Drug Approval and Drug Effectiveness

Glen I. Spielmans[1,2] and Irving Kirsch[3,4]

[1]Department of Psychology, Metropolitan State University, Saint Paul, Minnesota 55108

[2]Department of Counseling Psychology, University of Wisconsin, Madison, Wisconsin 53706;
email: glen.spielmans@metrostate.edu

[3]Program in Placebo Studies, Beth Israel Deaconess Medical Center, Harvard Medical School,
Boston, Massachusetts 02215

[4]School of Psychology, University of Plymouth, Plymouth, PL4 8AA United Kingdom;
email: i.kirsch@hull.ac.uk

Keywords

antidepressant, antipsychotic, Food and Drug Administration, clinical
significance, quality of life

Abstract

Data on the efficacy and safety of psychiatric medicines should form the
foundation of evidence-based treatment practices. The US Food and Drug
Administration (FDA) reviews such data in determining whether to approve
new treatments, and the published literature serves as a repository for ev-
idence on treatment benefits and harms. We describe the FDA review of
clinical trials, examining the underlying logic and legal guidelines. Several
FDA reviews provide evidence that the agency requires only minimal effi-
cacy for psychiatric drugs. Further, in some instances, the FDA has relied
on secondary rather than primary outcomes and has discounted the findings
of negative studies in its review of antidepressant and antipsychotic medi-
cations. The published literature provides another lens into the safety and
efficacy of treatments. We describe how treatment efficacy is systematically
overstated and treatment-related harms are understated in the scientific lit-
erature. Suggestions are provided to improve public access to underlying
safety and efficacy data and for the FDA to potentially improve its review
process.

Contents

INTRODUCTION

This article describes the assessment of drug effectiveness and its relation to drug approval. Although much of the included information and our recommendations have wide application, we focus especially on antidepressant assessment and approval. Clinical psychologists have a particular interest in psychotropic drugs, and antidepressants are the most widely prescribed class of psychotropic medication.

In the United States, antidepressants are the most popular treatment for major depressive disorder (Olfson et al. 2002). They are also the third-most prescribed class of medications and are the most-used prescription drugs among adults ages 18–44 (Pratt et al. 2011). Most treatment guidelines recommend these drugs for the first-line management of depression, particularly for patients with severe depressive symptoms (e.g., Am. Psychiatr. Assoc. Work Group Major Depress. Disord, Dep. Veterans Aff. Manag. Major Depress. Disord. Work. Group 2009). Some recent guidelines have also endorsed adding atypical antipsychotic medication to an antidepressant treatment regimen that is not yielding sufficient benefit (Patkar & Pae 2013).

In the United States, the Food and Drug Administration (FDA) examines evidence pertaining to a drug's efficacy and safety when deciding whether to approve a medication for marketing as a treatment. Some might assume that the FDA separates the wheat from the chaff, only approving medications that have demonstrated impressive efficacy. However, thorough meta-analytic reviews over the past 15 years have cast doubt on the clinical significance of antidepressant treatment, suggesting that an FDA stamp of approval does not equate to substantial benefits relative to placebo for most people taking antidepressants (Kirsch & Sapirstein 1998; Kirsch et al. 2002, 2008; Turner et al. 2008b). When weighing the evidence for approving an antidepressant, the FDA does not consider data on such outcomes as quality of life or how people function in work and social settings. Indeed, the extent to which antidepressants help people achieve meaningful change in interpersonal relationships, improve occupational functioning, and attain improved overall well-being is not clearly established (Bech 2005, Greener & Guest 2005, Healy 1999,

Papakostas et al. 2004). With such outcomes ignored in the FDA review process, there is little incentive for sponsors to examine how their drugs target these variables.

A clear understanding of the FDA approval process and its relationship to actual drug efficacy and safety should be beneficial for clinicians, researchers, and consumers of mental health services. This review focuses primarily on antidepressants because of their wide use and relevance in clinical psychology, but the issues raised herein apply throughout all of psychiatry and much of general medicine. We concentrate on issues pertaining to efficacy but place a strong secondary emphasis on the evaluation of drug safety, given that the value of any treatment lies at the intersection of its potential to heal and its potential to harm.

CURRENT FOOD AND DRUG ADMINISTRATION POLICY

Statistical Significance

Prior to marketing a drug for a specific condition, a pharmaceutical firm must receive approval from the FDA. After testing a prospective drug in animals, a drug undergoes a three-phase human testing process, which may culminate in marketing approval. Several trials are conducted as part of an FDA New Drug Application (NDA). In phase I research, a psychiatric drug is tested in normal volunteers and/or people with relevant psychiatric diagnoses to examine side effects and how the drug is absorbed, distributed, excreted, and metabolized. Phase II studies aim to produce preliminary evidence of efficacy and are often open label or single blind in design. These studies should provide some idea of an effective dose range. Phase III studies provide a more rigorous examination of efficacy, comparing the investigational treatment to a placebo control in a double-blind manner.

The FDA antidepressant clinical trial guidelines suggest that "preferably three to five" phase III studies should compare the new drug with a placebo and an already established compound (e.g., an antidepressant with demonstrated greater efficacy than placebo) (Food Drug Admin. Cent. Drug Eval. Res. 1977). These phase III trials are characterized by the FDA as "adequate and well-controlled" trials. According to federal law governing new drug approval, "Reports of adequate and well-controlled investigations provide the primary basis for determining whether there is 'substantial evidence' to support the claims of effectiveness for new drugs" (Code Fed. Regul. 2013). Thus, results of phase III trials are paramount to determining whether a psychiatric drug receives FDA approval. In some instances, the FDA does not rely on placebo-controlled trials to determine efficacy, but for psychiatric medicines, such trials are nearly always the basis of efficacy evidence. The FDA often convenes advisory committees that weigh in on a drug's efficacy and safety profile in clinical trials. These committees typically vote on whether a drug should receive approval; the FDA generally heeds their advice.

When evaluating a trial, FDA reviewers typically consider a study's results positive when the primary outcome reflects a statistically significant advantage over placebo. In psychiatric drug research, results are often mixed, with some studies finding a statistically significant benefit for an investigational drug and others not finding such a benefit. The FDA has interpreted federal law as requiring drug approval if it is found to be safe, the labeling is accurate, and at least two trials have shown positive results—regardless of how many studies have been conducted (Food Drug Admin. 1990, Food Drug Admin. Cent. Drug Eval. Res. 1998).

This was clearly illustrated in the debate over whether Pfizer's antidepressant drug sertraline should win FDA approval. Pfizer conducted five placebo-controlled trials, with three finding negative results on the primary outcome, one finding positive results, and one finding mixed results but that was considered mostly positive by the FDA (positive when results were pooled across dosages but not for two of three doses examined individually) (Food Drug Admin. Cent. Drug

Eval. Res. 1991). Bob Hammer, a psychiatry and statistics professor who served on an advisory committee regarding sertraline's potential FDA approval, interpreted the results as follows:

> If all we had was the two outpatient studies and they fairly clearly showed some sertraline effect, we would have what I interpret as the criteria necessary for us to say go ahead and approve the drug. That is, we have more than one well-documented study that demonstrates an effect. So the question is how do we interpret these two positive results in the context of several more studies that fail to demonstrate that effect? I am not sure I have an answer to that but I am not sure that the law requires me to have an answer to that—fortunately or unfortunately. That would mean, in a sense, that the sponsor could just do studies until the cows come home until he [sic] gets two of them that are statistically significant by chance alone, walks them out and says that he has met the criteria. (Food Drug Admin. 1990)

In the same committee meeting during which Hammer made comments above, Paul Leber, director of the FDA's Division of Neuropharmacological Drug Products, stated that by law, sertraline must be approved when there is "more than one investigation which is adequate and well controlled which would allow experts—experts by experience, training and background—to reach a conclusion that the drug is effective" (Food and Drug Admin. 1990).

An FDA historian (Junod 2012) describes the underlying legal requirements as follows: "The fact that terms such as 'preponderance of evidence' or 'evidence beyond a reasonable doubt' were not used indicates that Congress did not intend to set the bar for efficacious new drug approvals too high. New drugs did not have to be superior to other drugs on the market nor did 'substantial evidence' mean evidence 'so strong as to convince everyone'." Thus, overwhelming evidence of efficacy is clearly not required. Indeed, the FDA's standard of "substantial evidence" of efficacy has been described as between a "scintilla and a preponderance" (Junod 2012).

Much of the American public is unaware of the FDA's relatively low bar for efficacy. Indeed, 39% of a nationally representative sample endorsed that the "FDA only approves prescription drugs that are extremely effective," and 25% believed that "only extremely effective drugs can be advertised to consumers" (Schwartz & Woloshin 2011). To our knowledge, no formal research exists regarding how mental health professionals, including psychologists, interpret the meaning of FDA approval.

Clinical Significance

The FDA's weighing of evidence does not consider clinical significance. If a drug is found to be more effective than a placebo by a statistically significant margin twice, then FDA approval is essentially guaranteed provided that no major safety issues emerge. Whether a drug achieves a statistically significant advantage over a placebo depends on three items: (*a*) the mean difference on the primary outcome for the drug versus the placebo, (*b*) the sample size, and (*c*) the within-group variance. It is well established that the mean difference favoring antidepressant treatment is modest—typically less than the three points on the widely used Hamilton Rating Scale for Depression (HAM-D) that has been proposed as a criterion of clinical significance (Kirsch et al. 2008, Natl. Inst. Health Care Excell. 2004). With a sufficiently large sample size, trials are well powered—some might say overpowered—to detect a statistically significant advantage that is clinically insignificant (Carroll et al. 2004, Hochster 2008).

The FDA does not consider clinical significance when reviewing new drug applications. The National Institute for Health and Care Excellence (NICE), the organization that sets official treatment guidelines for the National Health Service in the United Kingdom, set a drug-placebo difference of three points on the HAM-D or a standardized mean difference of 0.5 as its cutoff

for determining clinical significance (Natl. Inst. Health Care Excell. 2004). A large meta-analysis of second-generation antidepressants [selective serotonin reuptake inhibitors (SSRIs) and other newer drugs] versus placebo found a standardized mean difference effect size of $d = 0.31$ (Turner et al. 2008b), which was interpreted by two of the investigators as "measurable and significant" (Turner & Rosenthal 2008). A similar meta-analysis of antidepressants versus placebo generated an effect size of $d = 0.32$ (Kirsch et al. 2008), and the authors interpreted the treatment benefit as clinically insubstantial. Most researchers likely agree that an antidepressant effect size of $d = 0.19$ [just below the criterion of $d = 0.20$ set for a small effect by Cohen (1988)] is clinically meaningless and that $d = 0.50$ is meaningful. However, the effects generated for most psychiatric drugs when used for their FDA-approved purposes fall somewhere between these goalposts. Antipsychotic drugs are widely believed to have impressive treatment benefits when used for treating schizophrenia. Yet in FDA registration trials, the pooled effect of second-generation antipsychotic drugs over placebo was only 0.44 (Turner et al. 2012). Complicating matters further, treatment-related risks and the financial costs of treatment must also be considered in any overall risk-benefit examination. Debate will likely continue regarding where cutoffs for clinical significance should be set, and as Cohen (1988) noted, these are likely to be different from one domain to another. Despite the difficulty in quantifying treatment benefit, few would disagree that some manner of assessing clinical significance is needed, and for the moment, the NICE criteria are the only ones that have been adopted by an official policymaking agency.

