SPECIAL ARTICLE

Selective Publication of Antidepressant Trials and Its Influence on Apparent Efficacy

Erick H. Turner, M.D., Annette M. Matthews, M.D., Eftihia Linardatos, B.S., Robert A. Tell, L.C.S.W., and Robert Rosenthal, Ph.D.

From the Departments of Psychiatry (E.H.T., A.M.M.) and Pharmacology (E.H.T.), Oregon Health and Science University; and the Behavioral Health and Neurosciences Division, Portland Veterans Affairs Medical Center (E.H.T., A.M.M., R.A.T.) — both in Portland, OR; the Department of Psychology, Kent State University, Kent, OH (E.L.); the Department of Psychology, University of California–Riverside, Riverside (R.R.); and Harvard University, Cambridge, MA (R.R.). Address reprint requests to Dr. Turner at Portland VA Medical Center, P3MHDC, 3710 SW US Veterans Hospital Rd., Portland, OR 97239, or at turnere@ohsu.edu.

N Engl J Med 2008;358:252-60.
Copyright © 2008 Massachusetts Medical Society.

ABSTRACT

BACKGROUND

Evidence-based medicine is valuable to the extent that the evidence base is complete and unbiased. Selective publication of clinical trials — and the outcomes within those trials — can lead to unrealistic estimates of drug effectiveness and alter the apparent risk–benefit ratio.

METHODS

We obtained reviews from the Food and Drug Administration (FDA) for studies of 12 antidepressant agents involving 12,564 patients. We conducted a systematic literature search to identify matching publications. For trials that were reported in the literature, we compared the published outcomes with the FDA outcomes. We also compared the effect size derived from the published reports with the effect size derived from the entire FDA data set.

RESULTS

Among 74 FDA-registered studies, 31%, accounting for 3449 study participants, were not published. Whether and how the studies were published were associated with the study outcome. A total of 37 studies viewed by the FDA as having positive results were published; 1 study viewed as positive was not published. Studies viewed by the FDA as having negative or questionable results were, with 3 exceptions, either not published (22 studies) or published in a way that, in our opinion, conveyed a positive outcome (11 studies). According to the published literature, it appeared that 94% of the trials conducted were positive. By contrast, the FDA analysis showed that 51% were positive. Separate meta-analyses of the FDA and journal data sets showed that the increase in effect size ranged from 11 to 69% for individual drugs and was 32% overall.

CONCLUSIONS

We cannot determine whether the bias observed resulted from a failure to submit manuscripts on the part of authors and sponsors, from decisions by journal editors and reviewers not to publish, or both. Selective reporting of clinical trial results may have adverse consequences for researchers, study participants, health care professionals, and patients.

MEDICAL DECISIONS ARE BASED ON AN understanding of publicly reported clinical trials.[1,2] If the evidence base is biased, then decisions based on this evidence may not be the optimal decisions. For example, selective publication of clinical trials, and the outcomes within those trials, can lead to unrealistic estimates of drug effectiveness and alter the apparent risk–benefit ratio.[3,4]

Attempts to study selective publication are complicated by the unavailability of data from unpublished trials. Researchers have found evidence for selective publication by comparing the results of published trials with information from surveys of authors,[5] registries,[6] institutional review boards,[7,8] and funding agencies,[9,10] and even with published methods.[11] Numerous tests are available to detect selective-reporting bias, but none are known to be capable of detecting or ruling out bias reliably.[12-16]

In the United States, the Food and Drug Administration (FDA) operates a registry and a results database.[17] Drug companies must register with the FDA all trials they intend to use in support of an application for marketing approval or a change in labeling. The FDA uses this information to create a table of all studies.[18] The study protocols in the database must prospectively identify the exact methods that will be used to collect and analyze data. Afterward, in their marketing application, sponsors must report the results obtained using the prespecified methods. These submissions include raw data, which FDA statisticians use in corroborative analyses. This system prevents selective post hoc reporting of favorable trial results and outcomes within those trials.

How accurately does the published literature convey data on drug efficacy to the medical community? To address this question, we compared drug efficacy inferred from the published literature with drug efficacy according to FDA reviews.

METHODS

DATA FROM FDA REVIEWS

We identified the phase 2 and 3 clinical-trial programs for 12 antidepressant agents approved by the FDA between 1987 and 2004 (median, August 1996), involving 12,564 adult patients. For the eight older antidepressants, we obtained hard copies of statistical and medical reviews from colleagues who had procured them through the Freedom of Information Act.[19] Reviews for the four newer antidepressants were available on the FDA Web site.[17,20] This study was approved by the Research and Development Committee of the Portland Veterans Affairs Medical Center; because of its nature, informed consent from individual patients was not required.

