RESEARCH

Effect of reporting bias on meta-analyses of drug trials: reanalysis of meta-analyses

OPEN ACCESS

Beth Hart *Doris Duke clinical research fellow and medical student*[1], Andreas Lundh *PhD student*[2], Lisa Bero *professor*[1]

[1]Department of Clinical Pharmacy, Institute for Health Policy Studies, University of California, San Francisco, 3333 California St, Suite 420, San Francisco, CA 94118, USA; [2]Nordic Cochrane Centre, Rigshospitalet and University of Copenhagen, Copenhagen, Denmark

Abstract

Objective To investigate the effect of including unpublished trial outcome data obtained from the Food and Drug Administration (FDA) on the results of meta-analyses of drug trials.

Design Reanalysis of meta-analyses.

Data sources Drug trials with unpublished outcome data for new molecular entities that were approved by the FDA between 2001 and 2002 were identified. For each drug, eligible systematic reviews containing at least one meta-analysis were identified by searches of Medline, Embase, and the Cochrane Library in November 2010.

Selection criteria Eligible systematic reviews were done after FDA approval of the drug, were published in English, and had outcomes and comparators that were the same as those of the trials with unpublished FDA trial outcomes, and the characteristics of participants in the systematic reviews were consistent with the FDA approved indication for the drug. Clinical guidelines, conference proceedings, duplicate systematic reviews, and systematic reviews in which included trials were not referenced or that combined trials across multiple drug classes were excluded. Systematic reviews using non-standard meta-analytic techniques (such as Bayesian or network meta-analyses) and those that used inappropriate or invalid methods for calculation of summary statistics (such as unweighted pooled analyses) were also excluded.

Data extraction Two authors independently extracted data from both the published systematic reviews and the FDA's medical and statistical reviews of the trials submitted to FDA.

Main outcome measure Summary statistics (risk ratios, odds ratios, or weighted mean differences) for relevant outcomes with and without unpublished FDA trial data.

Results 42 meta-analyses (41 efficacy outcomes, one harm outcome) for nine drugs across six drug classes were reanalysed. Overall, addition of unpublished FDA trial data caused 46% (19/41) of the summary estimates from the meta-analyses to show lower efficacy of the drug, 7% (3/41) to show identical efficacy, and 46% (19/41) to show greater efficacy. The summary estimate of the single harm outcome showed more harm from the drug after inclusion of unpublished FDA trial data.

Conclusion The effect of including unpublished FDA trial outcome data varies by drug and outcome. Unpublished FDA trial outcome data should be available and included in meta-analysis. Making these data easily accessible is particularly important because the effects of including unpublished data vary.

Introduction

Systematic reviews or meta-analyses of clinical trials are one of the foundations of healthcare and clinical practice guidelines informed by evidence.[1 2] Bias in the design, conduct, or reporting of clinical trials can result in inaccuracies in meta-analyses or guidelines, and subsequent errors in clinical practice. Reporting bias can take multiple forms, including publication bias, which is the tendency for published trials to be more likely to report statistically significant results than non-significant results.[3 4] Outcome reporting bias, or the selective publication of some but not all outcome data from a trial, is another form of reporting bias.[5]

Much of the evidence of reporting bias has been found in trials testing the efficacy of new drugs. Reporting bias has been detected by comparing publications of drug trials in the scientific literature with trial results submitted to drug regulatory authorities, trial protocols, and internal trial reports obtained through litigation.[6-9] Comparison of the published papers with the unpublished data found that entire trials were not reported and that when trials were reported, outcomes in the published report were deleted, added, or changed compared with the unpublished data. Outcome data that favour the efficacy of the drug are more likely to be published.[8-10]

When unfavourable results of drug trials are not published, meta-analyses and systematic reviews that are based only on published data may overestimate the efficacy of the drugs. Little is known about the effect of reporting bias on systematic reviews in general. The Outcome Reporting Bias in Trials (ORBIT) study examined the prevalence of outcome reporting bias in trials included in Cochrane reviews and found that

approximately half of the reviews did not seem to include all data from the relevant trials.[11] Furthermore, in about a quarter of the reviews, treatment effects from meta-analyses were reduced by 20% or more by adjustment for outcome reporting bias.[11]

The effect of unpublished data on the results of meta-analyses has been studied extensively for antidepressants. Inclusion of unpublished outcome data from trials of antidepressants in meta-analyses decreased the efficacy and increased the harms of the drugs.[12 13] Whether these findings about the effect of reporting bias on meta-analyses of antidepressants are generalisable to other classes of drugs is not known.

We have previously shown that reporting bias exists across a variety of drug classes.[8] This study expands on our previous work by investigating whether the selective publication of drug efficacy trials submitted to the Food and Drug Administration (FDA) affects a key "downstream" aspect of the medical literature, meta-analyses. Given that very little published information exists on newly approved drugs, the failure to include unpublished trial data in meta-analyses could make them particularly vulnerable to overestimated or imprecise treatment effects. On the other hand, obtaining and extracting unpublished trial data from drug regulatory authorities with current procedures is time consuming and difficult and may not be feasible owing to the poor reporting of the data. Therefore, assessing the effect of such data on meta-analyses across a variety of drug classes is important.

We investigated the effect of including unpublished data obtained from the FDA on the results of meta-analyses of trials of drugs. We hypothesised that inclusion of unpublished data in meta-analyses would decrease drugs' efficacy and increase their harms compared with meta-analyses that did not include the unpublished data.