Failed Studies

The FDA distinguishes between negative trials and failed trials. In phase II and III trials, an investigational antidepressant is often compared to both a placebo and an established antidepressant. In such trials, if neither the investigational antidepressant nor the established antidepressant outperforms placebo, the study is written off as "failed," and its results are considered invalid. The trial is said to lack "assay sensitivity." The logic is as follows: Suppose a three-arm trial compared a new drug to fluoxetine (an established antidepressant) and a placebo. Fluoxetine is a known antidepressant, so it if failed to outperform a placebo, then something must have been wrong with the trial itself—perhaps rating scales were done incorrectly or the sample was not appropriate. Essentially, the adequacy of the study is judged by its outcome rather than its methods. Indeed, two senior FDA officials said that this type of three-arm study was preferred because it "allows a clear distinction between a *drug* that does not work (the standard agent is superior to placebo but the new drug is not) and a *study* that does not work (neither the standard drug nor the new drug is superior to placebo)" (Temple & Ellenberg 2000, p. 457).

Judging a study by its outcome as opposed to its methods turns the scientific method on its head. The idea of assay sensitivity is based on the assumption that drugs that are established as efficacious have unquestionably demonstrated strong superiority to placebos (Otto & Nierenberg 2002). This is clearly false. More trials showed negative than positive results in applications for FDA approval for bupropion sustained-release (SR), citalopram, sertraline, and vilazodone (Dinh et al. 2010, Turner et al. 2008b). Nine of the 16 studies submitted for paroxetine yielded either negative or questionable results, and half of the 10 submitted mirtazapine studies showed negative results (Turner et al. 2008b). Yet each of these drugs is an established, FDA-approved antidepressant. Should negative trials of new drugs be deleted from the scientific record because both the new and the "established" drug failed to beat a placebo? The logic of assay sensitivity appears akin to a baseball or football team counting only its victories over the course of a season because the losses must have been fluke events due to poor officiating, injuries, the effect of the crowd, and other events not related to the actual skill of the team. Antidepressants are associated with a modest effect

size benefit over placebos; effect sizes across trials vary from study to study, and such variation is to be expected. Thus, some studies may yield impressive effects while other studies of the same compound will yield an effect that fails to achieve statistical significance. There is no compelling reason to suspect that a trial with negative results was done more poorly than a trial yielding a statistically significant positive finding. Furthermore, studies in which the placebo response is higher are more likely to result in an established antidepressant failing to outperform the placebo. Thus, the reliance on the principle of assay sensitivity leads the FDA to disproportionately rely on trials in which the placebo response is lower. Indeed, pharmaceutical companies and researchers are searching for ways of lowering the placebo response (e.g., excluding data from centers that have a "nonplausible" placebo response rate) in the hope of making it easier to show a statistically significant drug effect (Merlo-Pich et al. 2010, Rutherford & Roose 2013).

Both negative outcomes and the issue of assay sensitivity were on display in the FDA's 2011 approval of vilazodone, in which five phase II studies all failed to demonstrate efficacy of the drug (Dinh et al. 2010). In three of these trials, an approved antidepressant also failed to significantly outperform a placebo, whereas two trials limited to comparing vilazodone to a placebo found no benefit for the drug. Yet the two phase III studies comparing vilazodone to a placebo were positive, leading to the drug's approval. Vilazodone beat a placebo in two of seven trials and is considered an effective antidepressant. Several FDA staff members wrote an article in which they describe how the drug was approved. Germane to the current discussion, it was mentioned that lack of assay sensitivity caused three negative trials to be discarded. The following comment was made regarding the two trials comparing vilazodone to a placebo without also including an established antidepressant comparison: "The 2 trials lacking an active control group could be considered 'negative' trials for vilazodone, but this is not certain in the absence of an active control to confirm assay sensitivity" (Laughren et al. 2011, p. 1168). Similarly, a trial comparing nefazodone and a placebo yielded negative results. The FDA reviewer, in contrasting the negative results of the trial to another trial in which nefazodone was significantly superior to a placebo, wrote, "The inclusion of an active control arm may have provided some explanation for the lack of efficacy" (Food Drug Admin. Cent. Drug Eval. Res. 1994, p. 88). **Table 1** describes how the FDA appears to interpret clinical trial results as positive, negative, or failed depending on the design and trial results.

Table 1   How the US Food and Drug Administration (FDA) apparently interprets results from premarketing trials

Trial type	Result on primary outcome[a]	FDA evaluation
Two-arm: experimental drug, placebo	Negative	Likely "failed study" due to lack of assay sensitivity
	Positive	Positive
Three-arm: experimental drug, established drug, placebo	Negative for experimental drug Negative for established drug	"Failed study" due to lack of assay sensitivity
	Positive for experimental drug Positive for established drug	Positive
	Positive for experimental drug Negative for established drug	Positive
	Negative for experimental drug Positive for established drug	Negative

[a]Positive indicates a statistically significant benefit on the primary outcome measure; negative indicates lack of a statistically significant benefit on the primary outcome measure.

Table 2 Antidepressant trials with negative results on primary outcome but considered supportive of efficacy by the US Food and Drug Administration (FDA)

Trial: drug	Primary outcome; description of FDA interpretation
HGFR: olanzapine/fluoxetine combination (Laughren 2007b, Shelton et al. 2001)	Negative on HAM-D; FDA determined MADRS was a more appropriate measure given that it is purportedly less likely to conflate potential increases in appetite and sleep with antidepressant efficacy. However, FDA reported no such analyses of individual MADRS or HAM-D items in this trial.
HGIE: OFC (Corya et al. 2006, Laughren 2007b)	Negative on MADRS; FDA and sponsor agreed that a posthoc analysis including only participants who had failed two prior antidepressant trials during current depressive episode would be included in the analysis. This posthoc analysis was positive.
X065: fluoxetine for pediatric depression (Emslie et al. 1997)	Negative on combined final CDRS score <28 and endpoint CGI score of 1 or 2; the sponsor (Eli Lilly) later selected a different primary outcome on which positive results were found.
HCJE: fluoxetine for pediatric depression (Emslie et al. 2002)	Negative on reduction in CDRS scores of ≥30%; results were positive on some secondary measures.

Abbreviations: CDRS, Children's Depression Rating Scale (Poznanski & Mokros 1995); CGI, Clinical Global Impression rating scale; HAM-D, Hamilton Rating Scale for Depression (Hamilton 1960); MADRS, Montgomery-Asberg Rating Scale for Depression (Montgomery & Asberg 1979); OFC, olanzapine/fluoxetine combination.

Three-arm trials in which a new drug outperforms a placebo—but in which the established drug is significantly superior to the new drug—appear to be interpreted as positive by FDA. For instance, in trials underlying its FDA application, the antipsychotic drug iloperidone was compared to established agents and a placebo. Its efficacy was worse than existing agents, yet because it outperformed a placebo (depending on which patients were included), it received approval (Turner et al. 2012).

Primary and Secondary Outcomes

Because sponsors collect data on several outcomes, they are instructed to specify a single, primary efficacy outcome measure a priori in order to avoid simply cherry-picking an efficacy measure post hoc on which statistically significant results were obtained. Yet even when a trial obtains a null result on the primary outcome, it may still be interpreted as supporting the efficacy of a drug. **Table 2** lists several instances in which this has occurred in the case of drugs approved as antidepressants. Fluoxetine is one of only two FDA-approved antidepressants for children, yet in both trials upon which its approval was based, the prespecified primary outcome yielded no significant advantage over a placebo (Mosholder 2001). Eli Lilly submitted five studies regarding the efficacy of the combination drug olanzapine/fluoxetine for patients who did not demonstrate an adequate response to antidepressant treatment. One trial clearly demonstrated positive results. None of the other four trials achieved a statistically significant change on the primary outcome measure (Laughren 2007b). However, positive results on a secondary measure in one study and use of only a subset of participants in a second study led to both trials being deemed as providing supportive efficacy evidence. It is disconcerting that the FDA accepted these trials as supporting the drug's efficacy, particularly in the context of a medication tied to substantial weight gain, elevated metabolic laboratory values, frequent sedation, and edema (Spielmans et al. 2013).

Ignoring Negative Efficacy Results

In addition to changing primary outcomes and analyzing only a subset of patients, the FDA has sometimes ignored negative trials entirely. The antipsychotic drug quetiapine was approved for treating schizophrenia in 1997. An extended-release formulation of quetiapine was later created,

which received FDA approval for treating schizophrenia in May 2007. The statistical review notes that its sponsor, AstraZeneca, submitted three studies—and that only one had a positive result. Nonetheless, the drug was approved, as the FDA informed AstraZeneca that only one positive study was needed given that a shorter-acting version of the drug had already received FDA approval (Laughren 2007a). The FDA statistical review described three studies. In one (study 133), three doses of the extended release and one dose of the standard release version of the drug failed to beat placebo; given that 565 patients were enrolled, there was sufficient statistical power to detect an advantage for the drug (Dinh 2007). In a second large trial (study 132), all three doses of extended-release quetiapine outperformed a placebo. In a third large trial (study 41), only one dose (600 mg) of extended-release quetiapine was superior to placebo, with two doses of extended-release quetiapine and two doses of standard-release quetiapine not outperforming placebo (Dinh 2007). The preponderance of the evidence was negative—most comparisons of varying dosages yielded negative results—yet the drug was approved.

The approval of bupropion SR was based on even less impressive evidence. An immediate-release form of bupropion was approved as an antidepressant in 1985. Several years later, Burroughs Wellcome filed for approval to market a SR form of the drug. Three large placebo-controlled trials were conducted in which several doses were used. Only the 300 mg dose was used consistently in each of the three trials; this dosage outperformed the placebo in only one of three studies (study 203), and in the study where it was statistically superior, it barely beat the placebo ($p = 0.04$, $d = 0.27$). The FDA reviewer interpreted study 203 as negative given that the statistically significant benefit on the HAM-D would have not been positive had controls for multiple statistical tests been implemented. By pooling across the three trials (not normally done at the FDA), the FDA reviewer found a statistically significant effect on the HAM-D for the 300 mg SR dose, yet no effect was found on the HAM-D item assessing depressed mood. We performed a meta-analysis of the data provided in the FDA review and found a very modest effect size of $d = 0.17$ on the HAM-D for the 300 mg dose. A prior meta-analysis including doses of 300 mg and above found the same overall effect (Turner et al. 2008b). The FDA reviewer wrote that "the critical issue here is not whether or not bupropion works or whether or not the SR formulation works, but rather, what is the effective dose range that should be recommended to clinicians" (Food Drug Admin. Cent. Drug Eval. Res. 1994–1995, p. 58). The FDA initially rejected the drug but then approved the SR formulation a few months later. The letter indicating drug approval mentioned that "it was decided internally" that data from trials on the immediate-release form of the drug established "an effectiveness link" between the immediate-release form and the SR form of the drug. In other words, the FDA chose to ignore that its reviewer considered all three SR trials as negative and instead extrapolated efficacy data from a different formulation of the drug. The prescribing information for bupropion SR notes that "there are not as yet independent trials demonstrating the antidepressant effectiveness of the sustained-release formulation of bupropion" but describes clinical trials showing the efficacy of the immediate-release version and also notes that the drugs are bioequivalent (GlaxoSmithKline 2013b, p. 6). Thus, consumers and prescribers could logically (yet incorrectly) conclude that the SR form has demonstrated efficacy. The prescribing information contains no mention of the negative trials of the efficacy of bupropion SR.