From the FDA reviews of submitted clinical trials, we extracted efficacy data on all randomized, double-blind, placebo-controlled studies of drugs for the short-term treatment of depression. We included data pertaining only to dosages later approved as safe and effective; data pertaining to unapproved dosages were excluded.

We extracted the FDA's regulatory decisions — that is, whether, for purposes of approval, the studies were judged to be positive or negative with respect to the prespecified primary outcomes (or primary end points).[21] We classified as questionable those studies that the FDA judged to be neither positive nor clearly negative — that is, studies that did not have significant findings on the primary outcome but did have significant findings on several secondary outcomes. Failed studies[22] were also classified as questionable (for more information, see the Methods section of the Supplementary Appendix, available with the full text of this article at www.nejm.org). For fixed-dose studies (studies in which patients are randomly assigned to receive one of two or more dose levels or placebo) with a mix of significant and nonsignificant results for different doses, we used the FDA's stated overall decisions on the studies. We used double data extraction and entry, as detailed in the Methods section of the Supplementary Appendix.

DATA FROM JOURNAL ARTICLES

Our literature-search strategy consisted of the following steps: a search of articles in PubMed, a search of references listed in review articles, and a search of the Cochrane Central Register of Controlled Trials; contact by telephone or e-mail with the drug sponsor's medical-information department; and finally, contact by means of a certified letter sent to the sponsor's medical-information department, including a deadline for responding in writing to our query about whether the study results had been published. If these steps failed to reveal any publications, we concluded that the study results had not been published.

We identified the best match between the FDA-reviewed clinical trials and journal articles on the basis of the following information: drug name, dose groups, sample size, active comparator (if used), duration, and name of principal investigator. We sought published reports on individual studies; articles covering multiple studies were excluded. When the results of a trial were reported in two or more primary publications, we selected the first publication.

Few journal articles used the term "primary efficacy outcome" or a reasonable equivalent. Therefore, we identified the apparent primary efficacy outcome, or the result highlighted most prominently, as the drug–placebo comparison reported first in the text of the results section or in the table or figure first cited in the text. As with the FDA reviews, we used double data extraction and entry (see the Methods section of the Supplementary Appendix for details).

STATISTICAL ANALYSIS

We categorized the trials on the basis of the FDA regulatory decision, whether the trial results were published, and whether the apparent primary outcomes agreed or conflicted with the FDA decision. We calculated risk ratios with exact 95% confidence intervals and Pearson's chi-square analysis, using Stata software, version 9. We used a similar approach to examine the numbers of patients within the studies. Sample sizes were compared between published and unpublished studies with the use of the Wilcoxon rank-sum test.

For our major outcome indicator, we calculated the effect size for each trial using Hedges's g — that is, the difference between two means divided by their pooled standard deviation.[23] However, because means and standard deviations (or standard errors) were inconsistently reported in both the FDA reviews and the journal articles, we used the algebraically equivalent computational equation[24]:

$$g = t \times \sqrt{\frac{1}{n_{drug}} + \frac{1}{n_{placebo}}}\ .$$

We calculated the t statistic[25] using the precise P value and the combined sample size as arguments in Microsoft Excel's TINV (inverse T) function, multiplying t by –1 when the study drug was inferior to the placebo. Hedges's correction for small sample size was applied to all g values.[26]

Precise P values were not always available for the above calculation. Rather, P values were often indicated as being below or above a certain threshold — for example, $P<0.05$ or "not significant" (i.e., $P>0.05$). In these cases, we followed the procedure described in the Supplementary Appendix.

For each fixed-dose (multiple-dose) study, we computed a single study-level effect size weighted by the degrees of freedom for each dose group. On the basis of the study-level effect-size values for both fixed-dose and flexible-dose studies, we calculated weighted mean effect-size values for each drug and for all drugs combined, using a random-effects model with the method of DerSimonian and Laird[27] in Stata.[28]

Within the published studies, we compared the effect-size values derived from the journal articles with the corresponding effect-size values derived from the FDA reviews. Next, within the FDA data set, we compared the effect-size values for the published studies with the effect-size values for the unpublished studies. Finally, we compared the journal-based effect-size values with those derived from the entire FDA data set — that is, both published and unpublished studies.

We made these comparisons at the level of studies and again at the level of the 12 drugs. Because the data were not normally distributed, we used the nonparametric rank-sum test for unpaired data and the signed-rank test for paired data. In these analyses, all the effect-size values were given equal weight.

