had the patient been given placebo. Is this a clinically significant value for number needed to treat? The answer depends on one's view of the consequences of suboptimal treatment of major depressive disorder. Kraemer and Kupfer have noted that the more serious the clinical consequences of nonresponse, the higher the threshold number needed to treat is likely to be for clinical significance (14). For example, the number needed to treat associated with the use of cyclosporine, a breakthrough therapy for the prevention of organ rejection, is 6.3 (14). Failure to respond to cyclosporine may result in death or severe disability. For patients with major depressive disorder, the adverse social, economic, and health consequences of nonresponse may justify the risks associated with antidepressant treatment, even in milder presentations of the illness. Second, the optimal clinical management of patients with major depressive disorder is simply not addressed in FDA registration trials. Due to exclusion criteria, very limited efficacy data exist for patients whom clinicians would consider "severely depressed," e.g., patients requiring hospitalization due to active suicidality. Thus, specific clinical recommendations for severely ill major depressive disorder patients on the basis of these data are inappropriate. Third, decades of research have documented the acute and long-term benefit of nonpharmacological therapies, such as structured psychotherapies for mild, moderate, and potentially even severe major depressive disorder (16). However, we are unaware of empirical data to support the view that non-pharmacological therapies should always be *preferred* to antidepressant medication for the acute treatment of major depressive episodes. Patient preferences, economic factors, provider specialty (primary care versus psychiatric versus non-medical mental health professional), and risk/benefit considerations will continue to dictate choice of initial therapy.

While the use of antidepressant medication for acute depressive episodes continues to be debated, there is stronger evidence for the efficacy of antidepressants for the prevention of relapse or recurrence following the acute and/or continuation phases of treatment (17–19). However, the paucity of long-term (≥6 months) placebo-controlled, randomized trials in major depressive disorder is a serious limitation of the evidence base (20). Failure to mandate that antidepressants show long-term safety and benefit (due to concern that such requirements would severely hinder the introduction of new agents) requires that many of the best designed and executed maintenance studies be conducted by academic investigators supported by NIMH or private foundations (18). These informative, yet complex and costly, studies are in jeopardy without substantially increased programmatic funding from federal agencies.

How Can We Improve Experimental Therapeutics for Major Depression?

In 2001, one of the authors (D.S.C.) served as scientific director for an NIMH advisory body charged with formulating a comprehensive Strategic Plan for Mood Disorders (21). The workgroups were comprised of nationally recognized scientific experts, mem-

bers of the National Advisory Mental Health Council, representatives of consumer and advocacy groups, and NIMH staff. In the intervening years, what tangible progress has been achieved in high-priority areas? Implementation of several major initiatives has been largely successful, including integration of pharmacogenomics research with NIMH-supported practical clinical trials to identify single nucleotide polymorphisms and haplotypes that index both therapeutic response and adverse events (22–24). The NIMH Human Genetics Initiative provides an ongoing valuable resource by making biomaterials (DNA samples and cell lines) and clinical data available to the broader scientific community (25).

Progress has been slower in areas related to antidepressant treatment discovery. A major initial recommendation was to support the formation of targeted clinical trial networks to conduct proof-of-concept studies of therapeutic compounds and to validate novel outcome measures, instruments, and biomarkers (26). These NIH-developed networks, highly successful in several other NIH institutes focused on AIDS and cancer, would facilitate innovative drug development based on rational pathophysiology. The impressive successes in HIV therapeutics over the past decade suggest that a focused, targeted approach based on strong funding infrastructure from NIH, as well as industry and nongovernmental organizations, is critical to success (27). To facilitate drug discovery, in 2005 NIMH established novel grant mechanisms that encouraged partnerships between NIMH, academia, and industry such as the Cooperative Drug Discovery Group (CDDG). The CDDG's aim was to test novel mechanism agents in patient populations and perform early proof-of-concept studies of FDA-approved agents in different clinical populations. Although the CDDG program has been discontinued, a similar program will continue to support projects that fill the gap between preclinical drug discovery and large effectiveness trials (28). It is clear that if we are to replicate the therapeutic successes in other areas of medicine, a substantial commitment of federal resources for the establishment of clinical trial networks for experimental therapeutics for major depressive disorder is required. Networks comprised of disease-focused clinical research centers, such as a recently developed network at Massachusetts General Hospital, might facilitate recruitment of research participants with greater illness validity and would offer alternatives to the current Clinical Research Organization-based system, which incentivizes quick enrollment of symptomatic volunteers.

Below we offer several additional recommendations to enhance transparency and foster generativity in antidepressant drug discovery.

1. Industry-sponsored clinical trial protocols submitted for FDA review should include a section detailing publication strategy. At a minimum, this section would include a projected timetable for manuscript(s) submission and list of contributing authors. The FDA lacks the regulatory authority to mandate manuscript submission. However, the FDA currently requires drug manufacturers to submit periodic post-approval drug safety reports as part of postmarketing surveillance procedures and could also require evidence of manuscript submission for all phase 3 trials.