Long-Term Efficacy

Antidepressants are typically approved for short-term use based on trials ranging from four to eight weeks. If depressive symptoms show substantial improvement in the short term, the FDA considers separately whether a drug is efficacious for maintaining such improvement. Pharmaceutical firms design maintenance-phase trials using the randomized withdrawal design,

in which participants who initially responded to treatment are randomly assigned to receive either continuing treatment with the same drug or to receive a placebo. In nearly all such trials, the switch to placebo treatment is abrupt.

These trials are generally quite similar in design. Patients typically receive an open-label antidepressant treatment. Those who reach a certain response threshold are then randomly assigned to continue the same treatment or to take a placebo. Participants are then followed for a time between several months to a year or two. A meta-analysis of 31 such trials found pooled relapse rates of 41% on placebo versus 18% on antidepressants (Geddes et al. 2003). Such results are taken as evidence that antidepressants are efficacious in the long term.

The problem of antidepressant discontinuation syndrome complicates the interpretation of such trials (El-Mallakh & Briscoe 2012). Antidepressant withdrawal symptoms can be mistaken as relapse, thereby inflating the relapse rate for patients randomized to be switched to placebo. Several studies have examined the impact of abrupt placebo substitution, though few have assessed depressive symptoms using standard measures. One trial examined the effects of abrupt placebo substitution for fluoxetine, paroxetine, or sertraline for a period of five to eight days (Rosenbaum et al. 1998). Participants and clinicians were blinded to placebo substitution. About one-third of patients on sertraline or paroxetine experienced worsening on the HAM-D to the point where they were labeled as having a depressive relapse using strict criteria (HAM-D ≥ 7). Using more relaxed criteria (HAM-D ≥ 12), 14% of sertraline-withdrawn patients and 20% of paroxetine-withdrawn patients had a depressive relapse on the HAM-D. Reinstatement of treatment eliminated these negative mood effects at the aggregate level; it was not reported whether some patients continued to have withdrawal-induced mood problems. Patients discontinuing fluoxetine had much lower rates of withdrawal-associated problems; this is likely due to its long half-life. Michelson and colleagues (2000) found similar results in examining placebo substitution for fluoxetine, paroxetine, or sertraline. A third trial comparing placebo substitution for either fluoxetine or paroxetine produced similar findings (Judge et al. 2002). Another trial examining paroxetine found similar effects at one week; however, paroxetine withdrawal was not associated with increased depression scores at two-week follow-up, although increases in anxiety were still significant (Montgomery et al. 2004).

A meta-analysis of 30 randomized withdrawal antidepressant studies found a statistically significantly lower risk of depressive relapse for patients taking antidepressants at 3-, 6-, 9-, and 12-month follow-ups. However, the relative advantage for antidepressants plateaued at 3 months, with the later follow-up periods finding no additional benefit beyond what had been achieved in the first 90 days (Kaymaz et al. 2008). Hence the enhanced relapse rate associated with switching to placebo is limited to the first 3 months. Elevated risk of relapse following medication discontinuation is not unique to antidepressants; similar phenomena have also been observed after removal of popularly used medications in bipolar and psychotic disorders (Franks et al. 2008, Goodwin et al. 2011, Moncrieff 2006, Tsai et al. 2011, Viguera et al. 1997).

In sum, there is much evidence that the randomized withdrawal method of examining antidepressant maintenance efficacy conflates short-term and protracted antidepressant withdrawal with long-term antidepressant efficacy. As noted in a thorough review of such studies, "researchers must be aware that additional research is required to clarify whether antidepressants are actually preventing future depressive episodes or simply preventing antidepressant withdrawal phenomena" (El-Mallakh & Briscoe 2012, p. 106).

Continuation trials are less likely to conflate withdrawal effects with efficacy. These trials simply follow antidepressant- and placebo-treated participants who demonstrated an adequate response during acute-phase treatment (typically six weeks) over a longer time frame. A meta-analysis of eight continuation trials found that 92% of antidepressant participants maintained response

compared to 79% for placebo. Although this is a statistically significant difference, it is much more modest than what has been observed in randomized withdrawal trials (Khan et al. 2008).

PUBLICATION AND REPORTING BIAS

When considering to what extent an antidepressant is efficacious, one might ask whether the drug is FDA approved for depression, examine the published medical literature, or investigate a broader base of data, including results unpublished in medical journals. For many psychiatric drugs (and medications in general), these sources are likely to provide substantially different information. An antidepressant can be FDA approved and appear to possess robust efficacy in published research, yet the totality of data may cast doubt on its utility.

Efficacy

Trials submitted to the FDA are one source of data regarding efficacy. However, few people read FDA reviews, and many FDA reviews are not publicly accessible. Furthermore, after marketing approval, clinical trial data are typically not submitted to the FDA. Clinicians and researchers are much more likely to consult the published literature than FDA reviews to learn about the merits of various treatments.

Unfortunately, the published literature often fails to accurately represent the underlying data. An analysis of antidepressant trial data submitted to Swedish regulators indicated that half of the 42 completed clinical trials remained unpublished and that trials with positive results were more likely to attain publication (Melander et al. 2003). Ninety percent of trials with positive results were published, compared with only 29% of trials with negative results.

In the eyes of drug sponsors, psychopharmacology trials are marketing tools. Even prior to their official approval for marketing, a publication plan for new drugs is developed, which leverages data on safety and efficacy to influence the targeted audience of prescribers. For instance, an internal Pfizer memo stated that "High quality and timely publications optimize our ability to sell Zoloft [sertraline] most effectively" (Clary 2000). The same document makes it clear that the data from sponsored drug trials belong to the company and the "purpose of data is to support, directly or indirectly, marketing of our product" (Spielmans & Parry 2010).

The antidepressant efficacy literature offers a rare comprehensive glimpse into discrepancies between the published literature and the underlying clinical trial evidence. An ambitious meta-analysis contrasted data contained in FDA reviews of 12 antidepressants to the journal articles based on the same underlying data. Data from all antidepressants approved by the FDA from 1987 to 2004 were included. Of trials yielding positive results according to the FDA, 97% were published in a medical journal. Some trials generated a "questionable" outcome, which found negative or uncertain results on the primary outcome but on which some of the secondary outcomes were positive; "failed studies" were also considered as generating "questionable" outcomes in this analysis. The results of half of these trials were not published in medical journals, whereas half were published but were written up as if the results were clearly positive. Clouding the picture yet further, only one-third of studies finding negative outcomes on the primary measure were published, and five of eight such published articles were written as if the study had a positive outcome. The magnitude of benefit for antidepressants over placebos was inflated by 32% (from $d = 0.31$ to $d = 0.41$) when comparing the published results to all trial results lodged at the FDA (Turner et al. 2008b).

Spinning negative data into a positive journal article involved tactics such as not reporting data from all participants (such as those who dropped out due to lack of efficacy or because adverse events were excluded); reporting data from only one site of a multisite trial; reporting data for something called an "efficacy subset," apparently a euphemism for scrubbing inconvenient data

from the larger dataset; and switching primary outcomes post hoc (Turner et al. 2008a). For each of the 12 antidepressants approved from 1987 to 2004, the overall effect size was smaller in data extracted from the FDA reviews than from the journal articles associated with each drug. Thus, this manipulation of data is not limited to a handful of fluke occurrences; rather, for antidepressants it is standard operating procedure.

A similar study examined discrepancies between data reported to the FDA in antipsychotic trials for schizophrenia and the subsequent journal articles originating from the same studies (Turner et al. 2012). Relative to trials of antidepressants, journal articles more closely matched the underlying data, but there were still several studies in which the data reported to the FDA differed from the data published in journal articles.

A handful of antidepressant studies in children and adolescents trickled into journals from the late 1990s to early 2000s, all of which claimed that the drugs possessed a significant advantage over placebo. However, it turns out that the majority of antidepressant trials in youths found no statistically significant advantage over placebo and that, pooled together, the overall effect in terms of reducing depressive symptoms is very small (Bridge et al. 2007). The positive trials were published reasonably quickly, whereas negative results were either not published at all or often took several years between the date of last data collection and being published in a journal (Reyes et al. 2011). Furthermore, data regarding participants becoming suicidal while taking antidepressants were sometimes not included or were reported unclearly in subsequent journal articles (Healy & Cattell 2003, Jureidini et al. 2008). Thus, the medical literature was well out of line with the underlying data.

Internal AstraZeneca documents showed consternation within the company regarding the questionable efficacy of its antipsychotic drug quetiapine, yet data were presented that painted a much more favorable picture (Spielmans & Parry 2010). In 2000, at the annual convention of the American Psychiatric Association, favorable data were presented regarding quetiapine. A meta-analysis of four studies found that quetiapine was more likely to induce treatment response among patients with schizophrenia than its older generic competitor, haloperidol (Schulz 2000). In an accompanying press release, the author of the presentation stated, "I hope that our findings help physicians better understand the dramatic benefits of newer medications like Seroquel [quetiapine], because, if they do, we may be able to help ensure patients receive these medications first" (Olson 2009). However, an internal document described the results of research comparing the two compounds, and it concluded that quetiapine actually possessed weaker efficacy than haloperidol (AstraZeneca 2000). The company document was produced in March 2000, two months before the presentation of quetiapine's efficacy. An email regarding the internal data analysis, from a publications manager at AstraZeneca, stated in part, "The data don't look good. In fact, I don't know how we can get a paper out of this" (Tumas 2000). In response to a journalist's inquiry, the lead researcher on the 2000 presentation conceded that the claim regarding quetiapine being "significantly superior" was an exaggeration, yet maintained that the data analysis was accurate (Olson 2009).

AstraZeneca also conducted a comparative trial known as study 15. In this study, patients in partial to full remission of schizophrenia were randomly assigned to receive either haloperidol or quetiapine. At the end of the one-year trial, patients in the haloperidol group were significantly better than patients in the quetiapine group in terms of symptom ratings and fewer psychotic relapses. These negative results were not published. Instead, as noted in an internal email, cherry-picking occurred (Tumas 1999). On some measures of cognitive functioning, quetiapine was superior to haloperidol, which was the basis for a publication in the respected journal *Schizophrenia Research* (Velligan et al. 2002). The abstract included the statement, "Treatment with quetiapine at higher doses relative to haloperidol appears to have a positive impact on important domains of cognitive performance that have been found to predict role function and community outcomes in

patients with schizophrenia" (p. 239). Although the paper suggested community outcomes were better for patients taking quetiapine, it did not mention the increased risk of psychotic relapse and the poorer scores on symptom measures compared to haloperidol. According to a Google Scholar search, this article had been cited by 185 subsequent publications as of October 22, 2013—so the positive data on cognitive measures show up nearly 200 times in the scientific literature, whereas the negative efficacy data from the same study are invisible in the so-called scientific evidence base upon which physicians are supposed to base their prescribing decisions.