RESULTS

STUDY OUTCOME AND PUBLICATION STATUS

Of the 74 FDA-registered studies in the analysis we could not find evidence of publication for 23 (31%) (Table 1). The difference between the sample sizes for the published studies (median, 153 patients) and the unpublished studies (median, 146 patients) was neither large nor significant (5% difference between medians; $P=0.29$ by the rank-sum test).

The data in Table 1 are displayed in terms of the study outcome in Figure 1A. The questions of whether the studies were published and, if so, how the results were reported were strongly related to their overall outcomes. The FDA deemed 38 of the 74 studies (51%) positive, and all but 1 of the 38 were published. The remaining 36 studies (49%) were deemed to be either negative (24 studies) or questionable (12). Of these 36 studies, 3 were pub-

lished as not positive, whereas the remaining 33 either were not published (22 studies) or were published, in our opinion, as positive (11) and therefore conflicted with the FDA's conclusion. Overall, the studies that the FDA judged as positive were approximately 12 times as likely to be published in a way that agreed with the FDA analysis as were studies with nonpositive results according to the FDA (risk ratio, 11.7; 95% confidence interval [CI], 6.2 to 22.0; P<0.001). This association of publication status with study outcome remained significant when we excluded questionable studies and when we examined publication status without regard to whether the published conclusions and the FDA conclusions were in agreement (for details, see the Supplementary Appendix).

Overall, 48 of the 51 published studies were reported to have positive results (94%; binomial 95% CI, 84 to 99). According to the FDA, 38 of the 74 registered studies had positive results (51%; 95% CI, 39 to 63). There was no overlap between these two sets of confidence intervals.

These data are broken down by drug and study number in Figure 2A. For each of the 12 drugs, the results of at least one study either were unpublished or were reported in the literature as positive despite a conflicting judgment by the FDA.

NUMBER OF STUDY PARTICIPANTS

As shown in Table 1, a total of 12,564 patients participated in these trials. The data from 3449 patients (27%) were not published. Data from an additional 1843 patients (15%) were reported in journal articles in which the highlighted finding conflicted with the FDA-defined primary outcome. Thus, the percentages for the patients closely mirrored those for the studies (Table 1).

Whether a patient's data were reported in a way that was in concert with the FDA review was associated with the study outcome (Fig. 1B) (risk ratio, 27.1), which was consistent with the above-reported finding with the studies. Figure 2B shows these same data according to the drug being evaluated.

QUALITATIVE DESCRIPTION OF SELECTIVE REPORTING WITHIN TRIALS

The methods reported in 11 journal articles appear to depart from the prespecified methods reflected in the FDA reviews (Table B of the Supplementary Appendix). Although for each of these studies the finding with respect to the protocol-specified primary outcome was nonsignificant, each publication highlighted a positive result as if it were the primary outcome. The nonsignificant results for the prespecified primary outcomes were either subordinated to nonprimary positive results (in two reports) or omitted (in nine). (Study-level methodologic differences are detailed in the footnotes to Table B of the Supplementary Appendix.)

Table 1. Overall Publication Status of FDA-Registered Antidepressant Studies.

Publication Status	No. of Studies (%)	No. of Patients in Studies (%)
Published results agree with FDA decision	40 (54)	7,272 (58)
Published results conflict with FDA decision (published as positive)	11 (15)	1,843 (15)
Results not published	23 (31)	3,449 (27)
Total	74 (100)	12,564 (100)

EFFECT SIZE

The effect-size values derived from the journal reports were often greater than those derived from the FDA reviews. The difference between these two sets of values was significant whether the studies (P=0.003) or the drugs (P=0.012) were used as the units of analysis (see Table D in the Supplementary Appendix).

The effect sizes of the published and unpublished studies reviewed by the FDA are compared in Figure 3A. The overall mean weighted effect-size value was 0.37 (95% CI, 0.33 to 0.41) for published studies and 0.15 (95% CI, 0.08 to 0.22) for unpublished studies. The difference was significant whether the studies (P<0.001) or the drugs (P=0.005) were used as the units of analysis (Table D in the Supplementary Appendix).

The mean effect-size values for all FDA studies, both published and unpublished, are compared with those for all published studies, as shown in Figure 3B. Again, the differences were significant whether the studies (P<0.001) or the drugs (P=0.002) were used as units of analysis (Table D in the Supplementary Appendix).

