ROBUSTNESS OF RESULTS AND CONCLUSIONS IN SYSTEMATIC REVIEWS, TRIALS AND ABSTRACTS

Anders W. Jørgensen
PhD thesis

31 May 2011

The Nordic Cochrane Centre
Rigshospitalet

Faculty of Health Sciences
University of Copenhagen
# TABLE OF CONTENTS

Preface and acknowledgements .................................................................................................................. 3

Structure of the thesis ................................................................................................................................. 3
Acknowledgement .......................................................................................................................................... 3
Financial support ........................................................................................................................................ 3

Original Papers ........................................................................................................................................... 4

Abstract....................................................................................................................................................... 5
Danish Summary ......................................................................................................................................... 6

Introduction ................................................................................................................................................ 7

Systematic reviews .................................................................................................................................... 7
Bias ............................................................................................................................................................... 8
The abstracts role in dissemination of information .................................................................................... 9
Large losses to follow up, ITT and imputation of missing data ................................................................... 9
Multiple imputation .................................................................................................................................... 10
Objectives .................................................................................................................................................. 11

Methods and results ................................................................................................................................... 12

Paper 1 ....................................................................................................................................................... 12
  Discussion of paper 1 ................................................................................................................................. 18
Paper 2 ....................................................................................................................................................... 21
  Discussion of paper 2 ................................................................................................................................. 29
Paper 3 ....................................................................................................................................................... 31
  Discussion of paper 3 ................................................................................................................................. 39
Paper 4 ....................................................................................................................................................... 41
  Discussion of paper 4 ................................................................................................................................. 60
Paper 5 ....................................................................................................................................................... 62
  Discussion of paper 5 ................................................................................................................................. 67

Conclusion ................................................................................................................................................ 69

Implications ............................................................................................................................................... 70
Reccommendations for future research ..................................................................................................... 71

References ................................................................................................................................................. 72
PREFACE AND ACKNOWLEDGEMENTS

STRUCTURE OF THE THESIS
The structure of this thesis follows the guidelines from the Faculty of Health Sciences, University of Copenhagen. First I will give an introduction to the context of this thesis and present the objectives. The next part of the thesis consists of five articles, each followed by an additional discussion of the methods and findings in the context of observations done by other researchers on similar topics. Finally, I present an overall conclusion.

ACKNOWLEDGEMENTS
It is a great opportunity and a pleasure to thank those who made this thesis possible. I would like to thank my supervisor Peter C. Gøtzsche, who in the first place triggered me as a medical student into this field with his inspiring lectures about evidence based medicine, and secondly helped me with financial support and facilities, but most importantly, for his enthusiasm, drive and determination that also made this thesis possible.

I would also like to thank my co-authors for their constructive discussions and collaboration on the papers.

My gratitude goes to all my colleagues at The Nordic Cochrane Centre for their support and for making my three years as a PhD student a pleasant time.

Finally, I am forever indebted to my wife Anne who gave birth to our two wonderful girls and made moving to a new house possible during my PhD studies, but above all for her optimism, love and care.

FINANCIAL SUPPORT
I would like to thank The Nordic Cochrane Centre, The Danish Council for Independent Research in Medical Sciences (FSS), The Health Insurance Foundation, The Medical Research Foundation of the Capital Region of Denmark and Copenhagen University Hospital, which funded this PhD.
The thesis is based on the following papers:

Jørgensen AW, Hilden J, Gøtzsche PC. Cochrane reviews compared with industry supported meta-analyses and other meta-analyses of the same drugs: systematic review. BMJ 2006;333:782.


Jørgensen AW, Lundstrøm LH, Wetterslev J, Astrup A, Gøtzsche PC. We should not continue to use complete case analysis and single imputations like LOCF and BOCF. A comparison of results from different imputation techniques of missing data from an anti-obesity drug trial. (Manuscript to be submitted)

Bias in drug trials, e.g. introduced through lack of allocation concealment, lack of blinding, a large amount of missing outcome data, no intention-to-treat analysis and selective outcome reporting, is common and conclusions often favour the sponsor’s product in drug trials. Therefore, a critical appraisal of the research methodology and results is essential, but this is not an easy task when important details and data are kept unpublished.

The aim of this PhD was to explore the robustness of results and conclusions in reviews, randomised trials and abstracts. This was done in 4 studies. In the first study, we compared Cochrane reviews with industry-supported meta-analyses and other meta-analyses of the same drugs. In the second, we compared all industry-supported meta-analyses of drug-drug comparisons with those without industry support. In the third, we studied the balance between benefits and harms reported in abstracts on rofecoxib (Vioxx), and in the fourth, we compared results from different imputation methods for missing outcome data in an anti-obesity drug trial where we had access to individual patient data. A fifth paper, in which we describe how we succeeded to get access to unpublished trial reports possessed by the European Medicines Agency, has also been included in this thesis.

The results from the first two studies demonstrated that industry-supported reviews of drugs should be read with caution, as they were less transparent and had more favourable conclusions than other reviews. The third study found that the reporting of benefits and harms in abstracts on Vioxx was unbalanced. The fourth showed great variation in the results when different imputation techniques were used.

I conclude that systematic reviews, randomised trials and abstracts should be read with caution, especially when the research is industry supported, as important information is often not published. We suggest more transparency and access to unpublished trials and the raw data and trial protocols.
Bias i lægemiddelforsøg, f.eks. opstået ved utilstrækkelig skjult allokering, manglende blinding, manglende data, ingen 'intention-to-treat'-analyse og selektiv rapportering af forsøgsresultater, er almindelig, og konklusionerne er ofte til fordel for sponsorens lægemiddel. Derfor er det vigtigt, at man kritisk vurderer metoder og resultater, hvilket dog ikke er nemt, når vigtige detaljer og data ikke bliver offentliggjort.

Formålet med ph.d.-projektet, som blev opdelt i 4 delprojekter, var at undersøge pålideligheden af resultater og konklusioner i oversigtsartikler, artikler om lægemiddelforsøg og artikelresuméer. I det første projekt sammenlignede vi Cochrane-oversigter med firmastøttede og andre meta-analysen af de samme lægemidler. I det andet sammenligne vi firmastøttede meta-analyser af lægemiddelsammenligninger med ikke-frimastøttede meta-analyser. I det tredje undersøgte vi balancen mellem gavnlige og skadelige virkninger beskrevet i artikelresuméer om rofecoxib (Vioxx), og i det fjerde sammenligne vi forskellige imputationsmetoder for manglende data i et et forsøg med et lægemiddel mod fedme, hvor vi havde adgang til individuelle patientdata. En femte artikel, hvor vi beskriver hvordan det lykkedes os at få aktindsigt i hemmeligholdte forsøgsrapporter hos den Europæiske Lægemiddelstyrelse, er også indkluderet i ph.d.-afhandlingen.

Resultaterne fra de to første projekter viste, at man skal være særlig kritisk overfor firmastøttede meta-analyser, da de er sværere at gennemskue og oftere anbefaler det eksperimentielle lægemiddel end andre meta-analyser. I det tredje projekt fandt vi en ubalanceret beskrivelse af gavnlige og skadelige effekter i artikelresuméer om rofecoxib. Det fjerde viste en betydelig variation i resultaterne, når forskellige imputationsmetoder blev brugt.

Jeg konkluderer, at systematiske oversigtsartikler, artikler om lægemiddelforsøg og artikelresuméer bør læses med omhu, fordi vigtig information ofte ikke offentliggøres, og især når forskningen er støttet af firmaer. Jeg foreslår mere gennemsigtighet og fri adgang til ikke-offentliggjorte lægemiddelforsøg, individuelle patientdata og forsøgsprotokoller.
INTRODUCTION

Health professionals, including physicians, make a lot of decisions every day in their clinical practice about patients’ diagnoses and treatments. They try to choose the best test for diagnosing and the most effective intervention for treating diseases. A clinical decision is based on the setting, the patient’s values and expectations and the physician’s experience, skills, personality, and of course the knowledge, which primarily comes from the medical literature (1). However, to make informed decisions requires available and reliable information from unbiased research that seeks to answer clinical questions relevant to patients, but to get this kind of information is difficult because the amount of healthcare research is overwhelming (2) and most is not applicable to clinical decision-making.

Evidence-based medicine, defined as the integration of best research evidence with clinical expertise and patient values (3), is helpful in making informed decisions. Practicing evidence-based medicine is about trying to improve the quality of the information on which the decisions are based by avoiding information overload and by finding and applying the most useful information. A task that is not unique to medicine, but to all informed decisions (4).

Within the framework of evidence-based medicine, study designs have traditionally been arranged and graded hierarchically with randomised controlled trials (RCTs) ranking higher than observational studies (5,6). Unknown and uncontrollable prognostic factors can bias the result in observational studies if they are unbalanced, e.g. between the controls and cases. Due to randomisation the chance of unbalanced prognostic factors in RCTs is low and therefore the risk of bias is low.

SYSTEMATIC REVIEWS

Despite being retrospective, systematic reviews of RCTs rank higher than RCTs because they summarize, if possible in a meta-analysis, all the evidence from the available RCTs resulting in a higher degree of precision (5-7). The Cochrane Collaboration is an independent organisation that produces systematic reviews (Cochrane reviews) that aim at providing information for healthcare decisions by addressing the choices that politicians, clinicians and patients face and by reporting meaningful outcomes. To minimize bias and to promote transparency of methods, Cochrane reviews are based on pre-defined, peer-reviewed and published protocols (8).
BIAS

Systematic reviews and RCTs rank high in the hierarchy of evidence, but when conducted poorly the risk of bias is high. Major limitations in study methodology increase the risk of bias in these studies and lower their quality. For RCTs, these limitations in study design include lack of allocation concealment, lack of blinding, large losses to follow-up and failure to adhere to an intention-to-treat (ITT) analysis (9). If a meta-analysis is based on trials with a high risk of bias, the result from the meta-analysis will also have a high risk of being biased. For systematic reviews, poor searching and improper selection of studies increase the risk of bias. All this can be assessed if the papers of trials and systematic reviews report important details transparently, but this is not always the case.

A major pitfall for evidence-based medicine is missing or unpublished information. Published trials may only include some of the outcomes investigated and analyses actually undertaken, and if there is a systematic difference between the reported and unreported outcomes, it may lead to bias. Such selective outcome reporting bias has been demonstrated in studies of RCTs (10,11). The same is true for publication bias, where positive results get published more often than other results (12,13). This means that the information that doctors and other health personal rely on represent the most positive information and thus the informed decision becomes an illusion.

Part of the solution to publication and selective reporting bias is to let independent researchers have access to data at the medicines regulatory authorities. When manufacturers apply for marketing approval, they submit detailed reports of all trials, published or unpublished to the medicines agencies. Such invaluable information should be used to provide a more reliable basis for an informed decision.