Although negative data are often not published, positive data may be reported over and over again. When attempting to perform a meta-analysis of risperidone's outcomes for schizophrenia, one pair of authors found 20 articles describing results of randomized controlled trials, yet it appeared that these articles represented only 9 original trials, with the remaining publications simply republishing data that appeared elsewhere (Huston & Moher 1996).

A similar practice is "salami slicing," publishing portions of clinical trials across several papers in what are sometimes called least publishable units. For instance, a small number of clinical trials may collect data on numerous outcomes across several demographic groups. Many pooled analyses (aggregating data across multiple clinical trials) are often published separately, based on highly overlapping or identical underlying data (Melander et al. 2003). In the case of duloxetine for treating depression, several pooled analyses were published separately. Examples include an analysis finding no notable differences in safety and efficacy between African Americans and Caucasians, and another paper based on the same trials finding no notable differences between Hispanics and Caucasians. A further pooled analysis found no differences in treating women ages 40 to 55 compared to other groups. Several other examples can be found in a review of pooled analyses involving duloxetine for depression; data from 6 duloxetine clinical trials were recycled in at least 20 subsequent pooled analyses (Spielmans et al. 2010). Although a much smaller number of pooled analyses could communicate data just as clearly, the publication of a large number of pooled analyses helps widely disseminate key marketing messages regarding the efficacy and safety of a product across many journals. This practice has rarely been analyzed systematically, but for many psychiatric drugs, a five-minute database search will reveal that the number of pooled analyses exceeds the number of published clinical trials by a large margin.

Space limitations preclude a lengthier discussion of publication bias, selective reporting of outcomes, changing primary outcomes, duplicate publication of positive results, and pooled analyses that fail to address a new research topic. However, it is well documented that these issues occur commonly in both the psychiatric literature and across the medical literature as a whole (Chan et al. 2004, Eyding et al. 2010, Howland 2011, Hrobjartsson et al. 2005, McGauran et al. 2010, Spielmans et al. 2013, Spielmans & Parry 2010). These practices have been referred to as marketing-based medicine, evidence-biased medicine, and other unflattering terms. Whatever such tactics are called, they lead to a distorted picture of drug efficacy.

Safety

Incomplete or misleading reporting of adverse events in clinical trials is widespread. An analysis of clinical trials published in six highly cited medical journals found that both the severity of adverse events and statistical tests regarding adverse events were often not reported (Pitrou et al. 2009). The same group also found that in a substantial minority of trials, only a subset of adverse events were reported (e.g., only those occurring in at least 5% of participants). Adverse event reporting in psychopharmacology appears no better than in the general medical literature. Across a wide body of psychopharmacology trials, the median space devoted to adverse events was 0.10 journal pages per article (a median of only 7.1% of the reported results), significantly less than the space

devoted to names of authors and their affiliations (Papanikolaou et al. 2004). We are aware of no broad-based systematic review of adverse event reporting in psychopharmacology trials published since Papanikolaou and colleagues' paper in 2004. However, a recent meta-analysis of clinical trials examining antipsychotic augmentation of antidepressant treatment in depression found that 11 of 14 trials either did not report adverse events or only reported events that either occurred in (*a*) at least 5% of participants in any treatment group, (*b*) 10% of antipsychotic-treated participants, or (*c*) 13% of participants (Spielmans et al. 2013); of the 11 offending trial publications, 10 were published in 2005 or later, which suggests that reporting of adverse events has not improved substantially in recent years.

Clinical trials rarely assess adverse events systematically. In a typical trial, some sort of vague open-ended questioning is used to examine the presence of adverse events. This clearly leads to underreporting. One study enrolled 300 participants who were receiving outpatient treatment for depression. Participants were given a questionnaire that listed 31 antidepressant-related adverse events. Their treating psychiatrists completed progress notes as per their standard practice, which included blank lines on which adverse events could be written. Patients were much more likely to report adverse events than their treating clinicians: 90% of patients reported at least one adverse event, whereas their treating clinicians only reported at least one event in 26% of patients. Patients reported a mean of 7.4 adverse events; clinicians reported 0.6 such events (Zimmerman et al. 2010). When considering adverse events that occurred frequently or were rated as "troubling" by patients, the patient-clinician reporting gap was attenuated. However, patients still reported two to three times as many such events as did their treating psychiatrists.

The underreporting of adverse events in antidepressant clinical trials is clearly illustrated by side effects related to sexual functioning. Across five review articles focused on the safety of SSRIs within a few years after their release, only two reported the incidence of adverse events related to sexual functioning, with the two reviews reporting rates between 3% and 6% (Edwards & Anderson 1999). Randomized controlled trials of SSRIs often reported few, if any, adverse events related to sexual functioning. This likely relates to the sensitive nature of sexual experience. People often do not discuss sexual difficulties with close friends or family members; it seems unlikely that many people would report such experiences to nearly complete strangers without being specifically prompted. However, within a few years of the release of SSRIs, case reports of sexual side effects became commonly published (e.g., Kline 1989, Lydiard & George 1989). Systematic assessments of sexual functioning via questionnaires and interviews then occurred in several studies. In one study, 1,022 outpatients without a history of sexual dysfunction were asked brief questions regarding sexual dysfunction during each treatment visit. Among these patients, 59% reported sexual dysfunction after the onset of treatment, and 38% of those with sexual dysfunction reported that the related events were unacceptable and that they were at a "serious risk for noncompliance" with treatment as a result (Montejo et al. 2001). Another investigation found that whereas 14% of patients taking antidepressants spontaneously reported sexual adverse events, 58% of patients reported such events when directly questioned (Montejo-Gonzalez et al. 1997). A meta-analysis of trials that only (*a*) included participants whose sexual functioning was not impaired at baseline and (*b*) used some form of structured assessment of sexual dysfunction found that, as a group, over half of participants taking SSRIs reported at least one sexual side effect (Serretti & Chiesa 2009).

Current FDA labeling language (as of April 2013) openly admits to underestimating sexual side effects. For instance, various SSRIs provide prescribing information that indicate that changes in sexual desire, performance, or sexual satisfaction may be caused by either a psychiatric disorder or by SSRIs. The prescribing information states that "Reliable estimates of the incidence and severity of untoward experiences involving sexual desire, performance, and satisfaction are difficult to obtain, however, in part because patients and physicians may be reluctant to discuss them.

Accordingly, estimates of the incidence of untoward sexual experience and performance cited in product labeling are likely to underestimate their actual incidence" (e.g., Forest Lab. 2012, GlaxoSmithKline 2013a, Pfizer 2012). Escitalopram's prescription information lists sexual side effects as having a 20% incidence rate versus 3% to 4% for placebo in clinical trials (males) and 6% versus 1% to 2% for females. This rate listing is followed by the statement, "There are no adequately designed studies examining sexual dysfunction with escitalopram treatment" (Forest Lab. 2012). The fluoxetine prescribing information is even less detailed, mentioning that decreased libido was the only sexual side effect reported by at least 2% of participants taking fluoxetine (4% for fluoxetine versus <1% for placebo) (Eli Lilly 2013). This can be compared to the results of a meta-analysis reporting treatment-emergent sexual dysfunction in approximately 70% of patients on fluoxetine and 10% of patients on placebo (Serretti & Chiesa 2009). The detail of sexual side effect reporting varies across the prescribing information of SSRIs, but all appear to substantially underreport the frequency of sexual side effects relative to systematic studies, in which rates of treatment-emergent sexual dysfunction are as high as 80% (Serretti & Chiesa 2009).

Research on suicidality related to antidepressant treatment has been plagued by data manipulations and reporting biases. Prior to beginning randomized treatment with either a drug or a placebo, many trials include a washout phase of a week or two during which all participants receive placebo. This serves to at least partially remove the influence of any prior psychoactive medication as well as to remove participants who improve during the placebo washout from the randomized comparison. Several clinical trials reported suicidal acts occurring during the washout phase as if they occurred during the randomized portion of the study, thus artificially inflating the rates of suicidal behavior for participants taking the placebo (Healy & Whitaker 2003).

Five suicide attempts that occurred in pediatric sertraline trials did not appear in the published versions of the studies; in addition, a patient who committed suicide on sertraline and three others who discontinued sertraline due to suicidal ideation were not reported in publications (Healy & Cattell 2003). Treatment-emergent suicidality was reported under the euphemism of "emotional lability" in the publication of a pediatric paroxetine trial (Jureidini et al. 2008). Furthermore, several additional cases of "emotional lability" appeared in an internal report of the study and were not included in the final study publication (Jureidini et al. 2008).

Increasing Access to Data

In 1997, the FDA Modernization Act's Section 113 mandated the creation of a public resource in which data from efficacy trials would be lodged (Food Drug Admin. Mod. Act 1997). This database would provide a list of ongoing clinical trials in all areas of medicine. The National Library of Medicine, in partnership with the National Institutes of Health, created **www.ClinicalTrials.gov** in 2000. This was intended to serve as a readily searchable online database that included a section for study results. As evidence mounted that journal articles painted an overly optimistic picture of treatment efficacy and safety, the International Committee of Medical Journal Editors announced that, beginning with trials starting enrollment in July 2005, they would only publish trials that were registered at a site such as ClinicalTrials.gov. Furthermore, in 2007, FDA regulations were updated to mandate that results from trials lodged in ClinicalTrials.gov be published no later than 12 months after trial completion (24 months if the drug was under FDA review at the time of the study). With such regulations in place, data reporting should have improved notably. However, systematic studies indicate poor compliance with these legal requirements. One study examined a randomly selected sample of 677 trials that were marked as having data collection completed by the end of 2005. Over half of these trials were not associated with a publication in a journal at least two years after their completion. When researchers contacted the listed study official to

inquire about the publication status of studies that were apparently unpublished, their requests went unanswered 62% of the time (Ross et al. 2009).

Another study examined 738 completed trials registered with ClinicalTrials.gov, all of which should have reported their results within 12 months of study completion under federal law. Only 22% of these trials reported their results within the mandated time frame (Prayle et al. 2012). It is often perceived that failure to publish results is limited to private industry. Yet among 635 completed clinical trials sponsored by the National Institutes of Health, only 4% reported their results on ClinicalTrials.gov, and less than half of such studies had their results published in a medical journal within 30 months of study completion (Ross et al. 2012). Because clinical trial registries promoted transparency in data reporting, some hoped that such registries would increase the reporting of both relevant safety and efficacy data. However, problems clearly remain. Researchers examined a random sample of 500 ClinicalTrials.gov entries and were surprised that only one-quarter of them provided information about the number of patient deaths. Among 27 pairs of ClinicalTrials.gov entries and their associated journal publications, total deaths were apparent in only 15 studies, and the number in the ClinicalTrials.gov entry did not match the journal publication in 5 studies. In addition, 4 of 27 studies reported no deaths in the ClinicalTrials.gov entry but did report deaths in the journal article (Earley et al. 2013).

As noted previously, primary outcomes can be rejiggered after data are analyzed. If the original primary outcome is negative, then another outcome can be declared primary post hoc. This is obviously poor science but is surprisingly common. A comparison of ClinicalTrials.gov study protocols to their associated journal publications found that primary outcomes were changed in 31% of trials (Mathieu et al. 2009). A prior study compared published primary outcomes of trials with their Danish study protocols and found that 62% of trials had a discrepancy between the two (Chan et al. 2004). Both studies found that discrepancies between study protocols and published results tended to favor the reporting of statistically significant outcomes, with nonsignificant primary outcomes often being relegated to secondary status or not being published at all. These studies utilized different trial databases, so their results cannot be directly compared; however, they both indicate that published research results are often an overly optimistic spin on the underlying data (Chan et al. 2004, Mathieu et al. 2009).