For each of the 12 drugs, the effect size derived from the journal articles exceeded the effect size derived from the FDA reviews (sign test, P<0.001) (Fig. 3B). The magnitude of the increases in effect size between the FDA reviews and the published reports ranged from 11 to 69%, with a median

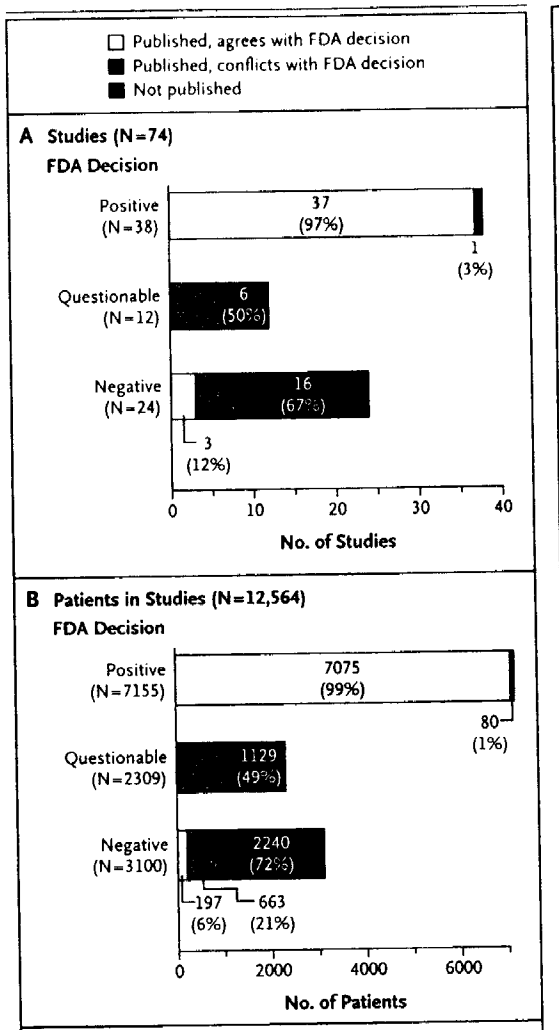


Figure 1. Effect of FDA Regulatory Decisions on Publication.

Among the 74 studies reviewed by the FDA (Panel A), 38 were deemed to have positive results, 37 of which were published with positive results; the remaining study was not published. Among the studies deemed to have questionable or negative results by the FDA, there was a tendency toward nonpublication or publication with positive results, conflicting with the conclusion of the FDA. Among the 12,564 patients in all 74 studies (Panel B), data for patients who participated in studies deemed positive by the FDA were very likely to be published in a way that agreed with the FDA. In contrast, data for patients participating in studies deemed questionable or negative by the FDA tended either not to be published or to be published in a way that conflicted with the FDA's judgment.

increase of 32%. A 32% increase was also observed in the weighted mean effect size for all drugs combined, from 0.31 (95% CI, 0.27 to 0.35) to 0.41 (95% CI, 0.36 to 0.45).

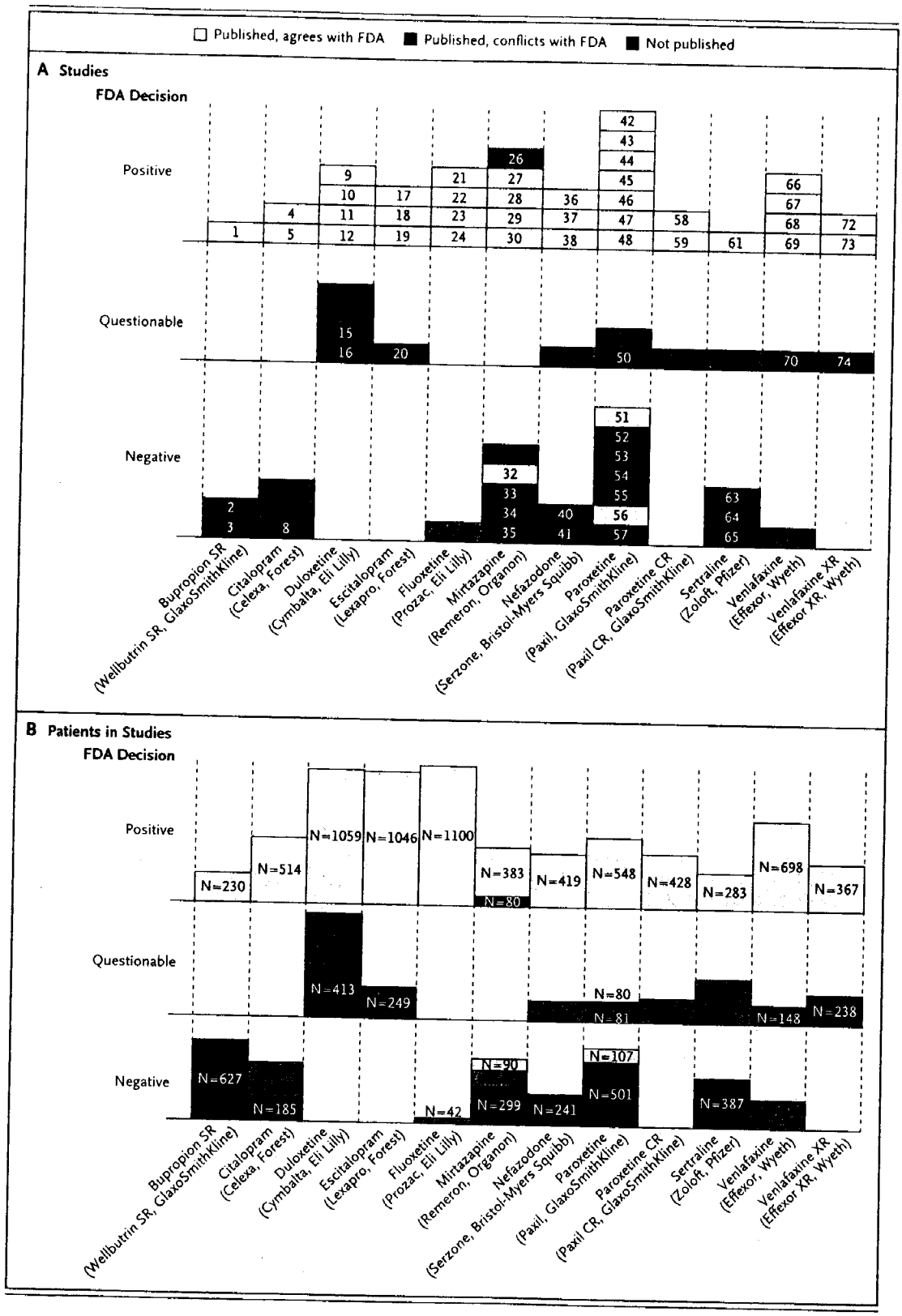
Figure 2 (facing page). Publication Status and FDA Regulatory Decision by Study and by Drug.