There are several motives for deliberate introduction of bias in research. Interests, financial and non-financial, may conflict with the responsibility to conduct, design and report trials objectively. Little research has been done to investigate non-financial conflict of interests (14), whereas numerous studies have documented that financial conflicts of interest may influence the study outcome, and the evidence has been synthesized in several systematic reviews (15-18).
THE ABSTRACT’S ROLE IN DISSEMINATION OF INFORMATION

When finding the best available evidence, the search often begins with databases. In addition to the title of the article, the databases usually include an abstract, which is a condensation of the information in the article. This makes the abstract the key entry to the information and the part of a paper that - second to the title - is most often read. Sometimes the abstract is the only published information about a trial, e.g. presented at a conference, and sometimes clinical decisions are based solely on abstracts (19,20). Abstracts are therefore a very important information source when evaluating an intervention, which underlines the importance of well-written abstracts. Since medical interventions and tests have the potential to be both beneficial and harmful, the benefits and harms must be equally prioritised and reported in the medical literature including the abstract (20).

LARGE LOSSES TO FOLLOW UP, AND IMPUTATION OF MISSING DATA

Attrition is inevitable in large long-term RCTs. Some subjects will discontinue the treatment due to harms, lack of benefit or other reasons, for example moving home, which makes participation in the trial troublesome. If there is a systematic difference in attrition between the intervention and the control group, the risk of bias is high if only the subjects who completed the trial are analysed. Intention-to-treat (ITT) is a strategy to reduce this bias and implies that all randomised subjects are included in the analyses, regardless of whether they received the treatment to which they were originally allocated, or were subsequently withdrawn or deviated from the protocol in other respects (21). In other words, an ITT analysis estimates the effect of the intention to give a particular treatment and thus mimics clinical practice where patients also discontinue treatments for the same reasons as stated above. Therefore, the principle of ITT is widely accepted and appears in several guidelines and recommendations (22,23). Frequently, however, the ITT approach is inadequately described and inappropriately applied (8,21). To make a proper ITT analysis, it is important that data are available for all patients, but missing data is a common problem in all types of medical research.

Attrition and missing data do not always introduce bias. If data are missing by chance then the available data can be regarded as a random sample of the ‘full’ data set and there is no need for imputation. But this situation is not common in trials and missing data must therefore be imputed.

In trials of anti-obesity drugs, the most used method for imputing missing data is ‘last observation carried forward’ (LOCF) (24). This method replaces missing measurements of body weight with the last measured weight. For example, if 5 weight measurements have been planned, but a patient only attends the first 3 visits, the missing weight measurements
on visits 4 and 5 are replaced by the measurement at visit 3. To assume that a subject’s weight is unchanged after dropping out of a trial is hard to justify, as most patients regain much of their lost weight within a short period of time after they stop treatment (25). Further, LOCF overestimates the precision of the effect estimate because the dataset is analysed as if it were a ‘complete’ dataset with no missing data (26). Therefore, LOCF is likely to be a biased analysis.

There are other and more reliable techniques for handling missing data (27). Attention has been drawn to multiple imputation (MI) (28) and it was recently recommended for handling missing data in obesity trials (24).

MULTIPLE IMPUTATION

Multiple imputation may be the most reliable technique (29). It is a stepwise procedure. In the first stage, multiple copies of the dataset are being created with the missing values replaced by imputed values. The imputed values are randomly drawn from their predictive distribution based on the observed data. Since we cannot know the true values of the missing data, the imputation procedure must inject appropriate variability into the multiple imputed values to account for the uncertainty when the missing values are predicted (which single imputation such as LOCF doesn’t do). In the second stage, the multiple data sets that have been generated are analysed with standard statistical methods and the results, which differ from dataset to dataset due to the variation introduced in the imputation of the missing values, are averaged. The final estimate is unbiased because we are averaging over the distribution of the missing data given the observed data and incorporating the uncertainty of the imputed missing values into the confidence intervals.
OBJECTIVES

I wanted to explore the robustness of results and conclusions in systematic reviews, randomised trials and abstracts and to get access to unpublished data at the European Medicines Agency. I undertook five projects with the aims:

1. To compare Cochrane reviews with industry-supported meta-analyses and other meta-analyses of the same drugs for differences in conclusions and methodological quality.

2. To assess all meta-analyses indexed on PubMed of drug-drug comparisons published in 2004 and compare those with industry support with those with other or no support.

3. To quantify how often benefits and harms in terms of gastrointestinal bleeding and cardiovascular thrombosis, respectively, were reported in abstracts on rofecoxib.

4. To analyse the weight change and compare the results when using four different methods for handling missing data in an randomised trial of an anti-obesity drug.

5. To describe the character of the arguments in our case when we applied for access to unpublished trial reports and protocols of two anti-obesity drugs at the European Medicines Agency.
The aim of the first study was to compare methodological quality and conclusions in Cochrane reviews with those in industry supported meta-analyses and other meta-analyses of the same drugs.
Cochrane reviews compared with industry supported meta-analyses and other meta-analyses of the same drugs: systematic review
Anders W Jørgensen, Jørgen Hilden, Peter C Gøtzsche

Abstract

Objective To compare the methodological quality and conclusions in Cochrane reviews with those in industry supported meta-analyses and other meta-analyses of the same drugs.

Design Systematic review comparing pairs of meta-analyses that studied the same two drugs in the same disease and were published within two years of each other.

Data sources Cochrane Database of Systematic Reviews (2003, issue 1), PubMed, and Embase.

Data extraction Two observers independently extracted data and used a validated scale to judge the methodological quality of the reviews.

Results 175 of 1596 Cochrane reviews had a meta-analysis that compared two drugs. Twenty four meta-analyses that matched the Cochrane reviews were found: eight were industry supported, nine had undeclared support, and seven had no support or were supported by non-industry sources. On a 0-7 scale, the median quality score was 7 for Cochrane reviews and 3 for other reviews (P < 0.01). Compared with industry supported reviews and reviews with undeclared support, Cochrane reviews had more often considered the potential for bias in the review—for example, by describing the method of concealment of allocation and describing excluded patients or studies. The seven industry supported reviews that had conclusions recommended the experimental drug without reservations, compared with none of the Cochrane reviews (P = 0.02), although the estimated treatment effect was similar on average (z = 0.46, P = 0.64). Reviews with undeclared support and reviews with no support had conclusions that were similar in cautiousness to the Cochrane reviews.

Conclusions Industry supported reviews of drugs should be read with caution as they were less transparent, had fewer reservations about methodological limitations of the included trials, and had more favourable conclusions than the corresponding Cochrane reviews.

Introduction

Bias in drug trials is common and often favours the sponsor’s product. Critical, systematic reviews that aggregate the available information in a neutral manner are therefore essential. Cochrane reviews aim to minimise bias and avoid conflicts of interest, and, on average, they may have greater methodological rigour than systematic reviews published in paper based journals. We therefore hypothesised that Cochrane reviews would be more transparent and less biased than industry supported systematic reviews. We aimed to compare Cochrane reviews with other meta-analyses of the same drugs, which we divided into those that had industry support, those with undeclared support, and those that had non-profit support or no support.

Methods

We searched for pairs that consisted of a Cochrane review and a similar review in a paper based journal. A Cochrane review was eligible if it used meta-analysis to compare at least two different drugs or classes of drugs; was published in the Cochrane Database of Systematic Reviews 2003, issue 1; could be matched with a meta-analysis of the same drugs and diseases published in full in a paper based journal within two years before or after the most recent substantive amendment of the Cochrane review; and had no authors in common with the Cochrane review.

We defined support by the pharmaceutical industry as provision of grants, authorship, or other major assistance such as help with the statistical analysis. We did not consider provision of references or unpublished trial reports as support.

One investigator (AWJ) hand searched all reviews in the Cochrane Database of Systematic Reviews 2003, issue 1 for drug comparisons. For each potentially eligible Cochrane review, we sought possibly eligible paper based reviews by searching PubMed (January 1966 to July 2003) for the same diseases and drugs combined with “meta-analysis” or meta-analysis[pt]. From online inspection of titles and abstracts, we selected meta-analyses for examination of the full text. When we found no match in PubMed, we searched Embase (WebSPIRS 5) (1980 to August 2003). When we found more than one match with the same type of support, we chose the one with the closest publication date to the Cochrane review.

AWJ and PCG independently assessed each pair of reviews in random order, by reading the Cochrane review first in half of the pairs. We used a pilot tested data sheet and resolved disagreements by discussion. We were not blinded. We extracted data on the date of the most recent substantive amendment to the Cochrane review and of the publication of the paper based review; names of relevant drugs and diseases; types of support; number and type of sources used to identify trials for the review; searches for unpublished trials; and descriptions of concealment of allocation, details of blinding, and excluded patients and trials.

We assessed the methodological quality of the reviews with Oxman and Guyatt’s index, which is a validated tool with nine
items considering the potential for bias and an overall assessment on a 0-7 scale. We judged the review authors’ conclusions by assessing whether the experimental intervention was recommended without reservations or whether it was not recommended or recommended only with reservations. As our data were paired, we compared quantitative data with the Wilcoxon-Pratt one sample rank sum test and binary data with a sign test; P values are two sided.

For the industry supported reviews, we also assessed whether the estimated treatment effects were different from those reported in the Cochrane reviews. We used the first reported outcome with data in the industry supported review that was also presented in the Cochrane review, and we also did the corresponding analysis in which we started with the first reported outcome in the Cochrane review. We calculated pooled comparative z scores, after adjustment for the number of patients contributing to the outcome and for the number of patients that were common to the two analyses (see appendix on bmj.com for details).

Results

The Cochrane Database of Systematic Reviews 2003, issue 1 contained 1596 Cochrane reviews, of which we excluded 1421, mostly because they did not compare drugs (figure). In 72 of the remaining 175 reviews, the Medline and Embase searches identified potentially eligible paper based reviews, of which we excluded 48, mostly because their publication date differed by more than two years from the last substantive update of the Cochrane review, resulting in 24 matched pairs. One Cochrane review was paired with a paper based review with industry support and also with one with undeclared support; one of these reviews presented only a subgroup analysis, and we therefore substituted it with an older review that contained the same trials. In eight of the 24 pairs the paper based reviews had industry support, in nine they had undeclared support, and in seven they had non-profit or no support.

The overall median quality score was 7 for the 24 Cochrane reviews and 3 for the other reviews (P < 0.001; table A on bmj.com). The mean years of publication for the three sets of pairs were 2000 versus 2000 for Cochrane reviews versus industry supported reviews, 2000 versus 1999 for Cochrane reviews versus reviews with undeclared support, and 2000 versus 2000 for Cochrane reviews versus reviews with non-profit or no support; the median differences in number of included trials were 0, 1, and 1 (table B on bmj.com).