In September 2012, the FDA was given responsibilities for enforcement of trial reporting requirements on ClinicalTrials.gov (Off. Comm. Food Drugs 2012). Specifically, the FDA is to determine whether clinical trial information has been submitted in a timely manner or whether information disseminated in clinical trial reports is "false and misleading." Whether adequate staffing and funding will be provided and whether the FDA's leadership perceives this as an important priority is not yet clear. Flustered by unsuccessful attempts to move data into the public realm, many researchers have strongly opined that greater access to data from clinical trials is sorely needed (e.g., Chan 2008, Gotzsche 2011, Tumber & Dickersin 2005). The latest large-scale effort has been led by a group named AllTrials, which proposes that all clinical trial protocols and data should be freely available to all (Alltrials.net 2013). We are hopeful that such efforts will eventually lead to much improvement in data reporting and thus a much firmer foundation for evidence-based medicine, but past efforts at reform have not planted seeds of optimism.

RECOMMENDATIONS

Requirements for New Drug Applications

Drugs demonstrating even quite small effects relative to placebo are approved if they can achieve statistical significance versus placebo in two trials, even if they have not achieved significance in

other trials. Furthermore, the clinical significance of the difference is not considered. These are but two of several problems with the current system, which is ripe for reform. We suggest the following changes in psychotropic drug approval:

1. The current FDA practice of requiring two positive trials—even if the preponderance of evidence casts doubt on a drug's efficacy—is scientifically nonsensical. In an era of containing costs and, supposedly, following evidence-based medicine, perhaps legislative support could be generated requiring the FDA to thoroughly weigh both negative and positive trials.

2. The concept of assay sensitivity in evaluating antidepressants should disappear. There is no basis to support the idea that trials in which both an experimental and an established antidepressant fail to outperform a placebo are somehow flawed and should be discarded (Otto & Nierenberg 2002). A pooled analysis of data across similarly designed placebo-controlled trials would incorporate more data than would relying only on phase III trials or on trials that are not counted as "failed." Such an analysis across trials should provide a broad-based assessment of efficacy on depression measures completed by clinical raters and participants while also incorporating data from such outcomes as social/occupational functioning and quality of life.

3. Information about negative as well as positive trials should be included in FDA-approved labeling. Currently this is not the agency's practice. The NDA for vilazodone, for example, included seven placebo-controlled efficacy trials, only two of which were positive. The FDA-approved label makes no mention of the negative trials. Instead, it merely states, "[T]he efficacy of VIIBRYD was established in two 8-week, placebo-controlled trials in adult patients with MDD" (Forest Lab. 2013). Similar problems apply to the labeling of other antidepressants. There has been some internal debate at the FDA regarding this issue. Paul Leber, who served as director of the FDA's Division of Neuropharmacological Drug Products, opined that "labeling that selectively describes positive studies and excludes mention of negative ones can be viewed as potentially 'false and misleading'" (Leber 1998).

4. Criteria for clinical effectiveness should be examined and eventually adopted. These should be based on continuous outcome measures rather than response or remission rates, as the latter increase the risk of false positives and create the potential for patients with very similar improvement (e.g., 50% versus 49% symptom improvement) to be characterized as very different (responders versus nonresponders). NICE (2004) has proposed a three-point mean difference on the HAM-D or a standardized mean difference of 0.50 as criteria for clinical significance of antidepressants. This effect size corresponds to the cutoff for a moderate effect as defined by Cohen (1988). Adopting this particular cutoff would raise the bar for efficacy given that neither FDA-approved second-generation antidepressants nor antipsychotics as a group met this cutoff in their registration trials (Turner et al. 2008b, 2012). However, there is currently little hard evidence on the appropriateness of various cutoffs for determining clinical significance on continuous outcome measures; we thus encourage further research to examine what might make an appropriate benchmark for clinical significance.

In FDA registration trials, 5 of 12 modern antidepressants yielded a benefit of less than $d = 0.30$ over placebo, and 9 of 12 such drugs provided a benefit of less than $d = 0.40$ (Turner et al. 2008b). Yet in head-to-head trials, the drugs with somewhat higher effects have not generated convincingly superior results to those drugs that found lesser benefits compared to a placebo (Gartlehner et al. 2011). This is not surprising given that antidepressant-placebo differences will vary both systematically according to study design features as well as randomly across studies. Even if all drugs are estimating the same effect, they will yield

various effects across a small number of FDA registration trials. Setting a cutoff for clinical significance thus potentially sets up a situation where drugs scoring above the threshold are not actually superior to other drugs that failed to meet the cutoff. However, this does not mean that clinical significance should be ignored. By the same token, approximately half of the trials of approved antidepressants submitted to the FDA failed to show statistically significant drug-placebo differences. As with adopting a cutoff for clinical significance, the conventional cutoff for statistical significance (i.e., $p < 0.05$) potentially sets up a situation where drugs scoring above the threshold are not actually superior to other drugs that failed to meet the cutoff.

Despite these caveats, we believe that statistical significance is still an important hurdle for a medication to clear. By the same token, clinical significance should also be a key benchmark in determining drug efficacy.

5. The randomized withdrawal design should not be used to assess long-term efficacy. The potential conflation of long-term efficacy with acute and longer-term drug withdrawal effects does not allow for reasonable conclusions to be made (El-Mallakh & Briscoe 2012). The use of a long-term continuation design is more appropriate. In such studies, participants who respond to either an antidepressant or a placebo in an acute-phase trial continue to receive the same treatment for several months to a year afterward.

6. Clinical significance should be interpreted in conjunction with potential harm. Rather than merely requiring that a certain level of statistical or clinical significance be demonstrated, harm/benefit analyses should be conducted. The question to be answered is whether the benefit produced by the drug outweighs its potential harm. In making this assessment, better methods of assessing side effects and other risks need to be used. Unless clinical trials assess adverse events more systematically, it will be difficult to compare safety profiles either across drugs or in relation to benefits, particularly during the FDA approval process. As described previously, sexual side effects were reported as rare in FDA approval trials for antidepressants but were later found to be quite common. It is clear that adverse event reporting in clinical trials, including those used for FDA approval, should be provided in detail and analyzed across trials. The use of systematic checklists and narratives as opposed to reliance on spontaneous report would be a substantial improvement. Clear operational definitions for various adverse events should also be provided for the sake of clarity. The CONSORT (Consolidated Standards of Reporting Trials) statement, a thorough guideline for safety reporting in clinical trials, has been proposed and endorsed by many experts in clinical trial design (Schulz et al. 2010). Adoption of such data-reporting guidelines for adverse events would be useful. Data from FDA clinical trials should be reported online in an easily accessible manner when a drug receives FDA approval.

7. Risk/benefit analyses should be interpreted in relation to those of alternative treatments that are available, including, but not limited to, other medications. To date, little difference in efficacy has been shown between one antidepressant and another, and consideration of adverse events has been suggested as a better criterion for treatment choice (Gartlehner et al. 2011). Comparisons of antidepressants with nondrug treatments also indicate comparable short-term benefits (Khan et al. 2012, Spielmans et al. 2011), so here, too, potential harm might provide a sound basis for treatment choice.

8. It is not clear whether the proliferation of antidepressants with similar modest efficacy, often-similar adverse event profiles, and similar purported mechanisms of action has done anything to improve mental health at a societal level. During the time of widely expanding antidepressant prescriptions, with a glut of SSRIs and SNRIs on the market, there is no evidence that depression at a societal level is improving. Rather, rates of mental health

disability and the societal costs attributed to mental illness continue to increase (Deacon 2013). Perhaps a new antidepressant should be required to demonstrate some sort of clear advantage over an existing treatment—lest another antidepressant enter the market with little to distinguish itself from generic products except a higher price tag and an accompanying marketing blitz.

9. The sole focus on clinician-rated depression severity measures is clearly problematic, as important measures such as self-reports of depressive symptoms, quality of life, and functional impairment are completely ignored in this process (Bech 2005, Healy 2000, Papakostas et al. 2004, Spielmans et al. 2013). There is no good reason to discard such measures when evaluating an antidepressant. An antidepressant should have broad-based efficacy in improving most or all of these measures to some extent. The amount of superiority over a placebo that can be expected in short-term trials on such measures is unclear at this point. However, it is clear that some drugs approved as antidepressants fare poorly on these measures. For instance, the antipsychotic drugs aripiprazole and quetiapine as well as an olanzapine/fluoxetine combination have been approved as adjunctive treatments for depression unresponsive to initial antidepressant intervention. Across three trials, aripiprazole outperformed a placebo by a small effect size of 0.23 on a quality-of-life measure; this modest effect becomes even smaller when patients who violated study protocol are excluded from analysis. Quetiapine and olanzapine/fluoxetine did not show any benefit over a placebo in terms of quality of life (Spielmans et al. 2013).

Quality-of-life measures were not reported in the FDA reviews of most modern antidepressants. Trials in the FDA application for duloxetine used quality-of-life measures and found positive results in about half of the comparisons with a placebo, depending on which measures are included (Siddiqui 2003). The FDA application for desvenlafaxine mentions quality-of-life measures being used but did not describe their results (Chen et al. 2007). It is encouraging that applications for these two relatively recent antidepressants included such measures, but it was disappointing they were not considered meaningful by the FDA. We believe these outcomes should be upgraded in terms of importance because the utility of an antidepressant that provides a statistically significant benefit on a depression measure yet fails to improve quality of life is questionable.

In trials of drugs to be marketed as treatments for cognitive symptoms of schizophrenia, the FDA is requiring use of a "coprimary" measure of functional capacity (Green 2007). The development of such measures has been thoughtful and deliberative, with leaders in the field acknowledging the challenges of finding a measure of real-life functional capacity capable of showing change during short-term trials (Bellack et al. 2007). Performance-based assessments of daily activities are often used to assess functional capacity in schizophrenia; patients who can perform basic tasks such as writing checks, role playing how to handle emergency situations, and planning a public transit trip are more likely to function independently (Mausbach et al. 2008). A recent validation study found that multiple functional capacity measures showed strong measurement properties, but with a substantial caveat. The correlations between functional capacity measures and a quality-of-life scale used to measure community functioning were low (Green et al. 2011). Although other research has linked functional capacity to community functioning (e.g., Mausbach et al. 2011), consistent evidence linking functional capacity to real-world functioning in the context of clinical trial settings is lacking. In other words, it is not known whether improvement on such measures in clinical trials actually relates to better quality of life. Sorting out how to measure (and enhance) functional outcomes in schizophrenia is an area of much current research interest. Unfortunately, it appears that a drug could garner FDA approval via a statistically

significant benefit on both cognitive symptoms and scores on a coprimary measure of functional capacity, yet leave patients with no substantial gain in terms of actual functioning or quality of life. There is precedent for such a possibility.