Methods

Sample of drugs with unpublished outcome data

We identified drug trials with unpublished outcome data in our previous study of new molecular entities that were approved by the FDA between 2001 and 2002.[8] We chose this time period to allow sufficient time for publication of the trials. New molecular entities contain novel active ingredients never before marketed in the United States. In contrast, "me-too" drugs, which are very similar to existing drugs, or combinations of previously approved drugs are not considered new molecular entities. Because of the novelty of new molecular entities, prescribers need complete and valid information on their efficacy and safety. The trial data supporting the efficacy and safety of new molecular entities is submitted in new drug applications. The reviews of these data are summarised in the FDA's medical and statistical reviews, which are publicly available, along with approval letters, at www.accessdata.fda.gov/scripts/cder/drugsatfda.

By comparing the FDA's medical and statistical reviews of the submitted data from drug trials with published trial reports, we identified 299 unpublished outcomes for 24 drugs. The unpublished outcomes resulted from lack of publication of entire trials (n=34 unpublished trials with 258 outcomes), as well as unpublished outcomes from published trial reports (n=41 unpublished outcomes). We will refer to these outcomes as unpublished FDA trial outcomes.

Identification of systematic reviews containing at least one meta-analysis

For each drug, we identified systematic reviews containing at least one meta-analysis (which we will call "systematic reviews") by searching Medline, Embase, the Cochrane Database of Systematic Reviews, and the Database of Abstracts of Reviews of Effects (DARE) in November 2010. We tailored the search strategy for each drug by combining terms for the drug's name (proprietary, generic, and international non-proprietary name), the FDA approved indication for the drug, and a previously validated search filter for the identification of systematic reviews and meta-analyses.[14 15]

Two authors independently screened titles and abstracts to assess eligibility. We retrieved full papers for relevant studies. Discrepancies were resolved by discussion among all authors.

Selection of systematic reviews for assessing effect of unpublished FDA trial outcomes

We included systematic reviews that contained at least one meta-analysis, were done after FDA approval of the drug (so any unpublished FDA trial outcome data would have been available), were published in English, had one or more outcomes that were the same as unpublished FDA trial outcomes, in which the comparator (active or placebo) was the same as the comparator in trials with unpublished FDA outcomes, and in which the characteristics of participants included were consistent with the FDA approved indication. We also required that the FDA trials with unpublished outcomes met all the inclusion and exclusion criteria of the systematic review.

We excluded clinical guidelines, conference proceedings, duplicate systematic reviews (in the event of duplicates, we included the first published review), and systematic reviews in which included trials could not be identified or that combined trials across multiple drug classes. We also excluded systematic reviews that used non-standard meta-analytic techniques that we would not be able to reproduce (such as Bayesian, network, or individual patient data meta-analyses) and those that used inappropriate or invalid methods for calculating summary statistics (such as unweighted pooled analyses).

For some drugs, multiple systematic reviews met our eligibility criteria. However, we did not consider systematic reviews of a particular drug to be independent of one another, as the trials they included often overlapped. For this reason, we used the following decision tree to select a single systematic review for each drug: firstly, we selected the systematic review with the greatest number of included trials; if tied, secondly, we selected the systematic review with the greatest availability of data for extraction (that is, whether the effect size and variance were reported for each trial included in the relevant meta-analyses); if still tied, thirdly, we selected the systematic review with the greatest number of relevant meta-analyses; if still tied, finally, we selected the systematic review containing meta-analyses with the largest sample sizes. We designed our decision tree in a way that emphasised both practical considerations and the selection of more robust meta-analyses with regard to the amount of evidence synthesised.

Data extraction from systematic reviews and FDA reviews

From each included systematic review, we selected all the meta-analyses that could potentially include the unpublished FDA trial data. Two authors independently extracted data from both the published meta-analyses and the FDA's reviews of the

submitted trials, which are publicly available on its website (www.accessdata.fda.gov/scripts/cder/drugsatfda).

From the published meta-analyses, we extracted data that we needed to recalculate the relevant meta-analyses (events, number of patients in each group, means, standard deviations, and so on). From the FDA reviews, we extracted data by using the same methods as described in the published meta-analysis to obtain the same data the authors of the meta-analysis would have extracted had the unpublished FDA trial outcome data been included in the meta-analysis. For example, we based use of intention to treat versus a per protocol analysis, definition of the intention to treat population, drug dosages, and time points extracted on methods described by the authors of the meta-analysis. If individual meta-analyses were done for different time points, dosages, or analytical techniques (such as intention to treat versus per protocol) for the same outcomes (for example, pain relief at one hour and two hours), we regarded each as a separate outcome for reanalysis.

We also extracted the following data: type of outcome (primary, secondary, not specified) stated in the meta-analysis and FDA review, the type of journal where the systematic review was published (dichotomised as medical journal or Cochrane Library), and the year of publication. We contacted authors of the published systematic reviews to request missing information. We had to contact two authors, and they provided all the requested information.