Panel A shows the publication status of individual studies. Nearly every study deemed positive by the FDA (top row) was published in a way that agreed with the FDA's judgment. By contrast, most studies deemed negative (bottom row) or questionable (middle row) by the FDA either were published in a way that conflicted with the FDA's judgment or were not published. Numbers shown in boxes indicate individual studies and correspond to the study numbers listed in Table A of the Supplementary Appendix. Panel B shows the numbers of patients participating in the individual studies indicated in Panel A. Data for patients who participated in studies deemed positive by the FDA were very likely to be published in a way that agreed with the FDA's judgment. By contrast, data for patients who participated in studies deemed negative or questionable by the FDA tended either not to be published or to be published in a way that conflicted with the FDA's judgment.

A list of the study-level effect-size values used in the above analyses — derived from both the FDA reviews and the published reports — is provided in Table C of the Supplementary Appendix. These effect-size values are based on P values and sample sizes shown in Table A of the Supplementary Appendix, which also lists reference information for the publications consulted.

DISCUSSION

We found a bias toward the publication of positive results. Not only were positive results more likely to be published, but studies that were not positive, in our opinion, were often published in a way that conveyed a positive outcome. We analyzed these data in terms of the proportion of positive studies and in terms of the effect size associated with drug treatment. Using both approaches, we found that the efficacy of this drug class is less than would be gleaned from an examination of the published literature alone. According to the published literature, the results of nearly all of the trials of antidepressants were positive. In contrast, FDA analysis of the trial data showed that roughly half of the trials had positive results. The statistical significance of a study's results was strongly associated with whether and how they were reported, and the association was independent of sample size. The study outcome also affected the chances that the data from a participant would be published. As a result of selec-


Published, agrees with FDA
Published, conflicts with FDA
Not published
A Studies
FDA Decision
Positive
Questionable
Negative
B Patients in Studies
FDA Decision
Positive
Questionable
Negative
N=230
N=514
N=1059
N=1046
N=1100
N=383
N=80
N=419
N=548
N=428
N=283
N=698
N=367
N=413
N=249
N=80
N=81
N=148
N=238
N=627
N=185
N=42
N=90
N=299
N=241
N=107
N=501
N=387
Bupropion SR (Wellbutrin SR, GlaxoSmithKline)
Citalopram (Celexa, Forest)
Duloxetine (Cymbalta, Eli Lilly)
Escitalopram (Lexapro, Forest)
Fluoxetine (Prozac, Eli Lilly)
Mirtazapine (Remeron, Organon)
Nefazodone (Serzone, Bristol-Myers Squibb)
Paroxetine (Paxil, GlaxoSmithKline)
Paroxetine CR (Paxil CR, GlaxoSmithKline)
Sertraline (Zoloft, Pfizer)
Venlafaxine (Effexor, Wyeth)
Venlafaxine XR (Effexor XR, Wyeth)