Cochrane reviews versus industry supported reviews (eight pairs) Cochrane reviews were of higher quality than industry supported reviews (P < 0.01). They also more often stated the search methods used to find studies (P = 0.06), avoided bias in the selection of studies (P = 0.03), reported criteria for assessing the validity of the studies (P = 0.03), used appropriate criteria in assessing the studies (P < 0.01) (table A), described methods of concealment of allocation (P = 0.02), and described excluded patients (P = 0.03) and studies (P = 0.03), and they used more sources to identify studies (P = 0.02) (table B).

One of the industry supported reviews had no conclusion, as it referred to physiological characteristics of the drug. The other seven reviews supported by industry all recommended the experimental drug without reservations, compared with none of the Cochrane reviews (P < 0.01). This difference was related to interpretation of the data (table) and consideration of costs. The authors of six of the eight Cochrane reviews had reservations about the quality or relevance of the trials or their findings, and two noted that the effect decreased with increasing sample size. Seven mentioned the higher cost of the experimental drug as a problem compared with none of the industry supported reviews, of which two claimed that economic analyses had shown that the experimental drug was cost effective. In contrast to these interpretations, the estimated treatment effect was similar, on average, in the pairs of reviews (pooled z = 0.46, P = 0.64; appendix and table C on bmj.com). However, the scatter of the comparative z scores was high (P = 0.02) (table B on bmj.com). The mean years of publication for the three sets of pairs were 1999 versus 1990 for Cochrane reviews versus industry supported reviews (P = 1.00) or efforts to avoid bias in the selection of studies (P = 0.22) and the recommendations were without reservations in one Cochrane review and in two other reviews (P = 1.00).

The paper based reviews were often biased and poorly done. One seemed to have included non-randomised studies. Another included retrospective studies, had arbitrary entry criteria, and seemed to have preferentially selected those studies and data that were in favour of the experimental drug; the outcome was analysed in eight different ways, and the biggest difference was emphasised. A third review presented a highly misleading result in the abstract that was based on an indirect comparison of treatment arms, and, in contrast to the Cochrane review, the authors failed to include all eligible studies and failed to detect sample size bias. A fourth review found no differences between the drugs but stated, with reference to an economic analysis: “Coupled with potential cost savings driven by the reduced need for hospitalization and revascularization procedures, it is not dif-
ficult to understand why clinicians, administrators and medical organizations representing the interests of physicians and the welfare of patients look favourably upon [low molecular weight heparin].

**Cochrane reviews versus reviews with non-profit or no support (seven pairs)**

We found no significant differences between Cochrane reviews and reviews with non-profit or no support (table A). The recommendations were without reservations in two Cochrane reviews and in one other review.

However, in three pairs of reviews the conclusions favoured different drugs. In one set, although the same trials were included, the authors had different views on the trade-off between benefits and harms. The Cochrane review strongly supported the use of warfarin in patients with atrial fibrillation with average or greater risk of stroke, whereas the paper based review strongly favoured antiplatelet drugs. In another set, the paper review found better control of haemorrhage with octreotide.

The Cochrane review did not find any differences but referred to another meta-analysis when it emphasised that terlipressin is the only drug for which a reduction in mortality has been found compared with no treatment. We have published a comment with this review, as we doubt that the effect on mortality is reliable.

In the third set, the paper review included only three double blind randomised trials and concluded that methadone should be used for heroin dependence; it noted in an addendum that LAAM (levomethadyl acetate hydrochloride) had been withdrawn from the market because of cardiotoxicity. The Cochrane review included 18 studies, three of which were cohort studies that were not analysed separately from the randomised trials, and concluded that LAAM seems to be more effective than methadone. A fourth Cochrane review found a major sample size bias but failed to make reservations in the abstract as regards the mortality benefit, which was based on all trials.

<table>
<thead>
<tr>
<th>References</th>
<th>Disease</th>
<th>Interventions</th>
<th>No of included trials (both reviews/C only if only)</th>
<th>No of included patients (C/I)</th>
<th>Significant difference favouring drug of interest</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>w1, w4</td>
<td>Acute coronary syndrome</td>
<td>Enoxaparin v unfractionated heparin</td>
<td>2/0/0</td>
<td>7081/7081</td>
<td>Yes/yes</td>
<td>No/no</td>
</tr>
<tr>
<td>w5, w6</td>
<td>Schizophrenia or similar psychoses</td>
<td>Amisulpride v typical antipsychotics</td>
<td>14/0/0</td>
<td>1702/1700</td>
<td>Yes/yes</td>
<td>Yes/yes</td>
</tr>
<tr>
<td>w7, w8</td>
<td>Rheumatoid arthritis</td>
<td>Celecoxib v other NSAIDs</td>
<td>4/1/1</td>
<td>4465/4191</td>
<td>No/yes</td>
<td>Yes/yes</td>
</tr>
<tr>
<td>w9, w10</td>
<td>Schizophrenia</td>
<td>Risperidone v haloperidol</td>
<td>8/4/0</td>
<td>2326/1047</td>
<td>Yes/yes</td>
<td>No/yes</td>
</tr>
<tr>
<td>w11, w12</td>
<td>Asthma</td>
<td>Inhaled fluticasone v budesonide and beclomethasone</td>
<td>14/16/0</td>
<td>7775/3564</td>
<td>Yes/yes</td>
<td>No/yes</td>
</tr>
<tr>
<td>w13, w14</td>
<td>Asthma</td>
<td>Salmeterol v theophylline</td>
<td>3/2/0</td>
<td>777/1330</td>
<td>No/yes</td>
<td>Yes/yes</td>
</tr>
<tr>
<td>w15, w16</td>
<td>Depression</td>
<td>Paroxetine v TCAs</td>
<td>15/17/24</td>
<td>5910/3758</td>
<td>NA/no</td>
<td>Yes/yes</td>
</tr>
<tr>
<td>w17, w18</td>
<td>Depression</td>
<td>Venlafaxine v other antidepressants</td>
<td>0/4/5</td>
<td>708/450</td>
<td>No/yes</td>
<td>NA/no</td>
</tr>
</tbody>
</table>
Discussion

We found that although some Cochrane reviews had clear methodological deficiencies, these were fewer, on average, than in reviews published in paper based journals.

Limitations

A minor limitation was that we could not be blinded, as the layout of Cochrane reviews is unique; blinding has little impact on extraction of data for reviews. A more important limitation is that our sample was small and needs to be replicated. Furthermore, we are affiliated with a Cochrane centre, and Cochrane reviews are done according to a handbook that was developed partly by Andy Oxman, who also participated in the development of the validated index that we used for evaluating methodological quality. To help readers to make their own judgments, we have provided details on the comparison with industry supported reviews (table) and added items to our data extraction sheet that are indisputably important for the reliability of reviews, such as adequate concealment of allocation.

Our findings correspond to another recent finding that Cochrane reviews assess methodological quality more often than do other reviews, although because of space constraints some paper based reviews might have been conducted better than was reported.

Industry supported review

The estimated treatment effects in industry supported reviews were similar to those of Cochrane reviews, but the former had uniformly positive recommendations for the experimental drug, without reservations about methodological limitations of the trials or costs, in contrast to none of the Cochrane reviews. This suggests that the main problem with industry supported reviews lies in how conclusions are formulated.

We compared pairs of similar reviews published within a narrow time frame. Other such pairwise comparisons have been anecdotal. A meta-analysis found a similar drop-out rate for fluoxetine as for tricyclic antidepressants (P = 0.4), whereas a company employee reported a marked difference in favour of fluoxetine (P < 0.001) in subsequent correspondence. This is surprising, as the industry supported meta-analysis contained fewer patients and included “data on file” reports, which are usually less favourable than published ones. In contrast to industry supported authors, authors of Cochrane reviews often cannot get access to such data, as indicated in two of the reviews.

As another example, a meta-analysis supported by Merck concluded in 2001 that no increased risk of arterial thrombosis existed with the company’s drug rofecoxib, but a meta-analysis not supported by industry showed an increased risk, which was apparent in publications available to the authors of the industry supported meta-analysis. Rofecoxib was withdrawn because of thromboses in 2004.

The influence of industry on trial reports is similar to our findings. A survey found that none of 56 trials of non-steroidal anti-inflammatory drugs supported by the manufacturer presented results that were unfavourable to the company. Another survey found that the conclusions recommended the experimental drug as the drug of choice five times as often if the trial was funded by for-profit organisations, even after adjustment for the effect size.

Reviews with undeclared support

The conclusions of paper based reviews with undeclared support were more cautious than those for industry supported reviews. We contacted the authors after we had assembled our data. Eight declared that they had not received any external funding or other type of support (which we exemplified as help with the statistical analyses), and one replied that he had not received any other financial support. We do not know whether these replies were comprehensive, and we suspect that some authors had received undeclared support or had allowed the company to review the paper and insert text, as suggested by the recommendation for low molecular weight heparin above.

Interpretation of financial support

The interpretation of financial support is not always straightforward. The authors of a paper based review that we classified as “non-profit support” noted that it was supported in part by public sources but did not describe the nature of the other part, and two of the authors had previously received “unrestricted grants” from the manufacturer of octreotide. The authors of the matching Cochrane review declared that they had no conflicts of interest but added that they had “no permanent financial contracts” with companies producing the comparator, terlipressin. Finally, one of the Cochrane reviews had industry support, as Upjohn had funded secondary analyses of the author’s own trial for use in the review.

The current policy in the Cochrane Collaboration is that industry support of Cochrane reviews is not acceptable.

Other problems with reviews

Less rigorously controlled studies than ours have reported on discrepant conclusions between systematic reviews assessing the same subject. The major reasons were incomplete searches, differential inclusion of trials, insufficient attention to the quality of the trials and to bias detection, and differences in interpretation. Some reviews missed more than half of the available trials, and a review of meta-analyses of analgesic interventions found that those with positive conclusions had lower quality scores on the Oxman and Guyatt index. A recent study of antihypertensive drugs found that the conclusions of meta-analyses were positive in 91% of the papers with financial ties and in 72% of other papers. However, the conclusions of Cochrane reviews also tend to be too positive.

Our examination of the comparative z scores revealed more scatter than expected, which indicates that some effect estimates might have been biased, and, furthermore, that the confidence interval in a meta-analysis generally exaggerates the precision in the underlying data.

Conclusions

Industry supported reviews of drugs are less transparent than Cochrane reviews and have few reservations about methodological limitations of the included trials; their conclusions should be read with caution. We believe that details of concealment of allocation, blinding, inclusion and exclusion criteria for trials, search strategies, and estimated effects in each included trial need to be reported to allow readers to judge the reliability of reviews. To improve transparency, access to the protocol should be available. Protocols for Cochrane reviews are published in the Cochrane Library, and protocols for other systematic reviews can be registered free of charge at the UK national research register through the Centre for Reviews and Dissemination in York, UK.