The FDA required a coprimary measure of global functioning in antidementia drug trials. However, FDA-approved drugs for dementia offer only a clinically marginal benefit over a placebo on both neuropsychiatric ratings and measures of functioning (Kaduszkiewicz et al. 2005, Trinh et al. 2003). Thus, requiring a coprimary measure is no guarantee that FDA-approved drugs will actually generate real-life benefit. We hope that the FDA will adopt standards wherein some sort of meaningful (not just statistically significant) benefit is required for new drugs. Quality of life and functional status may be appropriate measures among most psychiatric outpatients, although we acknowledge assessing such outcomes in short-term trials is challenging, particularly among patients with more severe psychiatric disability. Perhaps longer-term trials are needed to test for potential benefit on meaningful measures.

10. Clinical trials of antidepressants have also rarely included self-report measures of depressive symptom severity. When such measures have been included, they appear to show less of a drug-placebo difference than is obtained on clinician reports, though it is unclear whether this is due to rater bias or differing items across measures (Spielmans & McFall 2006). In any case, several self-reports of depression have solid psychometric properties in assessing depression, and such measures should comprise part of the efficacy portfolio for putative antidepressants.

An overarching problem across efficacy research for both medications and psychotherapy is the uncertain relationship between outcome measures and real-life referents (Kazdin 2006). For instance, it is unknown to what extent a two- or three-point advantage for an antidepressant over a placebo on the HAM-D translates into any particular real-world outcome. Given the dizzying number of individual-item score combinations that may result in the typical two- or three-point advantage for an antidepressant over a placebo, some small differences on depression rating scales probably reflect meaningful differences in a participant's life, whereas others do not. Having participants complete outcome measures with evidence of reliability and validity gives the appearance of science. But we know precious little about how change on these measures is calibrated with real-life improvement.

Required Efficacy and Safety Communication

Currently, the FDA is required to post NDAs online within one year after approval. However, the FDA is clearly not meeting its obligation in this regard. For instance, olanzapine/fluoxetine and quetiapine were both FDA approved in 2009 as adjunctive treatments for depression (in March and December, respectively). For each approved drug, the FDA website includes a main page that provides links to drug labels as well as approval-related documents, such as NDAs (for a guide on navigating the FDA website, see Turner 2013a). Because the olanzapine/fluoxetine NDA was not available on the FDA site, the first author of the present review filed a Freedom of Information Act request to obtain it in October 2012. After a delay of over six months, the NDA was received. The NDA for olanzapine/fluoxetine then appeared on the FDA website sometime between May 2 and July 11, 2013. The NDA for quetiapine as an adjunctive antidepressant is not on the FDA website as of this writing. However, those who are willing to spend substantial time digging through the FDA website can find the FDA's statistical review on quetiapine as presented in an advisory committee meeting prior to the drug's approval (Dinh et al. 2008).

We acknowledge that the FDA provides as much or more information to the public as do similar regulatory agencies in other nations, such as HealthCanada or the European Medicines Agency (EMA) (Turner 2013b). However, as early as 2014, the EMA may implement a policy mandating the release of raw data contained in trials submitted for regulatory approval. The EMA and various stakeholders were working through this process as the current review was being written.

When clinical trial summary data underlying a regulatory approval are not provided online, then the sponsoring drug firm is entirely in charge of deciding what information should be presented publicly. This often leads to the inflation of apparent drug efficacy data in medical journals (Turner et al. 2008b, 2012). For example, although the FDA review of vilazodone essentially ignored the negative findings of five efficacy trials, at least their existence and main results were acknowledged. This can be contrasted with a "comprehensive evaluation" funded by the sponsor of vilazodone, which did not even report the existence of these negative results (Reed et al. 2012).

A clear presentation of results in all individual trials on all outcome measures as well as a thorough pooled analysis across all trials would likely prove useful in clearly communicating information on drug efficacy. Such information should be posted concurrently with a drug's approval rather than awaiting posting within the mandated 12 months, which in current practice is enforced only occasionally.

CONCLUSION

The FDA's framework for evaluating clinical trials allows drugs with minimal efficacy in terms of symptomatic improvement—and no benefit in terms of quality of life or social functioning—to enter the marketplace as approved treatments. The published medical literature inflates the apparent efficacy of antidepressants (and other psychiatric drugs) while downplaying or altogether hiding adverse events. We have proposed a number of ways in which the FDA could raise its standards for approving drugs and more effectively disseminate data to clinicians, researchers, and potential consumers of medication.

A cynic may believe—perhaps with ample justification—that making our suggested changes would simply lead to other questionable maneuvers and subterfuge on the part of drug sponsors. Nonetheless, we believe that some attempt at reform is likely better than throwing up our hands and declaring that we must accept living with lax FDA drug approval standards and poor data reporting both by the FDA and in the wider literature.

SUMMARY POINTS

1. If a psychiatric drug outperforms a placebo by a statistically significant margin in two trials—even if several other trials have failed to demonstrate efficacy—the drug is considered efficacious by the FDA.

2. The FDA should begin weighing clinical significance in its drug approval process.

3. The FDA's use of "lack of assay sensitivity" to label some negative trials as "failed trials" lacks a solid scientific foundation.

4. The randomized withdrawal design to assess maintenance-phase drug efficacy confounds short- and long-term drug withdrawal with putative treatment benefits.

5. The FDA has sometimes considered studies that were negative on the primary outcome variable as providing positive evidence of efficacy.

6. The published literature overestimates drug efficacy and underestimates drug risks.

7. Clinical trials for new drug applications should include such measures as depression self-reports and assessments of quality of life; a true antidepressant should yield benefits on such outcomes.

8. The FDA should more consistently and promptly post data on newly approved drugs.

DISCLOSURE STATEMENT

LITERATURE CITED

Alltrials.net. 2013. All trials registered, all results reported. **http://www.alltrials.net/**

Am. Psychiatr. Assoc. Work Group Major Depress. Disord. 2010. *Practice Guideline for the Treatment of Patients with Major Depressive Disorder*. Washington, DC: Am. Psychiatr. Assoc. 3rd ed. **http://psychiatryonline.org/content.aspx?bookid=28&sectionid=1667485**

AstraZeneca. 2000. *Seroquel*[TM] (*quetiapine*). *Commercial support team—technical document* (*TD004*): *BPRS meta-analysis*. Anchorage, AK: Law Proj. Psychiatr. Rights. **http://psychrights.org/research/Digest/NLPs/Seroquel/OmnibusMSJExhibit2.pdf**

Bech P. 2005. Social functioning: Should it become an endpoint in trials of antidepressants? *CNS Drugs* 19:313–24

Bellack AS, Green MF, Cook JA, Fenton W, Harvey PD, et al. 2007. Assessment of community functioning in people with schizophrenia and other severe mental illnesses: a white paper based on an NIMH-sponsored workshop. *Schizophr. Bull.* 33:805–22

Bridge JA, Iyengar S, Salary CB, Barbe RP, Birmaher B, et al. 2007. Clinical response and risk for reported suicidal ideation and suicide attempts in pediatric antidepressant treatment: a meta-analysis of randomized controlled trials. *JAMA* 297:1683–96

Chan AW. 2008. Bias, spin, and misreporting: time for full access to trial protocols and results. *PLoS Med.* 5:e230

Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. 2004. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 291:2457–65

Chen Y, Yang P, Hung J. 2007. *Statistical review and evaluation of desvenlafaxine extended-release for major depressive disorder*. NDA 21-992. Washington, DC: FDA

Clary C. 2000. Zoloft: PSC update. Pfizer. Document from *Motus vs. Pfizer* case as cited in Moffat B, Elliot C. 2007. Ghost marketing: pharmaceutical companies and ghostwritten journal articles. *Perspect. Biol. Med.* 50:18–31

Code Fed. Regul. 2013. 21CFR314.126. Washington, DC: FDA. **http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=314.126**

Cohen J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Erlbaum. 2nd ed.

Corya SA, Williamson D, Sanger TM, Briggs SD, Case M, Tollefson G. 2006. A randomized, double-blind comparison of olanzapine/fluoxetine combination, olanzapine, fluoxetine, and venlafaxine in treatment-resistant depression. *Depress. Anxiety* 23:364–72

Deacon BJ. 2013. The biomedical model of mental disorder: a critical analysis of its validity, utility, and effects on psychotherapy research. *Clin. Psychol. Rev.* 33:846–61

Dep. Veterans Aff. Manag. Major Depress. Disord. Work. Group. 2009. Clinical practice guideline: management of major depressive disorder (MDD). Washington, DC: Dep. Veterans Aff., Dep. Defense. **http://www.healthquality.va.gov/mdd/MDD_FULL_3c1.pdf**

Dinh P. 2007. *Seroquel SR (quetiapine fumarate): treatment of schizophrenia. Statistical review and evaluation. NDA 22-047/N000.* Washington, DC: FDA. **http://www.accessdata.fda.gov/drugsatfda_docs/nda/2007/022047Orig1s000StatR.pdf**

Dinh P, Yang P, Hung HMJ. 2008. *Statistical review and evaluation of quetiapine for major depressive disorder.* Washington, DC: FDA. **http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/PsychopharmacologicDrugsAdvisoryCommittee/UCM161917.pdf**

Dinh P, Yang P, Hung HMJ. 2010. *Vilazodone for major depressive disorder. Statistical review and evaluation. NDA 22-567/N000.* Washington, DC: FDA. **http://www.accessdata.fda.gov/drugsatfda_docs/nda/2011/022567Orig1s000StatR.pdf**

Earley A, Lau J, Uhlig K. 2013. Haphazard reporting of deaths in clinical trials: a review of cases of ClinicalTrials.gov records and matched publications—a cross-sectional study. *BMJ Open* 3:1. doi:10.1136/bmjopen-2012-001963

Edwards JG, Anderson I. 1999. Systematic review and guide to selection of selective serotonin reuptake inhibitors. *Drugs* 57:507–33

Eli Lilly. 2013. Highlights of prescribing information for Prozac (fluoxetine hydrochloride). Indianapolis, IN: Eli Lilly

El-Mallakh RS, Briscoe B. 2012. Studies of long-term use of antidepressants: How should the data from them be interpreted? *CNS Drugs* 26:97–109

Emslie GJ, Heiligenstein JH, Wagner KD, Hoog SL, Ernest DE, et al. 2002. Fluoxetine for acute treatment of depression in children and adolescents: a placebo-controlled randomized clinical trial. *J. Am. Acad. Child Adolesc. Psychiatry* 41:1205–14

Emslie GJ, Rush J, Weinberg WA, Kowatch RA, Hughes CW, et al. 1997. A double-blind, randomized, placebo-controlled trial of fluoxetine in children and adolescents with depression. *Arch. Gen. Psychiatry* 54:1031–37

Eyding D, Lelgemann M, Grouven U, Harter M, Kromp M, et al. 2010. Reboxetine for acute treatment of major depression: systematic review and meta-analysis of published and unpublished placebo and selective serotonin reuptake inhibitor controlled trials. *BMJ* 341:c4737

Food Drug Admin. 1990. *Psychopharmacological Drugs Advisory Committee 33rd meeting.* Rockville, MD: FDA

Food Drug Admin. Cent. Drug Eval. Res. 1977. *Guidelines for the clinical evaluation of antidepressant drugs.* Washington, DC: FDA. **http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071299.pdf**

Food Drug Admin. Cent. Drug Eval. Res. 1991. *Statistical and medical reviews of sertraline.* Washington, DC: FDA. **http://drl.ohsu.edu/cdm/ref/collection/fdadrug/id/3287**

Food Drug Admin. Cent. Drug Eval. Res. 1994. *Statistical review for nefazodone. Major depressive disorder.* Washington, DC: FDA. **http://drl.ohsu.edu/cdm/ref/collection/fdadrug/id/3282**

Food Drug Admin. Cent. Drug Eval. Res. 1994–1995. *Bupropion sustained-release clinical review and related approval documents.* Washington DC: FDA. **http://drl.ohsu.edu/cdm/ref/collection/fdadrug/id/3280**

Food Drug Admin. Cent. Drug Eval. Res. 1998. *Guidance for industry: providing clinical evidence of effectiveness for human drugs and biological products.* Washington, DC: USDHHS, FDA, Cent. Drug Eval. Res., Cent. Biol. Eval. Res. **http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm078749.pdf**

Food Drug Admin. Mod. Act. 1997. *Public law 105–115.* Washington, DC: FDA **http://www.fda.gov/RegulatoryInformation/Legislation/FederalFoodDrugandCosmeticActFDCAct/SignificantAmendmentstotheFDCAct/FDAMA/FullTextofFDAMAlaw/default.htm#SEC**

Forest Lab. 2012. December. Highlights of prescribing information for Lexapro (escitalopram oxalate). New York: Forest Lab. **http://www.frx.com/pi/lexapro_pi.pdf**.