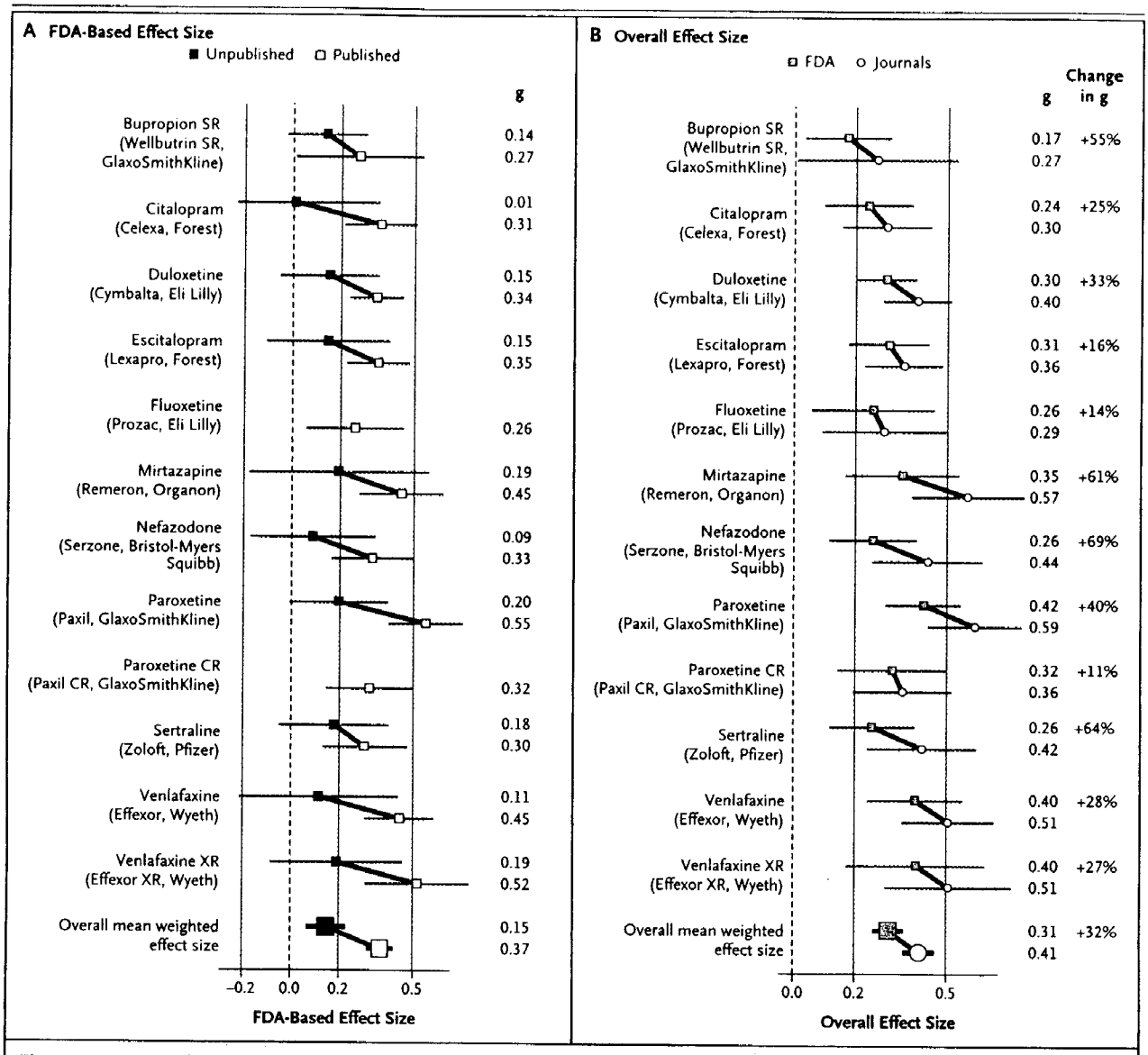


Figure 3. Mean Weighted Effect Size According to Drug, Publication Status, and Data Source.