We thank Stefan Leucht, Deborah J Cook, Deborah Goebert, Michael Ludwig, Max H Pittler, Chiel Springer, Steven L West, Frederick A Spencer, and Patrick Chien for providing information on support and funding of their studies and the Iberoamerican Cochrane Centre, Barcelona for providing study facilities.

Contributors: AWJ wrote the draft protocol and manuscript, and PCG contributed. AWJ and PCG extracted data. JH did the statistical analyses that
compared estimated treatment effects. All authors commented on the final manuscript. PCG is the guarantor.

Funding: None.

Competing interests: AWJ and PCG are affiliated with the Nordic Cochrane Centre. The views expressed in this article represent those of the authors and are not necessarily the views or the official policy of the Cochrane Collaboration. The views expressed in this article represent those of the authors and are not necessarily the views or the official policy of the Cochrane Collaboration. The views expressed in this article represent those of the authors and are not necessarily the views or the official policy of the Cochrane Collaboration.


What is already known on this topic

- Bias commonly occurs in trials of healthcare interventions and often favours the sponsor’s product
- Anecdotal reports have suggested that industry supported meta-analyses may also be more flawed than other meta-analyses

What this study adds

- Industry supported reviews were of lesser quality than Cochrane reviews of the same drugs and always recommended the experimental drug without reservations, which none of the Cochrane reviews did
- Industry supported meta-analyses of drugs were less transparent and had few reservations about methodological limitations of included trials
- Reviews with undeclared support and those with not for profit support or no support had similarly cautious conclusions to matched Cochrane reviews
- Industry supported meta-analyses of drugs were less systematic and had more restrictions on publication of adverse events and rofecoxib: cumulative meta-analysis. Lancet 2001;364:2011-2016.

Correspondence to: P C Gøtzsche pcg@cochrane.dk
DISCUSSION OF PAPER 1

In this study, we compared pairs of meta-analyses that consisted of Cochrane reviews closely matched to meta-analyses of the same drugs in reviews with different kinds of support published in paper based journals. Overall, we found that Cochrane reviews were of better methodological quality than the other reviews although some of the Cochrane reviews had clear methodological deficiencies.

We used an instrument developed by Andy Oxman and Gordon Guyatt to assess the overall methodological quality of the meta-analyses. Both researchers have also been involved with authoring the Cochrane Handbook, which is the official guide on how to conduct a Cochrane review. Therefore, their instrument has been criticised for favouring Cochrane reviews over other reviews explaining the differences in methodological quality between the set of reviews. However, to our knowledge it was the only validated tool among the more than 24 other tools available when we did the study and it was also the most used (30). Since then, the 11-item assessment tool AMSTAR has been developed and validated (31,32). AMSTAR addresses similar domains as the Oxman-Guyatt Index and therefore I do not believe that this tool would have altered our results.

We assessed the methodological quality of meta-analyses by reading the published paper only and one could therefore argue that we assessed the quality of reporting rather than the actual methods. This seems to be correct. An observational study found that authors of randomised controlled trials frequently used concealment of allocation and blinding, despite the failure to report these methods (33) and a comparison of the protocol and the published paper of 73 trials showed that the protocols described the blinding methods more precisely than the papers (34). However, the details reported may still be a good surrogate for the actual methodology. In an assessment of RCTs in breast cancer, Liberati and colleagues found that the average quality score was 50% of the maximum. After the principal investigators of the trials had been interviewed, the quality increased only marginally, to 57% (35). I am not aware of similar studies on meta-analyses and systematic reviews. One should also consider that when authors are asked what they did, their replies may not be completely accurate but might tend to be what ‘looks good.’

In a letter in BMJ criticising our work, the author argued that the difference in methodological quality between Cochrane reviews and paper based meta-analysis was mainly due to space restrictions in journals compared to almost no restrictions in Cochrane Library (36). Several studies have found that Cochrane reviews are of higher methodological quality on average than other reviews (37-41), but Cochrane reviews published in the Cochrane Library do not seem to differ from Cochrane reviews published
also in paper based journals (42,43). Therefore, word limitation is likely to have played a
minor role for our results, and a more reasonable explanation is that The Cochrane
Collaboration since its formation has been leading in setting the methodological standards
for systematic reviews and meta-analyses. Most of the authors of the PRISMA statement on
good reporting of systematic reviews (44), which most scientific journals adhere to, have
been or are involved with the collaboration.

We also compared the conclusions of Cochrane reviews and industry-supported reviews
and found that the latter had uniformly positive recommendations for the experimental drug
despite the results in the meta-analyses being similar. The industry-supported reviews had
no reservations regarding the risk of bias in the included trials or the cost of the drugs in
contrast to none of the Cochrane reviews. Reviews with undeclared support and reviews
with not for profit or no support had conclusions that were similar to the conclusions in
Cochrane reviews.

Our findings were reinforced by a study of a large cohort of meta-analyses on
antihypertensive drugs published before 2004, which found that meta-analyses with
favourable conclusions were associated with financial ties to a drug company and the
association increased when adjusting for methodological quality (45). Another study on
review articles and meta-analyses on passive smoking found that the conclusions were
strongly associated with author affiliation to the tobacco industry (46)

Numerous studies have documented that financial conflicts of interest influence the study
outcome and the evidence has been synthesized in several systematic reviews (15-18). In
industry-supported trials, both the conclusions and the results seem to favour the sponsor's
product (18) when compared to other trials. Differences in design such as study objectives
and the selection of dosage and type of administration of the study drugs can partly explain
this (47). Compared to other trials, industry-supported trials are more often placebo
controlled (48), probably because they are less expensive to conduct than active controlled
trials and because the medicines regulatory authorities almost only require this type of
research for marketing approval (48). We, as well as another study of meta-analyses (45),
did not find a relation between funding source and results. Different meta-analyses of the
same intervention and disease are likely to include the same studies since most trials in
meta-analyses are published, and despite the challenges in data extraction and analysis,
meta-analyses can often be replicated to a reasonable degree by other researchers (49,50).
Re-analysing trials requires the trial protocol and patient data on file and these are only
rarely publicly available. Therefore, the risk of undetected data manipulation, selective
outcome reporting and publication bias seems to be higher in trials than in meta-analyses,
and this may be a part of the explanation that the results in industry-funded trials, but not in
meta-analyses, are more in favour of the sponsor’s product compared to other sources of support.

The undetected bias in trials is the Achilles’ heel of meta-analyses (and systematic reviews). Turner et al. conducted two separate meta-analyses of the same antidepressant trials (51). One with trial results from FDA reviews and one with results from trial reports published in medical journals. The authors found that the effect size in the journals was 32% larger than when all FDA registered trials were included, ranging from 11 to 69% for individual drugs. In addition, 23 of 74 trials registered by FDA were not published and only one of these had positive results whereas the rest had negative or questionable results (51). Therefore, as suggested by Jefferson et al., meta-analyses and systematic reviews that rely on the published literature only run the risk of being an “unsolicited authoritative advertising for the drug industry” (52).
PAPER 2

The aim of this study was similar to the first, but with a broadened scope. We compared all industry-supported meta-analyses of drug-drug comparisons with those without industry support indexed in PubMed and published in 2004 for differences in methodological quality and conclusions.
Industry-supported meta-analyses compared with meta-analyses with non-profit or no support: Differences in methodological quality and conclusions

Anders W Jørgensen*, Katja L Maric, Britta Tendal, Annesofie Faurschou and Peter C Gøtzsche

Address: The Nordic Cochrane Centre, Department 3343, Rigshospitalet, Blegdamsvej 9, DK-2100 Copenhagen, Denmark

Email: Anders W Jørgensen* - awj@cochrane.dk; Katja L Maric - k.mrc@hotmail.com; Britta Tendal - bt@cochrane.dk; Annesofie Faurschou - af16@bbh.regionh.dk; Peter C Gøtzsche - pcg@cochrane.dk

* Corresponding author

Abstract

**Background:** Studies have shown that industry-sponsored meta-analyses of drugs lack scientific rigour and have biased conclusions. However, these studies have been restricted to certain medical specialities. We compared all industry-supported meta-analyses of drug-drug comparisons with those without industry support.

**Methods:** We searched PubMed for all meta-analyses that compared different drugs or classes of drugs published in 2004. Two authors assessed the meta-analyses and independently extracted data. We used a validated scale for judging the methodological quality and a binary scale for judging conclusions. We divided the meta-analyses according to the type of support in 3 categories: industry-supported, non-profit support or no support, and undeclared support.

**Results:** We included 39 meta-analyses. Ten had industry support, 18 non-profit or no support, and 11 undeclared support. On a 0–7 scale, the median quality score was 6 for meta-analyses with non-profit or no support and 2.5 for the industry-supported meta-analyses (P < 0.01). Compared with industry-supported meta-analyses, more meta-analyses with non-profit or no support avoided bias in the selection of studies (P = 0.01), more often stated the search methods used to find studies (P = 0.02), searched comprehensively (P < 0.01), reported criteria for assessing the validity of the studies (P = 0.02), used appropriate criteria (P = 0.04), described methods of allocation concealment (P = 0.05), described methods of blinding (P = 0.05), and described excluded patients (P = 0.08) and studies (P = 0.15). Forty percent of the industry-supported meta-analyses recommended the experimental drug without reservations, compared with 22% of the meta-analyses with non-profit or no support (P = 0.57).

In a sensitivity analysis, we contacted the authors of the meta-analyses with undeclared support. Eight who replied that they had not received industry funding were added to those with non-profit or no support, and 3 who did not reply were added to those with industry support. This analysis did not change the results much.

**Conclusion:** Transparency is essential for readers to make their own judgment about medical interventions guided by the results of meta-analyses. We found that industry-supported meta-analyses are less transparent than meta-analyses with non-profit support or no support.
Background
The primary source of knowledge used by physicians is the medical literature [1]. The amount of information on health care interventions is hardly manageable, however, and the need for systematic reviews and meta-analyses to aggregate the information and give a basis for rational decision-making is therefore obvious [2]. Systematic reviews and meta-analyses of randomised controlled trials are considered to be the highest level of evidence [3].

Bias in industry-sponsored drug trials is common and the sponsor's product is often favoured [4-7]. Concerning systematic reviews and meta-analyses, little information has been available on sponsor bias until recently. We found that, although the results were similar, industry-supported meta-analyses lacked scientific rigour and were more likely to recommend the experimental drug, compared to Cochrane reviews of the same disease and drugs [8]. A study on anti-hypertensive drugs found that meta-analyses with financial ties to only one drug company were associated with conclusions in favour of that company [9].

For this report, we assessed all meta-analyses indexed on PubMed of drug-drug comparisons published in 2004. We hypothesized that meta-analyses supported by the pharmaceutical industry are of poorer methodological quality and have conclusions favouring the experimental drug, compared to meta-analyses with non-profit or no support.