In the meantime, Data and Safety Monitoring Boards (DSMBs) should closely monitor publication status during their regular reviews of individual studies. DSMBs serve as ombudsmen of patients' welfare in clinical research and therefore should encourage timely submission of clinical trial results for publication. The FDA Amendment Act of 2007 requires phase 2 through phase 4 drug trials to be registered prospectively with clinicaltrials.gov prior to participant enrollment, and requires that summary results of primary and secondary outcomes be posted within one year of regulatory approval or trial conclusion. Mandatory clinical trial registration and web-based results reporting are steps in the right direction to foster transparency but will not address publication bias.

2. The FDA should scrutinize the total number of trials conducted for an investigational new drug in making an initial determination of approval for new drug applica-

tions. Package inserts for new antidepressants could be required to disclose the number of placebo-controlled trials conducted for an adequate trial duration at the FDA-approved dose range, along with a summary of trial results (positive, negative, or failed). Clinicians (and patients) have a right to know, for example, that the manufacturer of a new FDA-approved antidepressant performed a total of nine placebo-controlled trials for major depressive disorder, of which only two studies beat placebo. FDA approvals could be annotated with three grades: 1) approval with high enthusiasm, which would require at least 75% of trials to be positive, 2) approval with moderate enthusiasm (50% positive studies), and 3) approval with limited enthusiasm, which signifies that the drug achieved the minimal requirement for approval (two positive studies), but that the majority of studies were negative or failed trials. FDA-approved marketing materials, including direct-to-consumer advertising, could adopt these annotations.

3. Ultimately we need improved approaches to study depression to discover better antidepressants. This will require enhanced understanding of pathophysiological mechanisms associated with short-term therapeutic effects and mechanisms associated with long-term maintenance of benefit. Animal models for the latter are particularly needed. Personalized approaches to antidepressant trials that use biomarkers, including neurophysiological, neuroimaging, genetic, and neuropsychological techniques, are required to guide treatment. Approaches that consider family history and genetics, with identified biomarkers, may reduce heterogeneity and more precisely define phenotypic response patterns in groups of patients (29). Well-engineered small proof-of-concept trials with putative antidepressant agents of novel mechanisms beyond monoaminergic targets require support (30). These investigations can form the basis for more definitive large multicenter trials of potentially more effective antidepressant drugs.

References

1. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R: Selective publication of antidepressant trials and its influence on apparent efficacy. N Engl J Med 2008; 358:252–260
2. Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT: Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. PLoS Med 2008; 5:e45
3. Zimmerman M, Chelminski I, Posternak MA: Generalizability of antidepressant efficacy trials: differences between depressed psychiatric outpatients who would or would not qualify for an efficacy trial. Am J Psychiatry 2005; 162: 1370–1372
4. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, Niederehe G, Thase ME, Lavori PW, Lebowitz BD, McGrath PJ, Rosenbaum JF, Sackeim HA, Kupfer DJ, Luther J, Fava M: Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. Am J Psychiatry 2006; 163:1905–1917
5. Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG: Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA 2004; 291: 2457–2465
6. DeAngelis CD, Fontanarosa PB: Impugning the integrity of medical science: the adverse effects of industry influence. JAMA 2008; 299:1833–1835
7. Thase ME, Clayton AH, Haight BR, Thompson AH, Modell JG, Johnston AJ: A double-blind comparison between bupropion XL and venlafaxine XR: sexual functioning, antidepressant efficacy, and tolerability. J Clin Psychopharmacol 2006; 26:482–488
8. Lenox-Smith AJ, Jiang Q: Venlafaxine extended release versus citalopram in patients with depression unresponsive to a selective serotonin reuptake inhibitor. Int Clin Psychopharmacol 2008; 23:113–119
9. Khan A, Bose A, Alexopoulos GS, Gommoll C, Li D, Gandhi C: Double-blind comparison of escitalopram and duloxetine in the acute treatment of major depressive disorder. Clin Drug Investig 2007; 27:481–492
10. Temple R, Ellenberg SS: Placebo-controlled trials and active-control trials in the evaluation of new treatments. 1. Ethical and scientific issues. Ann Intern Med 2000; 133:455–463
11. Klein DF: Flawed meta-analyses comparing psychotherapy with pharmacotherapy. Am J Psychiatry 2000; 157:1204–1211
12. Ninan PT, Poole RM, Stiles GL: Selective publication of antidepressant trials. N Engl J Med 2008; 358:252–260
13. Melander H, Salmonson T, Abadie E, van Zwieten-Boot B: A regulatory apologia: a review of placebo-controlled studies in regulatory submissions of new-generation antidepressants. Eur Neuropsychopharmacol 2008; 18:623–627
14. Kraemer HC, Kupfer DJ: Size of treatment effect and their importance to clinical research and practice. Biol Psychiatry 2006; 59:990–996