Analysis

For the systematic reviews selected for recalculation (one per drug), we calculated summary statistics (risk ratios, odds ratios, or weighted mean differences) and the I^2 statistic (a measure of heterogeneity) for each meta-analysis with relevant outcomes both with and without unpublished FDA trial data. If the published meta-analysis already included unpublished data, we removed them and recalculated the summary statistic. We coded the resulting recalculated meta-analysis for each outcome, both with and without unpublished data, as favourable to the drug if it was statistically significant in favour of the FDA reviewed drug in the direction of greater efficacy or less harm (P<0.05, 95% confidence interval for difference excluding 0, or 95% confidence interval for ratio excluding 1); not favourable to the drug if it was statistically significant in favour of the comparator; null if it was not statistically significant; or unknown. We also coded a meta-analysis as favourable if it was a superiority outcome that was statistically significant in favour of the FDA reviewed drug or a non-inferiority or equivalence outcome for which the FDA reviewed drug and comparator had similar effects. We considered superiority outcomes that were not statistically significant or non-inferiority or equivalence outcomes that favoured the comparator to be not favourable. We reported the magnitude of the change in the result of the meta-analysis as a percentage change in the summary statistic after inclusion of all unpublished FDA trial data. For risk ratios and odds ratios, we calculated the percentage change of the log transformation as (log(E)−log(I))×100/log(E), where E=effect estimate excluding unpublished data and I=effect estimate including unpublished data. We calculated the log transformation for relative risks and odds ratios so that the point of "no effect" was equal to zero instead of 1, thus allowing for a calculation of percentage change. For weighted mean differences, we calculated the percentage change by using the formula (E−I)×100/E.

We reported the direction of the change in the results of the meta-analyses as showing an increase in efficacy when the point estimate of the summary statistic showed the drug to be more efficacious when the unpublished FDA trial data were included, a decrease in efficacy when the point estimate of the summary statistic showed the drug to be less efficacious when the unpublished FDA trial data were included, or no change (to two decimal places). For the single safety outcome, we noted if addition of the unpublished FDA trial data changed the point estimate to show more or less harm. We also noted any changes in statistical significance.

We calculated the proportion of unpublished FDA data in each recalculated meta-analysis by dividing the number of patients included from unpublished FDA trials by the total number of patients included in the meta-analysis.

We used RevMan 5.1 software to reanalyse each meta-analysis. We replicated the published meta-analysis with regard to the statistical method (Peto, Mantel-Haenszel, inverse variance), strategies for assessing heterogeneity, analysis model (fixed *v* random effects), and measure of effect (risk ratio, odds ratio, weighted mean difference). With regard to strategies for assessing heterogeneity, if the authors of the meta-analysis applied the results of a test of heterogeneity to determine their analysis model (for example, χ^2<0.1 will result in the application of a random effects analysis model), we would apply that same rule.

Imputation of data

When standard deviations for continuous outcomes were unavailable in the FDA trial data, we explored the possibility of imputing standard deviations from other statistical information in the FDA reviews. For one trial of aripiprazole, we imputed standard deviations as the pooled standard deviations that would produce the reported P values in a *t* test given the reported means and numbers of participants from the FDA trial data. For olmesartan, we imputed standard deviations for two trials by using the same methods for imputing data that were described in the published systematic review.

Sensitivity analysis

When the FDA trial data contained multiple analyses for a particular outcome and the methods of the meta-analysis did not use the most conservative method, we did a sensitivity analysis. For example, if the authors of the meta-analysis included per protocol data for their primary analysis, we would also extract per protocol data for our primary analysis. We would then extract intention to treat data for our sensitivity analysis to determine if the summary statistic was sensitive to the more conservative estimate of the effect of treatment.

Results

Selection of systematic reviews and drugs

Our search identified 1825 unique citations (figure⇓). After screening the titles and abstracts, we retrieved and screened 296 articles in full text. Of the full text articles screened, we excluded 259 because they did not meet our eligibility criteria, leaving a total sample of 37 systematic reviews (figure⇓). These 37 systematic reviews included nine drugs from our original sample of 24 drugs with unpublished FDA trial outcome data. For 15 drugs, we did not identify any systematic reviews that met our eligibility criteria. Our final sample, therefore, contained nine drugs: three for migraine, two antipsychotics, and one each for dementia/Alzheimer's disease, antihypertensive, antibiotic, and topical anti-inflammatory indications.

For eight of the nine drugs, more than one systematic review met our eligibility criteria. For six of these drugs, we selected the systematic review with the greatest number of included trials for reanalysis. For two of the drugs, systematic reviews were tied as to the greatest number of included trials and we selected the one with the most complete available data (that is, the effect size and variance was listed for each trial included in the relevant meta-analyses). We thus selected one systematic review for each of the nine drugs.

Characteristics of systematic reviews

As shown in table 1⇓, four of the nine systematic reviews included in the study were Cochrane reviews. The publication years ranged from 2003 to 2010. Seven of the nine systematic reviews were comparisons with placebo.

We recalculated the summary statistics for all meta-analyses in the systematic reviews for which unpublished FDA trial outcome data were available. A single systematic review can contain multiple meta-analyses, one for each outcome. The nine included systematic reviews for which unpublished FDA outcome data were available had a total of 41 efficacy outcomes and one safety outcome, creating a final sample of 42 meta-analyses reanalysed.

Thirty-eight per cent (16/42) of the outcomes that we reanalysed were designated as primary outcomes in the selected systematic reviews; 26% (11/42) were designated as primary outcomes in the FDA review. Nine of the 16 primary outcomes in the systematic reviews were not considered to be primary in the FDA reviews. Many outcomes were not designated as primary, secondary, or tertiary in both the systematic reviews (48%; 20/42) and FDA reviews (60%; 25/42).

Effect of unpublished FDA trial outcome data on meta-analyses

Table 2⇓ shows the summary statistics for relevant outcomes both with and without unpublished FDA trial data for each meta-analysis. All published summary statistics were replicated, with the exception of one (eletriptan pain relief at two hours—published effect estimate without unpublished data: relative risk 2.48 (95% confidence interval 1.99 to 3.11); our calculated effect estimate without unpublished data: relative risk 2.42 (1.97 to 2.98)). We used our calculation for the percentage change in the summary statistic calculation.