Methods
We collected meta-analyses that compared different drugs or classes of drugs, were published in full and in English. A meta-analysis was defined as a review with quantitatively combined data from at least two studies. Meta-analyses that we had authored ourselves and meta-analyses of placebo controlled trials were excluded.

Literature search
We searched PubMed for meta-analyses published in 2004 in English. We used a validated search strategy [10] and combined it with "Drug Therapy" [MeSH] OR "drug therapy" [Subheading] OR "Pharmaceutical Preparations" [MeSH] OR "drugs".

One author (KLM) reviewed the titles and abstracts of all potentially eligible meta-analyses, and if in doubt, the full text of the article was retrieved. Another author (AWJ) assessed all the included and 10% of the excluded meta-analyses for eligibility. Disagreements were resolved by discussion.

Data extraction
Two authors independently assessed the included meta-analyses. Data extraction was done unblinded. A third author resolved disagreements.

We used a pilot tested data sheet and extracted data on drugs and diseases; types of support; sources used to identify trials for the review; searches for unpublished trials; and descriptions of concealment of allocation, details of blinding, and excluded patients and trials.

We assessed the methodological quality of the reviews with the Oxman and Guyatt index, which is a validated tool with nine items considering the potential for bias and an overall assessment on a 0–7 scale [11-13]. Furthermore, we judged the review authors' conclusions by assessing whether the experimental intervention was recommended without reservations, or whether it was not recommended (or recommended with reservations) [6].

Data analysis
We divided the meta-analyses according to the type of support in 3 categories: industry-supported, non-profit support or no support, and undeclared support.

Industry support was defined as authorship, provision of grants to authors of the meta-analysis, or other major assistance such as help with the statistical analysis. We did not consider provision of references or unpublished trial reports as support.

We compared industry-supported meta-analyses with meta-analyses with non-profit support or no support with the Mann-Whitney test for categorical data and with Fisher’s exact test for binary data; P values are two-sided.

Results
The search in PubMed identified 1188 records of meta-analyses. Most were ineligible because they did not compare drugs (Figure 1). We included 39 meta-analyses, 10 of which were Cochrane reviews. Ten had industry support, 18 non-profit or no support, and 11 undeclared support (Table 1 and additional file 1: References of included meta-analyses).

Methodological quality
The median quality score was 6 for the 18 meta-analyses with non-profit or no support and 2.5 for the ten industry-supported meta-analyses (P < 0.01) (Table 2). More meta-analyses with non-profit or no support avoided bias in the selection of studies (P = 0.01); only five industry-supported meta-analyses reported specific inclusion criteria and two only included studies provided by the supporting company. Meta-analyses with non-profit or no support more often stated the search methods used to find studies.
(P = 0.02) and searched comprehensively (P < 0.01). Only four (40%) industry-supported meta-analyses searched for unpublished studies and five (50%) searched in the Cochrane Library and MEDLINE. For meta-analyses with non-profit or no support the numbers were 16 (89%) and 14 (78%) respectively. Meta-analyses with non-profit or no support more often reported criteria for assessing the validity of the studies (P = 0.02), used appropriate criteria (P = 0.04), described methods of allocation concealment (P = 0.05), described methods of blinding (P = 0.05), and

<table>
<thead>
<tr>
<th>Ref #</th>
<th>Medline and CLIB</th>
<th>Unpublished studies</th>
<th>Allocation concealment</th>
<th>Blinding</th>
<th>Excluded patients</th>
<th>Excluded studies</th>
<th>Recommends the experimental drug without reservations</th>
<th>Quality score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>✔</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
</tbody>
</table>

Non-profit or no support

<table>
<thead>
<tr>
<th>Ref #</th>
<th>Medline and CLIB</th>
<th>Unpublished studies</th>
<th>Allocation concealment</th>
<th>Blinding</th>
<th>Excluded patients</th>
<th>Excluded studies</th>
<th>Recommends the experimental drug without reservations</th>
<th>Quality score</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>14</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>16</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>17</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>18</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>19</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7</td>
</tr>
<tr>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
</tbody>
</table>

Support not declared

<table>
<thead>
<tr>
<th>Ref #</th>
<th>Medline and CLIB</th>
<th>Unpublished studies</th>
<th>Allocation concealment</th>
<th>Blinding</th>
<th>Excluded patients</th>
<th>Excluded studies</th>
<th>Recommends the experimental drug without reservations</th>
<th>Quality score</th>
</tr>
</thead>
<tbody>
<tr>
<td>29</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>32</td>
<td>✔</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>35</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>38</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>39</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>

CLIB: The Cochrane Library
described excluded patients (P = 0.08) and studies (P = 0.15).

Authors’ conclusion
Four industry-supported meta-analyses (40%) recommended the experimental drug without reservations, compared with four meta-analyses with non-profit or no support (22%) (P = 0.57). The supporting companies of all 4 meta-analyses that were recommending the experimental drug without reservation were also selling the drug, but not the control drug. This only applied to 3 of the 6 industry-supported meta-analyses that did not recommend or only recommended the experimental drug with reservations.

Post hoc sensitivity analyses
We did two sensitivity analyses. First, we excluded the 10 Cochrane reviews from the analysis, as these reviews had high quality scores (median of 7); nine from the meta-analyses with non-profit or no support and one from those with industry support. This resulted in a median quality score of 3 and 2 respectively (P = 0.06).

Second, we contacted the authors of the 11 meta-analyses with undeclared support. Four declared that they had not received any external funding or other type of support (which we exemplified as help with the statistical analyses), four replied that they had received funding from non-profit organisations and three did not reply. We added the eight who replied to the meta-analyses with non-profit or no support and the three who did not reply to the meta-
analyses with industry support. This sensitivity analysis did not change the results much, e.g. the median quality score was 5 for meta-analyses with non-profit or no support and 2 for industry-supported meta-analyses (P < 0.01)

Discussion
We found that meta-analyses with non-profit or no support are of better methodological quality on average than those with industry support. Lack of allocation concealment and blinding, and high attrition rates in randomised controlled trials may bias results of meta-analyses, but if the authors fail to describe these details, the reader is not able to judge if the meta-analysis is reliable. Most industry-supported meta-analyses failed on these counts; this agrees with results we have published previously [8].

Cochrane reviews
Cochrane reviews seem to have a better methodological quality on average than other meta-analyses [8,14-17].

In this study, ten Cochrane reviews contributed with high methodological quality mainly to meta-analyses with non-profit or no support, as only one Cochrane review was supported by the industry. The policy in the Cochrane Collaboration is that industry support of Cochrane reviews is not allowed [18].

Forty percent of the industry-supported meta-analyses and 22 percent of those with non-profit or no support recommended the experimental drug without reservation. This difference was not statistically significant, but we have previously found a significant difference in a comparison that was closely matched for drugs and diseases (P < 0.01) [8]. A large survey of 124 meta-analyses of antihypertensive drugs found that those with financial ties to only one drug company were significantly associated with conclusions in favour of that company [9].

Studies of trials have found similar results. A survey found that none of 56 trial reports of non-steroidal anti-inflammatory drugs supported by the manufacturer presented results that were unfavourable to the company [19]. Another survey found that trials funded by for profit organizations were more likely to recommend the experimental drug as the drug of choice (odds ratio = 5.3) compared with trials funded by non-profit organisations [6]. And recently a study of randomised trials of head-to-head comparisons of statins with other drugs found that not only the conclusions but also the results were in favour of the sponsor’s product [7].

Limitations
We were not blinded, as the layout of some papers, e.g. Cochrane reviews, is unique and impossible to blind.

---

Table 2: Methodological quality of meta-analyses.

<table>
<thead>
<tr>
<th>Questions</th>
<th>Supported by industry (N = 10)</th>
<th>Non-profit support or no support (N = 18)</th>
<th>P value*</th>
<th>Undeclared support (N = 11)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Were the search methods used to find evidence on the primary question stated?</td>
<td>6</td>
<td>18</td>
<td>0.02</td>
<td>6</td>
</tr>
<tr>
<td>2. Was the search for evidence reasonably comprehensive?</td>
<td>4</td>
<td>17</td>
<td>&lt;0.01</td>
<td>6</td>
</tr>
<tr>
<td>3. Were the criteria used for deciding which studies to include reported?</td>
<td>5</td>
<td>17</td>
<td>0.03</td>
<td>10</td>
</tr>
<tr>
<td>4. Was bias in the selection of studies avoided?</td>
<td>1</td>
<td>12</td>
<td>0.01</td>
<td>3</td>
</tr>
<tr>
<td>5. Were the criteria used for assessing the validity of the included studies reported?</td>
<td>3</td>
<td>15</td>
<td>0.02</td>
<td>5</td>
</tr>
<tr>
<td>6. Was the validity of all studies referred to in the text assessed using appropriate criteria?</td>
<td>1</td>
<td>10</td>
<td>0.04</td>
<td>1</td>
</tr>
<tr>
<td>7. Were the methods used to combine the findings of the relevant studies reported?</td>
<td>9</td>
<td>17</td>
<td>1.00</td>
<td>10</td>
</tr>
<tr>
<td>8. Were the findings of the relevant studies combined appropriately?</td>
<td>7</td>
<td>16</td>
<td>0.46</td>
<td>8</td>
</tr>
<tr>
<td>9. Were the conclusions made by the author(s) supported by the data reported?</td>
<td>9</td>
<td>15</td>
<td>1.00</td>
<td>9</td>
</tr>
<tr>
<td>Overall quality (1–7) (median score)</td>
<td>2.5</td>
<td>6</td>
<td>&lt;0.01</td>
<td>3</td>
</tr>
</tbody>
</table>

Number of meta-analyses that obtained an affirmative answer to the questions of the Oxman and Guyatt index; and median overall quality score.

* Supported by industry vs non-profit or no support.
However, blinding when assessing meta-analyses has little impact [20].

A substantial part of the meta-analyses with non-profit or no support were Cochrane reviews and these are made according to the Cochrane Handbook [21]. Andy Oxman participated in the development of the validated index that was used for evaluating methodological quality and he has also participated in developing the Cochrane Handbook.

Our definition of industry support does not distinguish between different amounts of support, and our judgement of support is based on details reported in the meta-analyses. This can theoretically lead to misclassification of the support, as industry support may range from very little to generous, and details about some types of support may be lacking more often than others. However, the definition is operational and we believe that it includes the most important types of industry support. Lack of details or transparency in meta-analyses may also have led to misjudgement of the methodological quality, and it has been argued that the methodological superiority of Cochrane reviews can be explained by the fact that there are no word limits in the Cochrane Library. However, the methodological quality of Cochrane reviews published in regular journals do not seem to differ from Cochrane reviews published in The Cochrane Library [16,17]. Furthermore, important methodological details should always be made available in journals with a word limit, either in the article itself, or in material on the journal’s website.

Conclusion
Transparency is essential for readers to make their own judgment about medical interventions guided by the results of meta-analyses. We found that industry-supported meta-analyses are less transparent than meta-analyses with non-profit support or no support.