Forest Lab. 2013. Highlights of prescribing information for Viibryd (vilazodone hydrochloride). New York: Forest Lab. **http://www.frx.com/pi/Viibryd_pi.pdf**

Franks M, Macritchie KA, Mahmood T, Young AH. 2008. Bouncing back: Is the bipolar rebound phenomenon peculiar to lithium? A retrospective naturalistic study. *J. Psychopharmacol.* 22:452–56

Gartlehner G, Hansen RA, Morgan LC, Thaler K, Lux L, et al. 2011. Comparative benefits and harms of second-generation antidepressants for treating major depressive disorder: an updated meta-analysis. *Ann. Intern. Med.* 155:772–85

Geddes JR, Carney SM, Davies C, Furukawa TA, Kupfer DJ, et al. 2003. Relapse prevention with antidepressant drug treatment in depressive disorders: a systematic review. *Lancet* 361:653–61

GlaxoSmithKline. 2013a. January. Paxil (paroxetine hydrochloride) prescribing information. Research Triangle Park, NC: GlaxoSmithKline. **http://www.accessdata.fda.gov/drugsatfda_docs/label/2012/020031s067,020710s031.pdf**

GlaxoSmithKline. 2013b. Wellbutrin SR (bupropion sustained-release) tablets prescribing information. Research Triangle Park, NC: GlaxoSmithKline. **http://www.accessdata.fda.gov/drugsatfda_docs/label/2013/018644s045,020358s052lbl.pdf**

Goodwin FK, Whitham EA, Ghaemi SN. 2011. Maintenance treatment study designs in bipolar disorder: Do they demonstrate that atypical neuroleptics (antipsychotics) are mood stabilizers? *CNS Drugs* 25:819–27

Gotzsche PC. 2011. Why we need easy access to all data from all clinical trials and how to accomplish it. *Trials* 12:249

Green MF. 2007. Stimulating the development of drug treatments to improve cognition in schizophrenia. *Annu. Rev. Clin. Psychol.* 3:159–80

Green MF, Schooler NR, Kern RS, Frese FJ, Granberry W, et al. 2011. Evaluation of functionally meaningful measures for clinical trials of cognition enhancement in schizophrenia. *Am. J. Psychiatry* 168:400–7

Greener MJ, Guest JF. 2005. Do antidepressants reduce the burden imposed by depression on employers? *CNS Drugs* 19:253–64

Hamilton M. 1960. A rating scale for depression. *J. Neurol. Neurosurg.* 23:56–61

Healy D. 1999. The three faces of the antidepressants: a critical commentary on the clinical-economic context of diagnosis. *J. Nerv. Ment. Dis.* 187:174–80

Healy D. 2000. The assessment of outcomes in depression: measures of social functioning. *Rev. Contemp. Pharmacother.* 11:295–301

Healy D, Cattell D. 2003. Interface between authorship, industry and science in the domain of therapeutics. *Br. J. Psychiatry* 183:22–27

Healy D, Whitaker C. 2003. Antidepressants and suicide: risk-benefit conundrums. *J. Psychiatry Neurosci.* 28:331–37

Hochster HS. 2008. The power of "p": on overpowered clinical trials and "positive" results. *Gastrointest. Cancer Res.* 2:108–9

Howland RH. 2011. Publication bias and outcome reporting bias: agomelatine as a case example. *J. Psychosoc. Nurs. Ment. Health Serv.* 49:11–14

Hrobjartsson A, Chan AW, Haahr MT, Gotzsche PC, Altman DG. 2005. Selective reporting of positive outcomes in randomised trials—secondary publication: a comparison of protocols with published reports. *Ugeskr. Laeger* 167:3189–91

Huston P, Moher D. 1996. Redundancy, disaggregation, and the integrity of medical research. *Lancet* 347:1024–26

Judge R, Parry MG, Quail D, Jacobson JG. 2002. Discontinuation symptoms: comparison of brief interruption in fluoxetine and paroxetine treatment. *Int. Clin. Psychopharmacol.* 17:217–25

Junod SW. 2012. *FDA and Clinical Drug Trials: A Short History*. Washington, DC: FDA. **http://www.fda.gov/%20AboutFDA/WhatWeDo/History/Overviews/ucm304485.htm**

Jureidini JN, McHenry LB, Mansfield PR. 2008. Clinical trials and drug promotion: selective reporting of study 329. *Int. J. Risk Saf. Med.* 20:73–81

Kaduszkiewicz H, Zimmerman T, Beck-Bornholdt H, den Bussche H. 2005. Cholinesterase inhibitors for patients with Alzheimer's disease: systematic review of randomised clinical trials. *BMJ* 331:321

Kaymaz N, van Os J, Loonen AJ, Nolen WA. 2008. Evidence that patients with single versus recurrent depressive episodes are differentially sensitive to treatment discontinuation: a meta-analysis of placebo-controlled randomized trials. *J. Clin. Psychiatry* 69:1423–36

Kazdin AE. 2006. Arbitrary metrics: implications for identifying evidence-based treatments. *Am. Psychol.* 61:42–49

Khan A, Faucett J, Lichtenberg P, Kirsch I, Brown WA. 2012. A systematic review of comparative efficacy of treatments and controls for depression. *PLoS ONE* 7:e41778

Khan A, Redding N, Brown WA. 2008. The persistence of the placebo response in antidepressant clinical trials. *J. Psychiatr. Res.* 42:791–96

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

Kirsch I, Moore TJ, Scoboria A, Nicholls SS. 2002. The emperor's new drugs: an analysis of antidepressant medication data submitted to the U.S. Food and Drug Administration. *Prev. Treat.* 5

Kirsch I, Sapirstein G. 1998. Listening to Prozac but hearing placebo: a meta-analysis of antidepressant medication. *Prev. Treat.* 1

Kline MD. 1989. Fluoxetine and anorgasmia. *Am. J. Psychiatry* 146:804–5

Laughren TP. 2007a. *Recommendation of approval action for Seroquel XR (quetiapine) tablets for the short-term treatment of schizophrenia*. Washington, DC: FDA. **http://www.accessdata.fda.gov/drugsatfda_docs/ nda/2007/022047Orig1s000SumR.pdf**

Laughren TP. 2007b. *Recommendation for approvable action for Symbyax (olanzapine/fluoxetine) for treatment of resistant depression (TRD)*. Washington, DC: FDA. **http://www.accessdata.fda.gov/drugsatfda_docs/ nda/2009/021520Orig1s012.pdf**

Laughren TP, Gobburu J, Temple RJ, Unger EF, Bhattaram A, et al. 2011. Vilazodone: clinical basis for the US Food and Drug Administration's approval of a new antidepressant. *J. Clin. Psychiatry* 72:1166–73

Leber P. 1998. *Approvable action on Forrest [sic] Laboratories Inc NDA 20–822 Celexa (citalopram HBr) for the management of depression*. Washington, DC: FDA. **http://www.accessdata.fda.gov/drugsatfda_docs/ nda/98/020822a_admindocs_corres_P1.pdf**

Lydiard RB, George MS. 1989. Fluoxetine-related anorgasmy. *South. Med. J.* 82:933–34

Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. 2009. Comparison of registered and published primary outcomes in randomized controlled trials. *JAMA* 302:977–84

Mausbach BT, Bowie CR, Harvey PD, Twamley EW, Goldman SR, et al. 2008. Usefulness of the UCSD performance-based skills assessment (UPSA) for predicting residential independence in patients with chronic schizophrenia. *J. Psychiatr. Res.* 42:320–27

Mausbach BT, Depp CA, Bowie CR, Harvey PD, McGrath JA, et al. 2011. Sensitivity and specificity of the UCSD Performance-based Skills Assessment (UPSA-B) for identifying functional milestones in schizophrenia. *Schizophr. Res.* 132:165–70

McGauran N, Wieseler B, Kreis J, Schuler YB, Kolsch H, Kaiser T. 2010. Reporting bias in medical research— a narrative review. *Trials* 11:37

Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. 2003. Evidence b(i)ased medicine—selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ* 326:1171–73

Merlo-Pich E, Alexander RC, Fava M, Gomeni R. 2010. A new population-enrichment strategy to improve efficiency of placebo-controlled clinical trials of antidepressant drugs. *Clin. Pharmacol. Ther.* 88:634–42

Michelson D, Fava M, Amsterdam J, Apter J, Londborg P, et al. 2000. Interruption of selective serotonin reuptake inhibitor treatment: double-blind placebo-controlled trial. *Br. J. Psychiatry* 176:363–68

Moncrieff J. 2006. Does antipsychotic withdrawal provoke psychosis? Review of the literature on rapid onset psychosis (supersensitivity psychosis) and withdrawal-related relapse. *Acta Psychiatr. Scand.* 114:3–13

Montejo AL, Llorca G, Izquierdo JA, Rico-Villademoros F. 2001. Incidence of sexual dysfunction associated with antidepressant agents: a prospective multicenter study of 1022 outpatients. Spanish Working Group for the Study of Psychotropic-Related Sexual Dysfunction. *J. Clin. Psychiatry* 62(Suppl. 3):10–21

Montejo-Gonzalez AL, Llorca G, Izquierdo JA, Ledesma A, Bousono M, et al. 1997. SSRI-induced sexual dys-function: fluoxetine, paroxetine, sertraline, and fluvoxamine in a prospective, multicenter, and descriptive clinical study of 344 patients. *J. Sex Marital Ther.* 23:176–94

Montgomery SA, Asberg M. 1979. A new depression scale designed to be sensitive to change. *Br. J. Psychiatry* 134:382–89

Montgomery SA, Kennedy SH, Burrows GD, Lejoyeux M, Hindmarch I. 2004. Absence of discontinuation symptoms with agomelatine and occurrence of discontinuation symptoms with paroxetine: a randomized, double-blind, placebo-controlled discontinuation study. *Int. Clin. Psychopharmacol.* 19:271–80