15. Khan A, Brodhead AE, Kolts RL, Brown WA: Severity of depressive symptoms and response to antidepressants and placebo in antidepressant trials. J Psych Res 2005; 39:145–150

16. DeRubeis RJ, Hollon SD, Amsterdam JD, Shelton RC, Young PR, Salomon RM, O'Reardon JP, Lovett ML, Gladis MM, Brown LL, Gallop R: Cognitive therapy vs medications in the treatment of moderate to severe depression. Arch Gen Psychiatry 2005; 62:409–416

17. Keller MB, Trivedi MH, Thase ME, Shelton RC, Kornstein SG, Nemeroff CB, Friedman ES, Gelenberg AJ, Kocsis, JH, Dunner DL, Hirschfeld RMA, Rothschild AJ, Ferguson JM, Schatzberg AF, Zajecka JM, Pedersen RD, Yan B, Ahmed S, Musgnung J, Ninan PT: The prevention of recurrent episodes of depression with venlafaxine for two years (PREVENT) study: outcomes from the 2-year and combined maintenance phases. J Clin Psychiatry 2007; 68:1246–1256

18. Reynolds CF III, Dew MA, Pollock BG, Mulsant BH, Frank E, Miller MD, Houck PR, Mazumdar S, Butters MA, Stack JA, Schlernitzauer MA, Whyte EM, Gildengers A, Karp J, Lenze E, Szanto K, Bensasi S, Kupfer DJ: Maintenance treatment of major depression in old age. N Engl J Med 2006; 354:1130–1138

19. Geddes JR, Carney SM, Davies C, Furukawa TA, Kupfer DJ, Frank E, Goodwin GM: Relapse prevention with antidepressant drug treatment in depressive disorders: a systematic review. Lancet 2003; 361: 653–651

20. Deshauer D, Moher D, Fergusson D, Moher E, Sampson M, Grimshaw J: Selective serotonin reuptake inhibitors for unipolar depression: a systematic review of classic long-term randomized controlled trials. CMAJ 178:1293–1301

21. Breaking Ground, Breaking Through: The Strategic Plan for Mood Disorders Research. NIH Publication No. 03-5121, January 2003

22. Perlis RH, Purcell S, Fava M, Fagerness J, Rush AJ, Trivedi MH, Smoller JW: Association between treatment-emergent suicidal ideation with citalopram and polymorphisms near cyclic adenosine monophosphate response element binding protein in the STAR*D study. Arch Gen Psychiatry 2007; 64:689–697

23. Hu XZ, Rush AJ, Charney D, Wilson AF, Sorant AJ, Papanicolaou GJ, Fava M, Trivedi MH, Wisniewski SR, Laje G, Paddock S, McMahon FJ, Manji H, Lipsky RH: Association between a functional serotonin transporter promoter polymorphism and citalopram treatment in adult outpatients with major depression. Arch Gen Psychiatry 2007; 64:783–792

24. Lekman M, Laje G, Charney D, Rush AJ, Wilson AF, Sorant AJ, Lipsky R, Wisniewski SR, Manji H, McMahon FJ, Paddock S.: The FKBP5-gene in depression and treatment response—an association study in the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) cohort. Biol Psychiatry 2008; 63:1103–1110

25. National Institute of Mental Health Center for Collaborative Genetic Studies on Mental Disorders. http://nimhgenetics.org/

26. Tamminga CA, Nemeroff CB, Blakely RD, Brady L, Carter CS, Davis KL, Dingledine R, Gorman JM, Grigoriadis DE, Henderson DC, Innis RB, Killen J, Laughren TP, McDonald WM, Murphy GM Jr, Paul SM, Rudorfer MV, Sausville E, Schatzberg AF, Scolnick EM, Suppes T: Developing novel treatments for mood disorders: accelerating discovery. Biol Psychiatry 2002; 52: 589–609

27. Wurtman RJ: What went right: why is HIV a treatable infection? Nat Med 1997; 3:714–717

28. Brady LS, Winsky L, Goodman W, Oliveri ME, Stover E: NIMH initiatives to facilitate collaborations among industry, academia, and government for the discovery and clinical testing of novel models and drugs for psychiatric disorders. Neuropsychopharm Reviews 2008; 1–15

29. Pangalos MN, Schechter LE, Hurko O: Drug development for CNS disorders: strategies for balancing risk and reducing attrition. Nat Rev Drug Dis 2007; 6:521–532

30. Agid Y, Buzsaki G, Diamond DM, Frackowiak R, Giedd J, Girault J-A, Grace A, Lambert JJ, Manji H, Mayberg H, Popoli M, Prochiantz A, Richter-Levin G, Somogyi P, Spedding M, Svenningsson P, Weinberger D: How can drug discovery for psychiatric disorders be improved? Nat Rev Drug Dis 2007; 6:189–201

SANJAY J. MATHEW, M.D.
DENNIS S. CHARNEY, M.D.