Competing interests
All authors are affiliated with the Nordic Cochrane Centre. We did not receive any funding or support for this study.

Authors’ contributions
AWJ drafted the manuscript and PCG helped. AWJ and KLM participated in writing the protocol, the literature search, data extraction and statistical analysis. BT, AF and PCG participated in data extraction. All authors reviewed and approved the final manuscript.

Additional material

Additional file 1
References of included meta-analyses
Click here for file
[http://www.biomedcentral.com/content-supplementary/1471-2288-8-60-S1.pdf]

References

Pre-publication history
The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1471-2288/8/60/prepub
DISCUSSION OF PAPER 2

Compared to the first project, the second was similar, but with a broadened scope. We assessed a cohort of meta-analyses from 2004, but apart from that we used the same methods and compared industry-supported meta-analyses with meta-analyses with non-profit or no support for differences in conclusions and methodological quality. Overall, the meta-analyses with non-profit or no support were of better methodological quality than industry-supported meta-analyses. The major contributor to a high median quality score was Cochrane reviews and when they were excluded, the differences were smaller but still apparent (P=0.06). Therefore, one may argue that the difference in methodological quality was more related to the high quality of Cochrane reviews than to source of support.

It is true that Cochrane reviews had a great impact on our results, but it seems that the relation between methodological quality and source of support was not only confounded by Cochrane reviews. From a publication by Yank et al. (45) in which the authors assessed meta-analyses of anti-hypertension drugs, it was possible for me to compare those with financial ties to one or more drug companies (n=72) with other meta-analyses (n=52). Only one of these meta-analyses was a Cochrane review. The authors also used the quality instrument by Oxman and Guyatt to assess methodological quality, but in a modified form, which meant that the meta-analyses with the lowest quality scored 0 and the ones with the highest quality scored 18. The median quality score of meta-analyses with financial ties to one or more drug companies was 4.5 and it was 9 for other meta-analyses (P=0.03, Mann-Whitney U test). The median quality score for meta-analyses with ties to only one company was 3.

Contrary to meta-analyses, industry-supported trials do not have poorer methodological quality than other trials and there is some evidence that they are of better quality (16). An explanation for this can be that some of the development of trial methodology has been done within the drug industry and that the regulatory authorities primarily require submission of high quality trials and only under special circumstances also meta-analyses (53). In contrast to this the development of meta-analyses and systematic reviews has been outside the medical industry. Over the recent two decades the Cochrane Collaboration, which is independent of the medical industry, has been the major contributor to the development of systematic reviews and meta-analyses.

It has been claimed that meta-analyses conducted by the industry are mainly for internal decision-making or regulatory purposes and from a drug company’s point of view - when conducting a meta-analysis of their own drugs - systematic searches of the literature is not important, as the company has access to all trials of that drug (54). This requires that the
company have a complete database of all trials and have access to the trials conducted by independent researchers, but this is not always the case.

When the authors of a Cochrane review on neuraminidase inhibitors for influenza searched for trials they identified a large oseltamivir trial by Roche Shanghai that Roche Basel claimed it did not know about. It was omitted from Roche’s list of 101 sponsored and supported trials despite the existence of an English language study report (52).

The ICH9 (International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use) guideline only mentions the possibilities of including a meta-analysis in the application for marketing approval without discussing the methodology of meta-analyses and systematic reviews. Therefore it seems that the medical industry has failed to acknowledge the important methodological aspects that are required when conducting meta-analyses enabling readers to make their own judgement about the results and conclusions. This is crucial, especially considering that industry meta-analyses seem to have more favourable conclusions than other meta-analyses.
In the third study, we quantified how often benefits and harms in terms of gastrointestinal bleeding and cardiovascular thrombosis were reported in abstracts on rofecoxib, how often the drug was described favourably, and how this pattern changed over time.
Unbalanced reporting of benefits and harms in abstracts on rofecoxib

Anders W. Jørgensen · Karsten Juul Jørgensen · Peter C. Gøtzsche

Received: 19 August 2009 /Accepted: 19 January 2010 /Published online: 17 February 2010
© Springer-Verlag 2010

Abstract

Purpose It was predicted from the mechanism of action that, compared to older non-steroidal anti-inflammatory drugs, rofecoxib (Vioxx) would reduce gastrointestinal bleeding, but also that it would increase the occurrence of cardiovascular thrombosis. From the patient’s point of view, both effects are important and should be investigated and reported similarly. We studied how they have been reported over time.

Methods We searched PubMed for abstracts on rofecoxib that commented on gastrointestinal bleeding or cardiovascular thrombosis or both. Two researchers, blinded to date of publication and authors, assessed the abstracts independently. We judged the authors’ view on rofecoxib and comments on gastrointestinal bleeding and thrombosis as being favourable, neutral or unfavourable towards rofecoxib.

Results We included 393 abstracts commenting on gastrointestinal bleeding (72%) and cardiovascular thrombosis (54%) or both. Before October 2000, all abstracts (n=27) mentioned only gastrointestinal bleeding and 89% were positive towards rofecoxib. The year before the withdrawal of rofecoxib (October 2003 to September 2004) (n=46), 59% of abstracts commented on gastrointestinal bleeding only, 17% on thrombosis only, 24% on both and 67% were still positive. From October 2006 to September 2007 (n=54), 13% mentioned gastrointestinal bleeding, 54% thrombosis, 33% mentioned both and only 11% were positive.

Conclusions The reporting of benefits and harms was not balanced and changed markedly over time. Knowledge of increased risk of thrombosis existed early on, but the harms came into focus too late, when the drug was already withdrawn, and when tens of thousands of patients had been harmed unnecessarily.

Keywords Rofecoxib · Cyclooxygenase-2 inhibitors · Abstracts · Bias · Thrombosis · Gastrointestinal bleeding

Introduction

Rofecoxib was marketed in 1999 as first-line treatment of osteoarthritis with the claim that it reduced pain as effectively as conventional non-steroidal anti-inflammatory drugs (NSAIDs), but with less gastrointestinal bleeding. However, the drug also caused serious thrombotic cardiovascular events [1, 2]. On 30 September 2004, Merck, the manufacturer, withdrew it from the market. Rofecoxib has been estimated to have caused the death of tens of thousands of patients because of thromboses [3].

The part of a paper that is most often read is the abstract and sometimes clinical decisions are based solely on abstracts [4, 5]. The recently published CONSORT guideline for abstracts states that any important adverse (or unexpected) effects of an intervention should be described in the abstract [5]. We quantified how often benefits and harms in terms of gastrointestinal bleeding and cardiovascular thrombosis, respectively, were reported in abstracts on rofecoxib, how often the drug was described favourably, and how this pattern changed over time.
Methods

Search strategy and data extraction

One author searched PubMed (24 September 2007) using the search terms “Vioxx OR rofecoxib”. All records with an abstract were assessed for inclusion by two observers independently. At the same time, they extracted the data using a pilot-tested data sheet. Any disagreements were settled by discussion. The observers were blinded to any information about authors and institutions and to the date of publication, and assessed only the title and the text of the abstract. The blinding was obtained by exporting the relevant parts of the PubMed records into Microsoft Excel.

Inclusion and exclusion criteria

We included abstracts that commented on the effect of rofecoxib on gastrointestinal bleeding or cardiovascular thrombosis or both, and contained a comment reflecting the authors’ view on rofecoxib. We accepted abstracts that implicitly referred to these harms by using phrases such as “gastrointestinal adverse effect” and “cardiovascular risk”. We also included abstracts that commented on the effect of cyclooxygenase 2 (COX-2) inhibitors and other NSAIDs in general when these included rofecoxib, albeit indirectly, e.g. “celecoxib had no tangible advantage in terms of serious gastrointestinal complications … overall mortality was higher with celecoxib than in the placebo group. The difference was similar to that observed in placebo-controlled trials of rofecoxib in Alzheimer’s disease.”

We excluded abstracts that only commented on harms that did not involve gastrointestinal bleeding or cardiovascular thrombosis, such as nausea or hypertension. We also excluded abstracts that did not have a comment reflecting the authors view on rofecoxib, e.g. “The manufacturer claims that in clinical studies rofecoxib inhibits COX-2 but not COX-1, has the power of high-dose NSAIDs—diclofenac and ibuprofen—and superior GI [gastrointestinal] safety profile compared to conventional NSAIDs.”

Abstracts of in vitro studies, animal studies, medical devices and pharmacokinetics were also excluded.

Evaluation of the authors’ view on rofecoxib

We categorised the authors’ view on rofecoxib as either favourable, neutral or unfavourable. If an active comparator was not used as a reference, we accepted placebo or statements that did not involve a comparator. The judgement was preferentially made using the conclusion of the abstract. If this was not possible, we used statements in the results section or elsewhere, e.g. “Because of its more favourable gastrointestinal toxicity profile compared with non-selective NSAIDs, rofecoxib is safer in patients …”.

We also judged the individual comments on gastrointestinal bleeding and cardiovascular thrombosis in the same way.

Analysis

We used graphs and descriptive statistics to assess how the reporting in abstracts changed over time. Because rofecoxib was withdrawn on 30 September 2004, our 1-year intervals are from October to September, e.g. year 2000 was defined as October 1999 to September 2000, both months included. For abstracts that only contained information about the year of publication, we used the date the citation was added to the PubMed database [EDAT].

Results

Our PubMed search identified 2,047 records and we included 393 abstracts. Most records were excluded because there was no abstract in PubMed or because the abstract did not contain a comment on gastrointestinal bleeding or cardiovascular thrombosis (Fig. 1). Twenty-nine of the excluded abstracts mentioned hypertension, but did not comment on cardiovascular thrombosis.

Reporting of harms over time

During the whole observation period, 181 of the included abstracts (46%) commented on gastrointestinal bleeding only, 110 (28%) on cardiovascular thrombosis only and 102 (26%) commented on both. Of the 283 (181+102) abstracts commenting on gastrointestinal bleeding, 141 (50%) used the explicit terms “ulcer”, “gastrointestinal bleeding” or “perforation”, or “serious gastrointestinal adverse effect”. The remaining 142 abstracts (50%) used the less explicit terms “gastrointestinal risk”, “gastrointestinal safety”, or “gastrointestinal adverse effect”. Of the 212 (110+102) abstracts commenting on cardiovascular thrombosis, 137 (65%) used the explicit terms “thrombosis”, “thrombolytic” or “thrombotic effect”, “myocardial infarction” or “stroke”. The remaining 75 abstracts (35%) used the less explicit terms “cardiovascular risk”, “cardiovascular safety” or “cardiovascular adverse effect”.