Mosholder A. 2001. *Statistical review for fluoxetine for pediatric OCD and depression*. NDA 18–936. Washington, DC: FDA. **http://www.antidepressantsfacts.com/doctor-Emslie-3.pdf**

Natl. Inst. Health Care Excell. 2004. Depression: management of depression in primary and secondary care. *Clin. Guidel. 23*. London: NICE

Off. Comm. Food Drugs. 2012. *Delegation of authority. A notice by the Food and Drug Administration.* Washington, DC: Fed. Regist. **http://www.federalregister.gov/articles/2012/09/26/2012-23598/office-of-the-commissioner-of-food-and-drugs-delegation-of-authority.html**

Olfson M, Marcus SC, Druss B, Elinson L, Tanielian T, Pincus HA. 2002. National trends in the outpatient treatment of depression. *JAMA* 287:203–9

Olson J. 2009. U doctor scrutinized over drug research. *St. Paul Pioneer Press*, March 18. **http://www.twincities.com/ci_11945154**

Otto MW, Nierenberg AA. 2002. Assay sensitivity, failed clinical trials, and the conduct of science. *Psychother. Psychosom.* 71:241–43

Papakostas GI, Petersen T, Mahal Y, Mischoulon D, Nierenberg AA, Fava M. 2004. Quality of life assessments in major depressive disorder: a review of the literature. *Gen. Hosp. Psychiatry* 26:13–17

Papanikolaou PN, Churchill R, Wahlbeck K, Ioannidis JP. 2004. Safety reporting in randomized trials of mental health interventions. *Am. J. Psychiatry* 161:1692–97

Patkar AA, Pae C. 2013. Atypical antipsychotic augmentation strategies in the context of guideline-based care for the treatment of major depressive disorder. *CNS Drugs* 27:29–37

Pfizer. 2012. December. Zoloft (sertraline hydrochloride) prescribing information. New York: Pfizer. **http://www.accessdata.fda.gov/drugsatfda_docs/label/2013/019839s079,020990s038lbl.pdf**

Pitrou I, Boutron I, Ahmad N, Ravaud P. 2009. Reporting of safety results in published reports of randomized controlled trials. *Arch. Intern. Med.* 169:1756–61

Poznanski E, Mokros H. 1995. *Children's Depression Rating Scale-Revised (CDRS-R)*. Los Angeles: West. Psychol. Serv.

Pratt LA, Brody DJ, Gu Q. 2011. *Antidepressant use in persons aged 12 and over: United States, 2005–2008.* NCHS Data Brief 76. Hyattsville, MD: Natl. Cent. Health Statist.

Prayle AP, Hurley MN, Smyth AR. 2012. Compliance with mandatory reporting of clinical trial results on ClinicalTrials.gov: cross sectional study. *BMJ* 344:d7373

Reed CR, Kajdasz DK, Whalen H, Athanasiou MC, Gallipoli S, Thase ME. 2012. The efficacy profile of vilazodone, a novel antidepressant for the treatment of major depressive disorder. *Curr. Med. Res. Opin.* 28:27–39

Reyes MM, Panza KE, Martin A, Bloch MH. 2011. Time-lag bias in trials of pediatric antidepressants: a systematic review and meta-analysis. *J. Am. Acad. Child Adolesc. Psychiatry* 50:63–72

Rosenbaum JF, Fava M, Hoog SL, Ascroft RC, Krebs WB. 1998. Selective serotonin reuptake inhibitor discontinuation syndrome: a randomized clinical trial. *Biol. Psychiatry* 44:77–87

Ross JS, Mulvey GK, Hines EM, Nissen SE, Krumholz HM. 2009. Trial publication after registration in ClinicalTrials.Gov: a cross-sectional analysis. *PLoS Med.* 6:e1000144

Ross JS, Tse T, Zarin DA, Xu H, Zhou L, Krumholz HM. 2012. Publication of NIH funded trials registered in ClinicalTrials.gov: cross sectional analysis. *BMJ* 344:d7292

Rutherford BR, Roose SP. 2013. A model of placebo response in antidepressant clinical trials. *Am. J. Psychiatry* 170:723–33

Schneider LS, Nelson JC, Clary CM, Newhouse P, Krishnan KRR. 2004. Sertraline and the Cheshire Cat in geriatric depression/Dr. Schneider and colleagues reply. *Am. J. Psychiatry* 161:759–61

Schulz KF, Altman DG, Moher D, CONSORT Group. 2010. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *PLoS Med.* 7:e1000251

Schulz SC. 2000. *Efficacy of quetiapine versus haloperidol and placebo in the short-term treatment of acute schizophrenia*. Presented at Am. Psychiatr. Assoc. Annu. Meet., Chicago

Schwartz LM, Woloshin S. 2011. Communicating uncertainties about prescription drugs to the public: a national randomized trial. *Arch. Intern. Med.* 171:1463–68

Serretti A, Chiesa A. 2009. Treatment-emergent sexual dysfunction related to antidepressants: a meta-analysis. *J. Clin. Psychopharmacol.* 29:259–66

Shelton RC, Tollefson GD, Tohen M, Stahl S, Gannon KS, et al. 2001. A novel augmentation strategy for treating resistant major depression. *Am. J. Psychiatry* 158:131–34

Siddiqui O. 2003. *Statistical review and evaluation for duloxetine for major depressive disorder.* NDA 21–427. Washington, DC: FDA. **http://www.accessdata.fda.gov/drugsatfda_docs/nda/2004/021427_s000_Cymbalta_Statr.pdf**

Spielmans GI, Berman MI, Linardatos E, Rosenlicht NZ, Perry A, Tsai AC. 2013. Adjunctive atypical antipsychotics for major depressive disorder: a meta-analysis of depression, quality of life, and safety outcomes. *PLoS Med.* 10:e1001403

Spielmans GI, Berman MI, Usitalo AN. 2011. Psychotherapy versus second-generation antidepressants in the treatment of depression: a meta-analysis. *J. Nerv. Ment. Dis.* 199:142–49

Spielmans GI, Biehn TL, Sawrey DL. 2010. A case study of salami slicing: pooled analyses of duloxetine for depression. *Psychother. Psychosom.* 79:97–106

Spielmans GI, McFall JP. 2006. A comparative meta-analysis of clinical global impressions change in antidepressant trials. *J. Nerv. Ment. Dis.* 194:845–52

Spielmans GI, Parry PI. 2010. From evidence-based medicine to marketing-based medicine: evidence from internal industry documents. *J. Bioethical Inq.* 7:13–29

Temple R, Ellenberg SS. 2000. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: ethical and scientific issues. *Ann. Intern. Med.* 133:455–63

Trinh NH, Hoblyn J, Mohanty S, Yaffe K. 2003. Efficacy of cholinesterase inhibitors in the treatment of neuropsychiatric symptoms and functional impairment in Alzheimer disease: a meta-analysis. *JAMA* 289:210–16

Tsai AC, Rosenlicht NZ, Jureidini JN, Parry PI, Spielmans GI, Healy D. 2011. Aripiprazole in the maintenance treatment of bipolar disorder: a critical review of the evidence and its dissemination into the scientific literature. *PLoS Med.* 8:1–13

Tumas J. 1999. *Re: 2 EPS abstracts for APA.* Email corresp. **http://s.wsj.net/public/resources/documents/WSJ-AstraZeneca-Ex14-090227.pdf**

Tumas J. 2000. *FW: meta analyses.* Email corresp. **http://stmedia.startribune.com/documents/Omnibus+MSJ+Exhibit+4.pdf**

Tumber MB, Dickersin K. 2005. Publication of clinical trials: accountability and accessibility. *J. Intern. Med.* 257:65–77

Turner EH. 2013a. How to access and process FDA drug approval packages for use in research. *BMJ.* 347:f5992. **http://www.bmj.com/content/347/bmj.f5992**

Turner EH. 2013b. Publication bias, with a focus on psychiatry: causes and solutions. *CNS Drugs* 27:457–68

Turner EH, Knoepflmacher D, Shapley L. 2012. Publication bias in antipsychotic trials: an analysis of efficacy comparing the published literature to the US Food and Drug Administration database. *PLoS Med.* 9:e1001189

Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. 2008a. Supplementary appendix Supplement to:Turner EH, Matthews AM, Linardatos E, et al. Selective publication of antidepressant trials and its influence on apparent efficacy. *N. Engl. J. Med.* 2008;358:252–60. **http://www.nejm.org/doi/suppl/10.1056/NEJMsa065779/suppl_file/nejm_turner_252sa1.pdf**

Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. 2008b. Selective publication of antidepressant trials and its influence on apparent efficacy. *N. Engl. J. Med.* 358:252–60

Turner EH, Rosenthal R. 2008. Efficacy of antidepressants. *BMJ* 336:516–17

Velligan DI, Newcomer J, Pultz J, Csernansky J, Hoff AL, et al. 2002. Does cognitive function improve with quetiapine in comparison to haloperidol? *Schizophr. Res.* 53:239–48

Viguera AC, Baldessarini RJ, Hegarty JD, van Kammen DP, Tohen M. 1997. Clinical risk following abrupt and gradual withdrawal of maintenance neuroleptic treatment. *Arch. Gen. Psychiatry* 54:49–55

Zimmerman M, Galione JN, Attiullah N, Friedman M, Toba C, et al. 2010. Underrecognition of clinically significant side effects in depressed outpatients. *J. Clin. Psychiatry* 71:484–90

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Clinical Psychology* articles may be
found at http://www.annualreviews.org/errata/clinpsy

ANNUAL REVIEWS

It's about time. Your time. It's time well spent.

New From Annual Reviews:

Annual Review of Organizational Psychology and Organizational Behavior

Editor: **Frederick P. Morgeson,** *The Eli Broad College of Business, Michigan State University*

The *Annual Review of Organizational Psychology and Organizational Behavior* is devoted to publishing reviews of the industrial and organizational psychology, human resource management, and organizational behavior literature. Topics for review include motivation, selection, teams, training and development, leadership, job performance, strategic HR, cross-cultural issues, work attitudes, entrepreneurship, affect and emotion, organizational change and development, gender and diversity, statistics and research methodologies, and other emerging topics.

Complimentary online access to the first volume will be available until March 2015.

Access this and all other Annual Reviews journals via your institution at www.annualreviews.org.

ANNUAL REVIEWS

It's about time. Your time. It's time well spent.

New From Annual Reviews:

Annual Review of Statistics and Its Application

Volume 1 • Online January 2014 • http://statistics.annualreviews.org

Editor: **Stephen E. Fienberg,** *Carnegie Mellon University*

Associate Editors: **Nancy Reid,** *University of Toronto*

Stephen M. Stigler, *University of Chicago*

The *Annual Review of Statistics and Its Application* aims to inform statisticians and quantitative methodologists, as well as all scientists and users of statistics about major methodological advances and the computational tools that allow for their implementation. It will include developments in the field of statistics, including theoretical statistical underpinnings of new methodology, as well as developments in specific application domains such as biostatistics and bioinformatics, economics, machine learning, psychology, sociology, and aspects of the physical sciences.

Complimentary online access to the first volume will be available until January 2015.

Access this and all other Annual Reviews journals via your institution at www.annualreviews.org.