Values for effect size are expressed as Hedges's g (the difference between two means divided by their pooled standard deviation). Effect-size values of 0.2 and 0.5 are considered to be small and medium, respectively.[29] Effect-size values for unpublished studies and published studies, as extracted from data in FDA reviews, are shown in Panel A. Horizontal lines indicate 95% confidence intervals. There were no unpublished studies for controlled-release paroxetine or fluoxetine. For each of the other antidepressants, the effect size for the published subgroup of studies was greater than the effect size for the unpublished subgroup of studies. Overall effect-size values (i.e., based on data from the FDA for published and unpublished studies combined), as compared with effect-size values based on data from corresponding published reports, are shown in Panel B. For each drug, the effect-size value based on published literature was higher than the effect-size value based on FDA data, with increases ranging from 11 to 69%. For the entire drug class, effect sizes increased by 32%.

tive reporting, the published literature conveyed an effect size nearly one third larger than the effect size derived from the FDA data.

Previous studies have examined the risk–benefit ratio for drugs after combining data from regulatory authorities with data published in journals.[3,30-32] We built on this approach by comparing study-level data from the FDA with matched data from journal articles. This comparative approach allowed us to quantify the effect of selective publication on apparent drug efficacy.

Our findings have several limitations: they are

restricted to antidepressants, to industry-sponsored trials registered with the FDA, and to issues of efficacy (as opposed to "real-world" effectiveness[33]). This study did not account for other factors that may distort the apparent risk–benefit ratio, such as selective publication of safety issues, as has been reported with rofecoxib (Vioxx, Merck)[34] and with the use of selective serotonin-reuptake inhibitors for depression in children.[3] Because we excluded articles covering multiple studies, we probably counted some studies as unpublished that were — technically — published. The practice of bundling negative and positive studies in a single article has been found to be associated with duplicate or multiple publication,[35] which may also influence the apparent risk–benefit ratio.

There can be many reasons why the results of a study are not published, and we do not know the reasons for nonpublication. Thus, we cannot determine whether the bias observed resulted from a failure to submit manuscripts on the part of authors and sponsors, decisions by journal editors and reviewers not to publish submitted manuscripts, or both.

We wish to clarify that nonsignificance in a single trial does not necessarily indicate lack of efficacy. Each drug, when subjected to meta-analysis, was shown to be superior to placebo. On the other hand, the true magnitude of each drug's superiority to placebo was less than a diligent literature review would indicate.

We do not mean to imply that the primary methods agreed on between sponsors and the FDA are necessarily preferable to alternative methods. Nevertheless, when multiple analyses are conducted, the principle of prespecification controls the rate of false positive findings (type I error), and it prevents HARKing,[36] or hypothesizing after the results are known.

It might be argued that some trials did not merit publication because of methodologic flaws, including problems beyond the control of the investigator. However, since the protocols were written according to international guidelines for efficacy studies[37] and were carried out by companies with ample financial and human resources, to be fair to the people who put themselves at risk to participate, a cogent public reason should be given for failure to publish.

Selective reporting deprives researchers of the accurate data they need to estimate effect size realistically. Inflated effect sizes lead to underestimates of the sample size required to achieve statistical significance. Underpowered studies — and selectively reported studies in general — waste resources and the contributions of investigators and study participants, and they hinder the advancement of medical knowledge. By altering the apparent risk–benefit ratio of drugs, selective publication can lead doctors to make inappropriate prescribing decisions that may not be in the best interest of their patients and, thus, the public health.

Dr. Turner reports having served as a medical reviewer for the Food and Drug Administration. No other potential conflict of interest relevant to this article was reported.

We thank Emily Kizer, Marcus Griffith, and Tammy Lewis for clerical assistance; David Wilson, Alex Sutton, Ohidul Siddiqui, and Benjamin Chan for statistical consultation; Linda Ganzini, Thomas B. Barrett, and Daniel Hilfet-Hilliker for their comments on an earlier version of this manuscript; Arifula Khan, Kelly Schwartz, and David Antonuccio for providing access to FDA reviews; Thomas B. Barrett, Norwan Moaleji and Samantha Ruimy for double data extraction and entry; and Andrew Hamilton for literature database searches.