Until and including September 2000, no abstracts commented on cardiovascular thrombosis (the first was published in November 2000), i.e. 100% (n=27) of the abstracts commented only on gastrointestinal bleeding (Fig. 2). The percentage of abstracts that only commented on gastrointestinal bleeding decreased to 59% in 2004 (n=46), before rofecoxib was withdrawn, and to 13% in 2007 (n=54). The percentage of abstracts that commented only on cardiovascular thrombosis increased from 0 to 17% in
2004 and to 54% in 2007. The percentage commenting on both gastrointestinal bleeding and cardiovascular thrombosis increased from 0 to 24% in 2004, and to 33% in 2007. The greatest change in reporting was seen immediately after the withdrawal of rofecoxib in 2004 (Fig. 2).

Authors’ general view on rofecoxib over time

Until and including September 2000 (n=27), the proportion of abstracts favouring rofecoxib was 89%. In 2004 (n=46), it was 67%. The greatest change was seen after the withdrawal of rofecoxib and in 2007 (n=54) where only 11% of the abstracts were positive (Fig. 3). We investigated the robustness of this result by including only those abstracts where our judgement was based on the conclusion section of the abstracts. This graph had a similar slope (dotted line in Fig. 3).

Authors’ views on harms before and after the withdrawal of rofecoxib

Two hundred and eleven abstracts (54%) were published before the withdrawal of rofecoxib in 2004 and 182 (46%) were published after.
Before the withdrawal, 193 (91%) abstracts commented on gastrointestinal bleeding, and 168 (87%) of them were favourable towards rofecoxib, 19 (10%) were neutral and 6 (3%) unfavourable. Fifty-six (27%) abstracts commented on cardiovascular thrombosis, and 5 (9%) of those were favourable towards rofecoxib, 31 (55%) neutral and 20 (36%) unfavourable.

After the withdrawal, the effect on gastrointestinal bleeding was mentioned in 89 (49%) abstracts, and 67 (75%) of those were favourable towards rofecoxib, 14 (16%) neutral and 8 (9%) unfavourable. The thrombotic effect was mentioned in 156 (86%) abstracts and, none of them were favourable towards rofecoxib, 26 (17%) neutral and 130 (83%) unfavourable.

Discussion

We found that most abstracts on rofecoxib reported only on the beneficial effect regarding less gastrointestinal bleeding, and that they were generally in favour of rofecoxib, from the introduction of the drug in 1999 to its withdrawal in 2004. After the withdrawal, most abstracts reported on the harmful effects, cardiovascular thrombosis, and few were in favour of rofecoxib.

Such findings might be expected for drugs with important but rare harms that are unknown when the drugs are introduced on the market and only discovered later. However, this is not the only explanation for our findings. It has been documented that the company suppressed cardiovascular harms in the scientific literature [6] and intimidated researchers and speakers who were critical of rofecoxib [6–8].

Before its introduction, it was predicted from the mechanism of action that the drug should reduce the incidence of gastrointestinal bleeding [9] but also increase the incidence of thrombosis, compared with non-selective NSAIDs [10–12]. Two trials conducted by Merck, 090 [3, 13, 14] and VIGOR [15], both showed that rofecoxib increased the risk of cardiovascular events significantly. However, the first trial, which ended in 1999, was not published in a scientific journal until 2006 [14]. The second trial was published in the New England Journal of Medicine, but the increased risk of myocardial infarction was interpreted as a beneficial aspirin-like prophylactic effect of the control NSAID [15]. This interpretation was speculative and was later refuted. Furthermore, three cases of myocardial infarction in the rofecoxib arm had been omitted from the paper [16].

In 2001, it was documented in a systematic review that COX-2 inhibitors increased the risk of cardiovascular events [17], and a cumulative meta-analysis of trials from 2004 showed that a clear relationship between rofecoxib and increased risk of myocardial infarction existed already.
Neurological disorders
Hemicrania continua
Schizophrenia
Sclerosis
Alzheimers dementia
Migraine
Premenstrual migraine

Surgery
Prevention of urethral strictures after TURP
Pre-medication for tonsillectomy
Pre-medication for uterine curettage
Hernia operations
Post CABG
Pre-medication for ear-nose-throat surgery in general
Minor dental surgery (e.g. removal of molars)
Minor orthopaedic surgery

Cancer
Treatment for glioblastoma multiforme
Protection against colorectal neoplasia in familiar polyposis
Treatment of malignant melanoma and sarcomas
Treatment of prostate cancer
Treatment of bone cancer
Treatment of breast cancer
Treatment of lung cancer

Other
Reduction of atherosclerosis among ACS-patients post-infarction
Congenial nephrogenous diabetes insipidus
Menstrual pain
Endometriosis
Non-bacterial prostatitis
Haemophilic arthropathy
Premenstrual acne
Prevention of ectopic ossification in arthroplasty

Fig. 4 Conditions for which the effect of rofecoxib was mentioned by the end of 2000 [18]. Two other meta-analyses did not find evidence of an increase in cardiovascular risk with rofecoxib, but they were conducted by employees of Merck [19, 20].

Over the studied time period, there has been an increased focus on harms [21–23] and the quality of reporting trial results in abstracts [5], which may have had an impact on our results. However, our sample of abstracts did not exclusively consist of trial abstracts, and the increased attention to harms does not explain the dramatic change in focus from beneficial to harmful effect when rofecoxib was withdrawn.

The safety data from trials on rofecoxib were far too positive compared to a real-world setting. None of the trials in the application for marketing approval were designed to evaluate the cardiovascular risk [6]. In fact, they included patients that had an unusually low cardiovascular risk. Medicare patients in Tennessee, who were treated with rofecoxib in clinical practice, had a baseline risk of getting a myocardial infarction that was eight times higher than that for the patients in the trials [18]. Patients at high risk of developing peptic ulcers were also often left out of trials on rofecoxib. In 2002, an analysis of cardiovascular adverse events was added to the protocols of three studies [23] including the one [2] that led to the withdrawal of rofecoxib [1]. This was considered breaking news and is likely to have initiated the change in focus from beneficial effects to harms.

Publishing and disseminating scientific papers on medical interventions is an important marketing strategy for the pharmaceutical industry [24], and Merck’s active role in the writing of journal articles is likely to have influenced how rofecoxib was portrayed and perceived by the clinicians. It is difficult to explore Merck’s role in more detail in relation to our results. Merck’s information control could have been clarified by looking at reporting in relation to type of financial support. We did not attempt to do this, as ghost authorship and other forms of support from drug companies are often not revealed in scientific papers [25], and Merck used guest and ghost authors for many of the papers on rofecoxib [26]. Merck also conducted a seeding trial [27], the ADVANTAGE trial, published in Annals of Internal Medicine [28], and sponsored the Australasian Journal of Bone and Joint Medicine, which looked like a peer-reviewed medical journal but was only a marketing tool. Most of the articles in the journal presented data favourable to Merck products, including rofecoxib, without disclosing sponsorship [29].

A strategy to increase drug sales that has been used by Merck [30] and many other drug companies is to stimulate off-label use [24, 31]. This may also be the case for rofecoxib [32] and could explain why many abstracts mentioned or evaluated the effect of rofecoxib in relation to other conditions than arthritis (Fig. 4). After having assessed one-third of the abstracts (n = 1,370), one observer decided to register the conditions (apart from arthritis) that rofecoxib was proposed for. These were mainly neurological disorders, cancer and pain related to minor surgery. The U.S. Food and Drug Administration approved rofecoxib for osteoarthritis, acute pain, primary dysmenorrhoea and rheumatoid arthritis.

We searched for abstracts in PubMed only, as PubMed is the most widely used database for medical research. It is likely that more abstracts would have been included if we had searched additional databases, but we would not expect it to have led to any important changes in our results.

It has been suggested that increases in blood pressure related to rofecoxib are a mechanism for the increase in the risk of cardiovascular events [33]. We excluded 26 abstracts for the reason that they only commented on hypertension, but they would not have changed the results much as they were scattered over the years 2001 to 2007.

We believe that if the reporting of benefits and harms in abstracts is unbalanced, doctors will get a false perception of the drug’s value. In particular, readers need information.
on deaths, and on harms that can be lethal, such as thromboses causes by COX-2 inhibitors. In the drug literature, there is plenty of evidence of flawed research [34–43], ghost-written articles [24, 25, 44–47], intimidation of researchers [44, 45, 48–56] and misleading and false statements in research papers and marketing [24, 44–47, 57–65].

We suggest that studies like ours should be done on other drugs than rofecoxib, preferably with newly marketed drugs associated with high expectations.

Conclusions

The basic principle of balanced reporting of benefits and harms seems to have been seriously distorted in abstracts on rofecoxib, although the harms were equally predictable as the benefits from the mechanism of action of the drug. Before the withdrawal of rofecoxib, abstracts mostly reported on gastrointestinal bleeding and were in favour of rofecoxib. The harms came in focus too late, when the drug had already been withdrawn, and when tens of thousands of patients had been harmed unnecessarily [3, 66, 67].

Funding

None.

References


52. Tanne JH (2007) FDA places “black box” warning on antidiabetes drugs. BMJ 334:1237


DISCUSSION OF PAPER 3

In the third project, we assessed the reporting of cardiovascular thrombosis and gastrointestinal bleeding in abstracts on the cyclooxygenase-2 inhibitor rofecoxib. Rofecoxib was marketed in 1999 under the trade name Vioxx as being as effective as conventional non-steroidal anti-inflammatory drugs (NSAIDs), but with less gastrointestinal bleeding, for treating osteoarthritis and rheumatoid arthritis. It was withdrawn due to serious thrombotic cardiovascular adverse events in 2004.

We found that, despite the manufacturer's awareness of an increased risk of cardiovascular thrombosis, the harm received little attention from the introduction of the drug to its withdrawal. In this period, most abstracts reported on the benefits of less gastrointestinal bleeding and they were in general in favour of rofecoxib. After withdrawal, most abstracts reported on the harmful effects, cardiovascular thrombosis, and few were in favour of rofecoxib.

This study did not have a control or reference drug and therefore the evolution of the descriptions of harms and benefits in abstracts cannot be categorised as more or less extreme than the 'average drug' marketed at the same time. However, a control drug was not necessary to illustrate that important harms of rofecoxib were underreported in abstracts. This is an important problem, as this skews the impression one gets of an intervention, and unfortunately, the problem is general, as it also exists in trials (55), review articles (56,57) and cost benefit analyses (58).

The change in reporting of harms and benefits over time for a particular drug or class of drugs has not previously been studied in abstracts. A more recent study about the evolution of trial methodology in psychopharmacological research found that the quality of reporting the results in abstracts improved over time (59), but it did not specify if the reporting of harms was considered a quality criterion. Nevertheless, there has been an increased focus on harms and the quality of reporting in abstracts, which may have influenced our findings to some degree, although it cannot explain the dramatic change in focus from a beneficial to a harmful effect after rofecoxib was withdrawn from the market.