Overall, addition of the unpublished FDA trial outcome data caused 46% (19/41) of the meta-analyses of each efficacy outcome to estimate decreased efficacy of the drug, 7% (3/41) to estimate the same drug efficacy, and 46% (19/41) to estimate increased drug efficacy. The one meta-analysis with a harm outcome estimated increased harm from the drug when unpublished data were added. The changes in estimates of effect varied widely by outcome, even for the same drug. For example, aripiprazole showed a 53% decrease in improvement of brief psychiatric rating scale score and a 166% increase in improvement of positive and negative syndrome scale score. For those outcomes showing an increase in drug efficacy when unpublished FDA trial outcome data were added, the median magnitude of change in summary statistic was 13% (range 2-166%). For those outcomes showing a decrease in drug efficacy with the unpublished data included, the median magnitude of change in summary statistic was 11% (range 1-53%).

For each drug with multiple unpublished outcomes, the direction of the effect of including the unpublished FDA trial outcome data varied; some meta-analyses changed to show more efficacy and some changed to show less efficacy of the drug. Only for

one drug (galantamine) did all outcomes show decreased efficacy after inclusion of unpublished data.

In the meta-analyses calculated without unpublished FDA data, 34 of 41 efficacy outcomes were statistically significant in favour of the drug and seven showed the drug was not significantly different from its comparator. When the unpublished FDA data were added to the meta-analyses, four of the seven outcomes that were not statistically significant became statistically significant in favour of the drug (frovatriptan, headache recurrence after response at four hours; olmesartan medoxomil 10 mg, change in trough systolic blood pressure; pimecrolimus, clear or almost clear eczema at two and four weeks). The other 34 outcomes remained statistically significant in favour of the drug.

Unpublished FDA trial outcome data comprised more than half of the data contained in the meta-analyses for 14/41 (34%) meta-analyses of efficacy outcomes. These meta-analyses had a median magnitude of change in estimated effect of 19% (range 2-109%). Unpublished FDA trial outcome data comprised less than half of the data contained in the meta-analyses for 27/41 (66%) meta-analyses. These meta-analyses had a median magnitude of change in estimated effect of 7% (range 0-166%).

Overall, heterogeneity, as measured by the I^2 statistic, decreased when unpublished data were added for seven of 42 meta-analyses. The I^2 statistic was unchanged for 23 of 42 meta-analyses and increased for 12 of 42 meta-analyses. Of the 12 meta-analyses for which I^2 increased, no increases led to changes in analysis model; five remained "fixed effects" and seven remained "random effects."

For 24% (10/42) of the recalculated meta-analyses, the unpublished FDA trial outcome data had already been added by the authors of the systematic review. To assess the effect of including the unpublished data under these circumstances, we removed the unpublished data and then added it back into the analysis. We recorded the summary statistics both with and without unpublished data. Of the meta-analyses that already had unpublished FDA data added in the publication, seven out of nine efficacy outcomes showed less benefit of the drug with the unpublished data added and one harm outcome showed more harm with the unpublished data added.

Sensitivity analyses

When the FDA trial data contained multiple analyses for a particular outcome and the methods of the meta-analysis did not use the most conservative method, we did a sensitivity analysis as shown in table 3⇓. None of these sensitivity analyses showed that the summary statistic was sensitive to the decisions about more or less conservative data extraction.

Discussion

We have documented that the addition of unpublished trial outcome data obtained from the Food and Drug Administration to published meta-analyses changes their results. We recalculated 42 meta-analyses (41 efficacy outcomes, one harm outcome) for nine drugs across six drug classes. The even distribution of increases and decreases in estimates of efficacy caused by the addition of unpublished FDA trial outcome data argues against our hypothesis that inclusion of unpublished data obtained from the FDA would decrease the efficacy of the drugs compared with meta-analyses that did not include the unpublished data. Overall, inclusion of the unpublished FDA trial outcome data changed 46% of the meta-analyses to show a decrease in efficacy of the drug and 46% to show an increase

in drug efficacy. The one meta-analysis with an unpublished harm outcome changed to show an increase in harm from the drug. The direction of the effect of including unpublished FDA trial outcome data varied by drug and outcome.

Meaning of study

Although inclusion of the unpublished FDA trial data changed the magnitude of the effect sizes, it changed the statistical significance of few of the individual meta-analyses. Changes in effect sizes may be more meaningful to clinicians and patients than changes in statistical significance. Furthermore, when the unpublished data comprise only a part of the meta-analysis, effect sizes can change substantially without affecting statistical significance. For example, for the migraine treatment eletriptan, the relative risk of being pain-free at one hour decreased by 25%, from a relative risk of 7.94 (2.88 to 21.87) to 4.70 (2.01 to 10.98), even though the confidence intervals overlap and uncertainty remains similar.

Meta-analyses in which more than half of the included data were unpublished showed larger changes in estimates of effect than did those containing smaller proportions of unpublished data. Therefore, when the occurrence of selective outcome reporting is high, meta-analyses of published drug trial data will overestimate or underestimate the efficacy of treatment more than will those in which the occurrence of selective outcome reporting is low. For meta-analyses that have a large proportion of unpublished FDA trial outcome data, an advantage of including the data, and thereby minimising bias, is that the larger number of events will increase the precision of the effect estimates.