REFERENCES

1. Hagdrup N, Falshaw M, Gray RW, Carter Y. All members of primary care team are aware of importance of evidence based medicine. BMJ 1998;317:282.
2. Craig JC, Irwig LM, Stockler MR. Evidence-based medicine: useful tools for decision making. Med J Aust 2001;174:248-53.
3. Whittington CJ, Kendall T, Fonagy P, Cottrell D, Cotgrove A, Boddington E. Selective serotonin reuptake inhibitors in childhood depression: systematic review of published versus unpublished data. Lancet 2004;363:1341-5.
4. Kyzas PA, Loizou KT, Ioannidis JP. Selective reporting biases in cancer prognostic factor studies. J Natl Cancer Inst 2005;97:1043-55.
5. Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H Jr. Publication bias and clinical trials. Control Clin Trials 1987;8:343-53.
6. Simes RJ. Confronting publication bias: a cohort design for meta-analysis. Stat Med 1987;6:11-29.
7. Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. BMJ 1997;315:640-5.
8. Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA 2004;291:2457-65.
9. Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. JAMA 1998;279:281-6.
10. Chan AW, Krleza-Jerić K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. CMAJ 2004;171:735-40.
11. Chan AW, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. BMJ 2005;330:753.
12. Lau J, Ioannidis JP, Terrin N, Schmid CH, Olkin I. The case of the misleading funnel plot. BMJ 2006;333:597-600.
13. Hayashino Y, Noguchi Y, Fukui T. Systematic evaluation and comparison of statistical tests for publication bias. J Epidemiol 2005;15:235-43.
14. Pham B, Platt R, McAuley L, Klassen TP, Moher D. Is there a "best" way to detect and minimize publication bias? An empirical evaluation. Eval Health Prof 2001;24:109-25.

15. Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. J Clin Epidemiol 2000; 53:1119-29.
16. Ioannidis JP, Trikalinos TA. The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. CMAJ 2007;176:1091-6.
17. Turner EHA. A taxpayer-funded clinical trials registry and results database. PLoS Med 2004;1(3):e60.
18. Center for Drug Evaluation and Research. Manual of policies and procedures: clinical review template. Rockville, MD: Food and Drug Administration, 2004. (Accessed December 20, 2007, at http://www.fda.gov/cder/mapp/6010.3.pdf.)
19. Committee on Government Reform, U.S. House of Representatives, 109th Congress, 1st Session. A citizen's guide on using the Freedom of Information Act and the Privacy Act of 1974 to request government records. Report no. 109-226. Washington, DC: Government Printing Office, 2005. (Also available at: http://www.fas.org/sgp/foia/citizen.pdf.)
20. Center for Drug Evaluation and Research. Drugs@FDA: FDA approved drug products. Rockville, MD: Food and Drug Administration. (Accessed December 20, 2007, at http://www.accessdata.fda.gov/scripts/cder/drugsatfda.)
21. International Conference on Harmonisation (ICH), European Medicines Agency (EMEA). Topic E9: statistical principles for clinical trials. Rockville, MD: Food and Drug Administration. (Accessed December 20, 2007, at http://www.fda.gov/cder/guidance/iche3.pdf.)
22. Temple R, Ellenberg SS. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: ethical and scientific issues. Ann Intern Med 2000;133:455-63.
23. Hedges LV. Estimation of effect size from a series of independent experiments. Psychol Bull 1982;92:490-9.
24. Rosenthal R. Meta-analytic procedures for social research. Newbury Park, CA: Sage, 1991.
25. Whitley E, Ball J. Statistics review 5: comparison of means. Crit Care 2002;6: 424-8.
26. Hedges LV, Olkin I. Statistical methods for meta-analysis. New York: Academic Press, 1985.
27. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7:177-88.
28. Stata statistical software, release 9. College Station, TX: StataCorp, 2005.
29. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. New York: Lawrence Erlbaum Associates, 1988.
30. Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. N Engl J Med 2007;356:2457-71. [Erratum, N Engl J Med 2007;357:100.]
31. Nissen SE, Wolski K, Topol EJ. Effect of muraglitazar on death and major adverse cardiovascular events in patients with type 2 diabetes mellitus. JAMA 2005;294: 2581-6.
32. Sackner-Bernstein JD, Kowalski M, Fox M, Aaronson K. Short-term risk of death after treatment with nesiritide for decompensated heart failure: a pooled analysis of randomized controlled trials. JAMA 2005;293:1900-5.
33. Revicki DA, Frank L. Pharmacoeconomic evaluation in the real world: effectiveness versus efficacy studies. Pharmacoeconomics 1999;15:423-34.
34. Topol EJ. Failing the public health — rofecoxib, Merck, and the FDA. N Engl J Med 2004;351:1707-9.
35. Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence b(i)ased medicine — selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. BMJ 2003;326:1171-3.
36. Kerr NL. HARKing: hypothesizing after the results are known. Pers Soc Psychol Rev 1998;2:196-217.
37. International Conference on Harmonisation — Efficacy. Rockville, MD: Food and Drug Administration. (Accessed December 20, 2007, at http://www.fda.gov/cder/guidance/#ICH_efficacy.)

Copyright © 2008 Massachusetts Medical Society.

VIEW CURRENT JOB POSTINGS AT THE NEJM CAREERCENTER
Visit our online CareerCenter for physicians at **www.nejmjobs.org** to see the expanded features and services available. Physicians can conduct a quick search of the public database by specialty and view hundreds of current openings that are updated daily online at the CareerCenter.