Comparison with other studies

The psychiatric drugs (aripiprazole, ziprasidone, galantamine) showed the most consistent changes; four of the five outcomes changed to show decreased efficacy when unpublished FDA trial data were included. Our findings for the psychiatric drugs are similar to those of previous studies of antidepressant trials showing that inclusion of unpublished outcome data in meta-analyses decreased efficacy and increased harms.[12] [13] Our study focused on unpublished data from efficacy trials submitted to the FDA. Turner has shown that for 12 antidepressants, the effect sizes of meta-analyses of the published efficacy data were 11-69% larger than the effect sizes of meta-analyses including the unpublished FDA data.[13] In our study, inclusion of the unpublished FDA trial outcome data decreased the efficacy measures by 22-53%.

Strengths and weaknesses of study

We were able to identify systematic reviews for only nine drugs for which unpublished FDA trial outcome data were available. Our previous study identified 24 drugs with unpublished FDA trial outcome data.[8] One reason for the lack of relevant systematic reviews for 15 of the drugs may be that reviewers are unaware of unpublished outcomes and do not include these outcomes in their protocols. Thus, selective reporting of FDA trial outcomes could affect systematic reviews by influencing the research questions that are asked, as well as the data included in the analyses.

A limitation of our study is that we identified selectively reported outcomes from efficacy trials submitted to the FDA.[8] Although the safety outcome of "any adverse event" was unreported from an efficacy trial of one drug (pimecrolimus), the unreported data for the other drugs consisted of efficacy outcomes only. We did not do a review of all safety data that

was submitted to the FDA to identify unpublished data on harms. Although our findings suggest that inclusion of unpublished FDA trial outcome data changes the results of meta-analyses of efficacy outcomes, we cannot determine the overall effect of unpublished data on the safety of drugs or on the risk-benefit ratio of each included drug.

Conclusions and policy implications

Controversy about including unpublished data from studies of drugs may stem from the belief that unpublished studies are not as methodologically rigorous as their published counterparts or from the stigma associated with the lack of peer review.[4] One survey showed that 78% of meta-analysts and methodologists thought that unpublished material should be included in meta-analyses, whereas only 47% of journal editors believed the same.[16] We had previously assessed risk of bias for all the unpublished FDA trials we identified and found them to have a low risk of bias.[8] In addition, all the unpublished data we included in the recalculated meta-analyses met the inclusion criteria of the original meta-analysis. Therefore, poor quality of unpublished FDA trial data is not a sufficient reason for excluding them from meta-analyses.

Unpublished FDA trial outcome data should be available and included in meta-analyses. Making these data easily accessible is particularly important because the effects of including unpublished data vary by drug. Variability across drugs may be due to the extent of selective outcome reporting or the proportion of unpublished data versus published data in the meta-analysis. As our findings suggest that the direction and magnitude of the effect of including unpublished drug trial data in meta-analyses cannot be predicted, the effect of including unpublished data must be measured for each drug and each outcome as the important differences may be found for some outcomes but not others.

Systematically identifying unpublished data from drug trials is difficult. In addition to data identified from FDA reviews, some data are available in data repositories,[17] [18] and the European Medicines Agency has recently granted access to clinical study reports.[19] In addition to these sources, mechanisms are being developed to promote the full reporting of trial data. For example, some journals, such as *Science*, require that raw data should be available to readers,[20] and public funders are moving towards requiring that data from sponsored research be made publicly available.[21] However, how such data are reported is important.[22] Journals could also consider demanding a list of all trial outcomes, the full protocol, and raw data as a requirement for publication. Groups that do systematic reviews and meta-analyses, such as the Cochrane Collaboration, should consider developing policies about whether regulatory databases (such as the FDA database) should be systematically searched for data to include in the meta-analyses.

Our findings show that data from FDA's medical and statistical reviews are a valuable source of unpublished data for meta-analyses of drugs. The advantage of the FDA reviews is that they are readily available on the internet, but the data obtained from them have several problems. We excluded some meta-analyses from recalculation because we could not find usable data for the unpublished outcomes in the FDA reviews. For example, means were often not reported with their standard deviations, changes from baseline were reported without baseline data or final endpoint data, or continuous data from the FDA review were dichotomised for the meta-analysis in a way that we were unable to replicate. The FDA reviews contain redacted data, and the text can be difficult to read. The reviews do not

adhere to a standard format, so the ease of finding data varies by drug. Whether the FDA reviews contain all data from the clinical study reports submitted to the FDA or a biased selection of the data is also unclear.

To ensure the accuracy and completeness of meta-analyses of the efficacy and safety of drugs, we urge the FDA and other drug regulatory authorities to make the full clinical study reports available to researchers, as the European Medicines Agency has recently agreed to do.[19] Even data from drug trials submitted to regulatory authorities are based on the sponsor's analysis and may, therefore, be biased. Therefore, access to study protocols and the raw data is necessary to minimise bias in the results of meta-analyses.[9]

1   Oxman AD, Schünemann HJ, Fretheim A. Improving the use of research evidence in guideline development: 8. Synthesis and presentation of evidence. *Health Res Policy Syst* 2006;4(20).

2   Schünemann HJ, Fretheim A, Oxman AD. Improving the use of research evidence in guideline development: 1. Guidelines for guidelines. *Health Policy Syst* 2006;4(13).

3   McGauran N, Wieseler B, Kreis J, Schüler Y, Kölsch H, Kaiser T. Reporting bias in medical research—a narrative review. *Trials* 2010;11(37).

4   Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database Syst Rev* 2009;1:MR000006.

5   Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan A-W, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS One* 2008;3:e3081.

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

7   Chan AW, Krleza-Jeric K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *CMAJ* 2004;171:735-40.

8   Rising K, Bacchetti P, Bero L. Reporting bias in drug trials submitted to the Food and Drug Administration: review of publication and presentation. *PLoS Med* 2008;5:e217, discussion e217.

9   Vedula S, Bero L, Scherer R, Dickersin K. Outcome reporting in industry-sponsored trials of gabapentin for off-label use. *N Engl J Med* 2009;361:1963-71.

10 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;326:1171-6.

11 Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010;340:c365.

12 Eyding D, Lelgemann M, Grouven U, Harter M, Kromp M, Kaiser T, et al. 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* 2010;341:c4737.

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

14 Montori VM, Wilczynski NL, Morgan D, Haynes RB. Optimal search strategies for retrieving systematic reviews from Medline: analytical survey. *BMJ* 2005;330:68.

15 Wilczynski NL, Haynes RB. Embase search strategies achieved high sensitivity and specificity for retrieving methodologically sound systematic reviews. *J Clin Epidemiol* 2007;60:29-33.

16 Cook D, Guyatt GH, Ryan G. Should unpublished data be included in meta-analyses? Current convictions and controversies. *JAMA* 2004;291:2457-65.

17 JISC. Dryad UK repository and infectious disease data. 2011. www.jisc.ac.uk/whatwedo/programmes/mrd/clip/dryaduk.aspx.

18 Pharmaceutical Research and Manufacturers of America. ClinicalStudyResults.org. 2011. www.clinicalstudyresults.org.

19 Gøtzsche P, Jorgensen A. Opening up data at the European Medicines Agency. *BMJ* 2011;342:d2686.

20 Science. General information for authors. 2011. www.sciencemag.org/site/feature/contribinfo/prep/gen_info.xhtml.

21 Boulton G, Rawlins M, Vallance P, Walport M. Science as a public enterprise: the case for open data. *Lancet* 2011;377:1633-5.

22 Hrynaszkiewicz I, Norton M, Vickers A, Altman D. Preparing raw clinical data for publication: guidance for journal editors, authors, and peer reviewers. *BMJ* 2010;340:c181.

Cite this as: *BMJ* 2012;344:d7202

What is already known on this topic

Reporting bias exists across a variety of drug classes

When unfavourable results of drug trials are not published, meta-analyses and systematic reviews that are based on only published data may overestimate the efficacy of drugs

What this study adds

Addition of unpublished trial outcome data to published meta-analyses changed their results

The direction of effect of including unpublished trial outcome data varied by drug and outcome

Unpublished trial outcome data should be available and included in meta-analyses; this is particularly important as the effects of including unpublished data are not predictable

Tables

Table 1| Characteristics of systematic reviews

Drug class and name (brand name)	Comparator	Journal type	Publication year
Migraine			
Almotriptan (Axert)	Placebo	Medical	2007
Eletriptan (Relpax)	Placebo	Medical	2007
Frovatriptan (Frova)	Placebo	Medical	2005
Antipsychotic			
Aripiprazole (Abilify)	Haloperidol	Cochrane	2010
Ziprasidone (Geodon)	Placebo	Medical	2003
Dementia/Alzheimer's disease			
Galantamine (Reminyl)	Placebo	Cochrane	2009
Antihypertensive			
Olmesartan medoxomil (Benicar)	Placebo	Cochrane	2009
Antibiotic			
Cefditoren pivoxil (Spectracef)	Penicillin VK	Medical	2004
Topical anti-inflammatory			
Pimecrolimus (Elidel)	Placebo	Cochrane	2007

Table 2| Effect of unpublished Food and Drug Administration (FDA) trial outcomes on meta-analyses

Drug class and name (brand name) and outcome	Proportion of unpublished FDA data in meta-analysis (%)*	Summary statistic		Change in summary statistic (%)†	Direction of change in efficacy
		Without unpublished FDA data	With unpublished FDA data		
Migraine					
Almotriptan (Axert):					
Headache relief at 2 hours	11	RR 1.63 (1.07 to 2.48)	RR 1.69 (1.26 to 2.27)	7	Increase
Pain-free response at 2 hours	11	RR 2.93 (1.68 to 5.12)	RR 2.80 (1.97 to 3.99)	4	Decrease
Eletriptan (Relpax):					
Pain relief at 30 minutes	51	RR 1.17 (0.29 to 4.80)	RR 1.24 (0.61 to 2.53)	37	Increase
Pain relief at 1 hour	22	RR 2.54 (1.95 to 3.31)	RR 2.23 (1.65 to 3.00)	14	Decrease
Pain relief at 2 hours	18	RR 2.42 (1.97 to 2.98)	RR 2.27 (1.89 to 2.73)	7	Decrease
Pain-free at 1 hour	26	RR 7.94 (2.88 to 21.87)	RR 4.70 (2.01 to 10.98)	25	Decrease
Pain-free at 2 hours	17	RR 4.83 (3.05 to 7.66)	RR 4.49 (3.00 to 6.71)	5	Decrease
Recurrence at 24 hours	22	RR 0.72 (0.59 to 0.87)	RR 0.68 (0.58 to 0.78)	17	Increase
Frovatriptan (Frova):					
Pain-free at 2 hours	8	RR 3.63 (2.45 to 5.38)	RR 3.80 (2.59 to 5.59)	4	Increase
Pain-free at 4 hours	8	RR 2.69 (2.19 to 3.30)	RR 2.69 (2.21 to 3.28)	0	No change
Headache response at 2 hours	8	RR 1.66 (1.45 to 1.90)	RR 1.68 (1.47 to 1.90)	2	Increase
Headache response at 4 hours	8	RR 1.81 (1.64 to 2.00)	RR 1.80 (1.64 to 1.99)	1	Decrease
Headache recurrence‡	6	RR 0.79 (0.62 to 1.02)	RR 0.74 (0.58 to 0.94)	28	Increase
Nausea at 2 hours	8	RR 0.88 (0.80 to 0.96)	RR 0.86 (0.79 to 0.94)	18	Increase
Photophobia at 2 hours	8	RR 0.83 (0.78 to 0.89)	RR 0.82 (0.77 t o 0.88)	7	Increase
Phonophobia at 2 hours	8	RR 0.87 (0.80 to 0.94)	RR 0.86 (0.80 to 0.93)	8	Increase
Nausea at 4 hours	8	RR 0.64 (0.57 to 0.71)	RR 0.64 (0.57 to 0.71)	0	No change
Photophobia at 4 hours	8	RR 0.65 (0.59 to 0.70)	RR 0.66 (0.60 to 0.71)	4	Decrease
Phonophobia at 4 hours	8	RR 0.68 (0.62 to 0.76)	RR 0.69 (0.62 to 0.76)	4	Decrease
Antipsychotic					
Aripiprazole (Abilify):					
Improvement in BPRS total score	14	WMD 1.07 (−2.09 to 4.22)	WMD 0.50 (−1.05 to 2.04)	53	Decrease
Improvement in PANSS total score	18	WMD 0.70 (−4.13 to 5.53)	WMD 1.86 (−2.21 to 5.93)	166	Increase
Ziprasidone (Geodon):					
Leaving study early—lack of efficacy	47	RR 0.59 (0.40 to 0.87)	RR 0.67 (0.52 to 0.87)	24	Decrease
Dementia/Alzheimer's disease					
Galantamine (Reminyl)					
Global rating (no change or improvement 32-36 mg/day)—OC	37	OR 2.04 (1.50 to 2.79)	OR 1.72 (1.35 to 2.20)	24	Decrease
Global rating (no change or improvement 32-36 mg/day)—ITT	37	OR 1.87 (1.42 to 2.45)	OR 1.63 (1.31 to 2.02)	22	Decrease
Antihypertensive					
Olmesartan medoxomil (Benicar):					
Change in trough SBP—10 mg	75	WMD −7.00 (−14.70 to 0.70)	WMD −9.59 (−12.95 to −6.23)	37	Increase
Change in trough SBP—20 mg	35	WMD −9.91 (−13.15 to −6.68)	WMD −9.91 (−12.41 to −7.41)	0	No change
Change in trough SBP—40 mg	73	WMD −13.00 (−20.55 to −5.45)	WMD −11.98 (−15.50 to −8.47)	8	Decrease
Change in trough DBP—5 mg	94	WMD −6.20 (−11.24 to −1.16)	WMD −4.70 (−5.77 to −3.62)	24	Decrease
Change in trough DBP—10 mg	92	WMD −5.40 (−10.07 to −0.73)	WMD −6.12 (−7.41 to −4.83)	13	Increase
Change in trough DBP—20 mg	54	WMD −7.11 (−8.94 to −5.28)	WMD −6.99 (−8.16 to −5.82)	2	Decrease
Change in trough DBP—40 mg	74	WMD −6.80 (−11.39 to −2.21)	WMD −7.35 (−9.46 to −5.24)	8	Increase
Antibiotic					
Cefditoren pivoxil (Spectracef):					

Table 2 (continued)

Drug class and name (brand name) and outcome	Proportion of unpublished FDA data in meta-analysis (%)*	Summary statistic		Change in summary statistic (%)†	Direction of change in efficacy
		Without unpublished FDA data	With unpublished FDA data		
Clinical cure rate	27	OR 2.29 (1.61 to 3.28)	OR 2.09 (1.55 to 2.82)	11	Decrease
Bacterial cure rate	26	OR 1.83 (1.37 to 2.44)	OR 1.87 (1.47 to 2.38)	4	Increase
Topical anti-inflammatory					
Pimecrolimus (Elidel) efficacy outcomes:					
Clear or almost clear eczema (IGA 0 or 1) at 1 week	55	RR 2.00 (1.06 to 3.76)	RR 2.78 (1.26 to 6.11)	48	Increase
Clear or almost clear eczema (IGA 0 or 1) at 2 weeks	55	RR 1.58 (1.00 to 2.52)	RR 2.20 (1.22 to 3.98)	72	Increase
Clear or almost clear eczema (IGA 0 or 1) at 3 weeks	52	RR 2.52 (1.65 to 3.84)	RR 2.72 (1.84 to 4.03)	8	Increase
Clear or almost clear eczema (IGA 0 or 1) at 4 weeks	55	RR 1.42 (1.00 to 2.03)	RR 2.08 (1.04 to 4.17)	109	Increase
Clear or almost clear eczema (IGA 0 - 1) at 6 weeks	68	RR 2.29 (1.43 to 3.66)	RR 2.03 (1.50 to 2.74)	15	Decrease
Mild or absent pruritus (pruritus score 0 or 1) at 1 week	15	RR 1.89 (1.51 to 2.35)	RR 1.92 (1.57 to 2.35)	3	Increase
Mild or absent pruritus (pruritus score 0 to 1) at 3 weeks	52	RR 2.10 (1.66 to 2.65)	RR 2.02 (1.69 to 2.42)	5	Decrease
Mild or absent pruritus (pruritus score 0 to 1) at 6 weeks	68	RR 2.17 (1.51 to 3.13)	RR 1.82 (1.48 to 2.25)	23	Decrease
Pimecrolimus (Elidel) safety outcome:					
Any adverse events	49	RR 0.85 (0.71 to 1.03)	RR 0.92 (0.82 to 1.02)	49	More harm

BPRS=brief psychiatric rating scale; DBP=diastolic blood pressure; IGA=investigator global assessment; ITT=intention to treat; OC=observed cases; PANSS=positive and negative syndrome scale; SBP=systolic blood pressure; RR=relative risk; OR=odds ratio; WMD=weighted mean difference.

*Calculated as number of participants from unpublished outcome(s) divided by total number of participants.

†RR, OR: $(\log(E)-\log(I))\times100/\log(E)$, where E=effect estimate excluding unpublished data and I=effect estimate including unpublished data; WMD: $(E-I)\times100/E$; absolute values are reported.

‡After response at 4 hours.

Table 3| Sensitivity analyses

Drug class and name (brand name) and outcome	Summary statistic with unpublished data	Sensitivity analysis
Antipsychotic		
Ziprasidone (Geodon)*:		
Leaving study early—lack of efficacy	RR 0.67 (0.52 to 0.87)	RR 0.64 (0.52 to 0.80)
Dementia/Alzheimer's disease		
Galantamine (Reminyl)†:		
Global rating (no change or improvement 32-36 mg/day)—OC	OR 1.72 (1.35 to 2.20)	OR 1.73 (1.36 to 2.21)
Global rating (no change or improvement 32-36 mg/day)—ITT	OR 1.63 (1.31 to 2.02)	OR 1.63 (1.31 to 2.02)
Antihypertensive		
Olmesartan medoxomil (Benicar) (dose)‡:		
Change in trough SBP mm Hg (10 mg)	WMD −9.59 (−12.95 to −6.23)	WMD −8.60 (−12.06 to −5.13)
Change in trough SBP mm Hg (20 mg)	WMD −9.91 (−12.41 to −7.41)	WMD −9.25 (−11.73 to −6.77)
Change in trough SBP mm Hg (40 mg)	WMD −11.98 (−15.50 to −8.47)	WMD −12.21 (−15.65 to −8.76)
Change in trough DBP mm Hg (5 mg)	WMD −4.70 (−5.77 to −3.62)	WMD −4.46 (−5.56 to −3.37)
Change in trough DBP mm Hg (10 mg)	WMD −6.12 (−7.41 to −4.83)	WMD −6.05 (−7.35 to −4.74)
Change in trough DBP mm Hg (20 mg)	WMD −6.99 (−8.16 to −5.82)	WMD −6.78 (−7.97 to −5.59)
Change in trough DBP mm Hg (40 mg)	WMD −7.35 (−9.46 to −5.24)	WMD −6.80 (−8.98 to −4.62)
Antibiotic		
Cefditoren Pivoxil (Spectracef)§:		
Clinical cure rate	OR 2.09 (1.55 to 2.82)	OR 1.62 (1.27 to 2.08)
Bacterial cure rate	OR 1.87 (1.47 to 2.38)	OR 1.64 (1.33 to 2.04)

DBP=diastolic blood pressure; ITT=intention to treat; OC=observed cases; OR=odds ratio; RR=relative risk; SBP=systolic blood pressure; WMD=weighted mean difference.

*Authors of meta-analysis received unpublished data directly from trial's sponsor and included events for one dosage in their meta-analysis of outcome "leaving study early—lack of efficacy;" events were summed across multiple dosages for other trial outcomes included in meta-analysis; therefore, primary analysis here replicated authors' findings by including unpublished data for one dosage; sensitivity analysis summed events across multiple dosages.

†Authors of meta-analysis received unpublished data directly from trial's sponsor for both outcomes assessed in this study; Food and Drug Administration (FDA) data differed from what authors of meta-analysis included in their meta-analysis; therefore, primary analysis here replicated authors' findings by adding sponsor's unpublished data to data originally used in meta-analysis; sensitivity analysis instead used data extracted directly from FDA reviews.

‡Authors of meta-analysis used per protocol analysis at post-treatment time point, which was used as primary analysis here; sensitivity analysis was done with unpublished data from intention to treat analysis at follow-up time point.

§Data from FDA medical review included in primary analysis and data from FDA statistical review in sensitivity analysis; statistical review was considered to be more conservative.

Figure
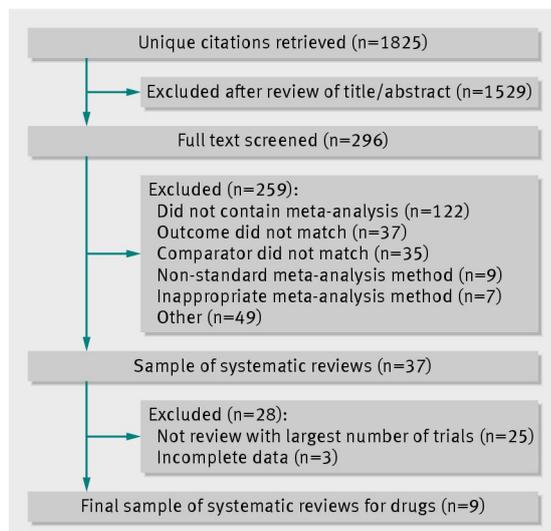
Flow chart for selection of systematic reviews containing at least one meta-analysis