Some abstracts are more informative than others, e.g. structured abstracts, which often allow more liberal word counts. Informative abstracts usually present the background, objectives, methods, results and conclusions of a survey or a trial and indicative abstracts presents fewer subheadings (60). Some of the abstracts that we included were indicative and due to the nature of those abstracts, important information may have been left out.
Therefore the style of abstracts plays an important role for the quality of information presented.

It can be discussed whether safety results should be presented in the abstract if they were not pre-planned as primary or secondary outcomes. Nevertheless, the CONSORT statement for abstracts mentions that authors should describe any important adverse (or unexpected) effects of an intervention, to which I agree. Harms are not always the primary outcome and generally receive too little attention, as illustrated by the late amendment of an analysis of cardiovascular adverse events to three study reports in 2002 (61). One of them that led to the withdrawal of rofecoxib began in 2000 (62), but a trial finished in 1999 had already shown an increased risk of cardiovascular adverse events. This trial, however, was first published in 2006 (63).
Our primary aim of the fourth study was to analyse the change in body weight and compare the results when using four different methods for handling missing data.
We should not continue to use complete case analysis and single imputations like LOCF and BOCF. A comparison of results from different imputation techniques of missing data from an anti-obesity drug trial.

Anders W Jørgensen (MD)¹, Lars H Lundstrøm (MD, PhD)², Jørn Weterslev (MD, PhD)², Arne Astrup (Professor, DrMedSci)³ and Peter C. Gøtzsche (Professor, DrMedSci)¹,⁴

¹The Nordic Cochrane Centre, Dept 3343, Rigshospitalet, Blegdamsvej 9, DK-2100 Copenhagen, Denmark
²Copenhagen Trial Unit, Copenhagen Centre of Clinical Intervention Research, Dept 3344, Rigshospitalet, Blegdamsvej 9, DK-2100 Copenhagen, Denmark
³Department of Human Nutrition, Faculty of Life Sciences, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg, Denmark
⁴Institute of Medicine and Surgery, Faculty of Health Sciences, University of Copenhagen, Blegdamsvej 3B, DK-2200 Copenhagen, Denmark

Corresponding author:
Anders W Jørgensen
The Nordic Cochrane Centre
Rigshospitalet
Blegdamsvej 9
DK-2100 Copenhagen
Denmark
e-mail: awj@cochrane.dk

Abstract: 222 words

Text: 3766 words
Abstract

Background: In randomised trials of medical interventions the most reliable analysis follows the intention-to-treat (ITT) principle. However, the ITT analysis requires that missing outcome data have to be imputed. Different imputation techniques may give different results and some may lead to bias. In anti-obesity drug trials, many data are missing and the most used imputation method is last observation carried forward (LOCF). LOCF is generally considered conservative, but there are more reliable methods such as multiple imputation (MI).

Objectives: To compare four different methods of handling missing data in a 60-week placebo controlled anti-obesity drug trial on topiramate.

Methods: We compared an analysis of complete cases with datasets where missing body weight measurements had been replaced using three different imputation methods: LOCF, baseline carried forward (BOCF) and MI.

Results: 561 participants were randomised. At the end of the trial, 85% of the participants were lost to follow up. Compared to placebo, there was a significantly greater weight loss with the topiramate in all analyses: 9.5 kg (SE 1.17) in the complete case analysis (N=86), 6.8 kg (SE 0.66) using LOCF (N=561), 6.4 kg (SE 0.90) using MI (N=561) and 1.5 kg (SE 0.28) using BOCF (N=561).

Conclusions: The different imputation methods gave very different results. LOCF was not conservative compared to the more reliable MI, in fact LOCF had a lower SE.

Introduction

Attrition has been described as the bane of clinical trials on anti-obesity drugs (1). In most studies more than one third of the subjects drop out after one year (2), and depending on the type of missingness the results will be biased (3). Missingness can be classified as missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR). Only when missingness is MCAR the use of complete case analysis of the data, obtained exclusively from participants with all data observed, may give an unbiased result because the analysis then is a random sample of the full dataset. However, in most studies MCAR is rarely the case (4) and when missingness is MAR, missingness depends on the observed data. Finally, when missingness is MNAR, it also depends on some unobserved data.

One of the most commonly applied methods for handling attrition in obesity research is ‘last observation carried forward’ (LOCF) (2). In an analysis of the outcome measure with LOCF a missing weight measurement of a participant at the end of trial is replaced by the participant’s last observed value. To assume that one’s weight is unchanged after dropping out of a trial seems hard to justify as participants tend to regain much of their lost weight within a short period of time after having stopped the intervention (5). Therefore, baseline carried forward (BOCF) has been proposed as a more reliable imputation strategy (6). However, both BOCF and LOCF overestimate the precision of the effect estimate because the dataset is analysed, after single imputation of the missing data, as if it was a ‘complete’ dataset with no missing data (7). Intuitively one has doubt in imputed data and thus the p-values and confidence interval should increase. Therefore BOCF and LOCF provide undue certainty of the effect estimate even under MCAR.
There are other and more reliable techniques for handling missing data (8), and they can also be used when missingness is MAR. Attention has been drawn to the multiple imputation (MI) technique (9), which for some time has been recommended for handling missing data in obesity trials (2). The multiple imputation technique is a stepwise procedure. First, based on the observed data a plausible multivariable distribution for the missing values is estimated and they are being replaced by values randomly drawn from this distribution resulting in a complete dataset. Second, the first step is repeated multiple times generating multiple datasets. Third, the datasets are then analysed separately producing multiple estimates, e.g. weight change, and fourth, the multiple estimates are pooled resulting in one single estimate. Compared to LOCF and BOCF the precision in this case will not be unduly overestimated because the uncertainty of the imputed values due to multiple sampling from the distribution is taken into account.

We had access to an individual patient dataset from a large randomised three armed weight loss trial that compared diet plus topiramate with diet plus placebo (10). Our primary aim was to analyse the weight change and compare the results by using four different methods for handling missing data. Thus we analysed the dataset of complete cases and datasets where missing weight measurements had been replaced by three different imputation methods LOCF, BOCF and MI. Our second aim was to report the results of the analysis of the primary outcome measure at the time-point specified in the trial protocol.

Material and Methods

The trial

The trial was a randomised weight maintenance trial (n=561) that compared placebo (n=187) with topiramate 96 mg (n=190) and 192 mg (n=184) per day. It was designed to proceed for a total of 82 weeks; a 8-week non-pharmacological low-calorie diet run-in phase that was followed by the randomisation, a 60-week intervention phase, hereafter a 14-week drug tapering period and lastly a follow-up period. The data included assessments of weight from 26 visits plus standard baseline values (age, height, sex etc.), a variety of blood sample analyses and measures of hip and waist circumferences. Each visit corresponded to a specific number of weeks in the trial. Subjects that withdrew from the study were tapered off intervention over two to three weeks and during that period their weight was measured. More details about the methodology have been published previously (10). However, the sponsor terminated the trial prematurely due to low tolerability to the drug, and therefore the publication only contained results on a subset of patients. The primary outcome measure being percent weight change from enrolment to end of intervention phase has never been published.

Data

Data was provided by the first author (AA) of a previous report of this trial (10) and imported to SPSS 18.0.

Missingness

We assessed the mechanism of missingness by using Little’s test (11) and by plotting weights of subjects with missing data with those with data.
Comparison of imputation methods
We used the data from baseline (randomisation) to week 60. For simplicity we pooled the topiramate arms. We analysed the mean weight change in the placebo and the pooled topiramate group and the difference between the two from baseline to week 60. We also analysed percentage change in the same way. The results were plotted against time for comparisons between the four following methods.

Complete case analysis
This did not involve any imputation and was an analysis of data from participants on intervention, i.e. all available data from baseline and to week 60. Patients that had their weights measured on the visit of interest e.g. week 60, but had missing weight measurements on other visits, e.g. week 40 and 52, were also included.

Last observation carried forward
We substituted missing weight measurements at 60 weeks with the last observed measurement. We allowed carrying forward the baseline value if this was the last observed measurement.

Baseline carried forward
We substituted missing weight measurements with the baseline weight.

Multiple imputation
We imputed the missing weight measurements with values of weight obtained by the ‘Fully conditional specification method’ in SPSS ver. 18. This is an iterative Markov chain Monte Carlo method that can be used when the pattern of missing data is monotone (i.e. a subject attends all visits till a visit is missed and never return) or none-monotone (12). We used a linear regression model that contained the variables: intervention group, sex, race, age and baseline values (height, waist circumference, plasma glucose, triglycerides, HDL-cholesterol, HDL/LDL-ratio, insulin, haemoglobin and haemoglobin A1c). Additionally we included variables for weight, but only at visits prior to the visit of interest. E.g. we only included weight measurements from baseline to week 20 for the analysis on week 20. We log-transformed weight to satisfy the normality assumption which seemed to hold when we assessed Q-Q plots and used the Kolmogorov–Smirnov test (p>0.2, df=561), but not the Shapiro–Wilk test (p=0.04, df=561). We also assessed convergence by plotting the means and standard deviations by iteration and imputation. The default number of 10 iterations in SPSS 18 seemed to be too low (web figure) as the standard deviation gradually increased up till and stabilized at 400 iterations. Therefore we chose 500 iterations and 10 imputations assuring an efficiency of the imputation of 99%.

The primary outcome measure at 60 weeks in the trial protocol
We estimated the mean percent weight change from enrolment to week 60 using the same methods as described above. However, the two topiramate groups receiving different dosages were not pooled. For completeness we also re-analysed the data on the subset of patients previously published (10). The subset was chosen because the authors wanted to reduce the risk of bias due to trial termination and consisted of participants who received at least one dose of study drug, provided at least one post-baseline efficacy evaluation, had the opportunity to complete 44 weeks of treatment before the study closedown announcement and allowed only data collected before the closedown announcement and
up to week 44 to be included in the analysis of efficacy. Using these criteria it was not possible to get the exact same subset, and therefore we allowed data collected 3 days after the closedown announcement to be included in the analysis. However, in our analysis the number of patients in the topiramate 192 mg group (n=105) differed from the original analysis (n=99).

Results
Details about the study population have previously been published (10). Baseline characteristics are available in web table 1 and include the variables used in our analyses after MI.

Missingness
Missing data on weight gradually increased from week 0 to 44 (from 0% to 27%), but onwards missingness increased much more (Figure 1). Only 15% (n=86) of the patients were still on treatment and had their weight recorded on week 60. The reason for missingness varied, but mostly because of trial termination (10). The mechanism was unlikely to be MCAR with a statistical significant Little’s test (P<0.01)(11). On most visits participants who missed the following visit seemed to weigh more than those who attended (Figure 1). During the 60 weeks only 13% of the participants attended all visits and the most frequently occurring patterns of missingness (66% of the participants) were monotone; 17%, 17%, 9% and 9% had missing visits onwards from week 44, 48, 52 and 56 (figure 3).

Comparison of imputation methods over time and with increasing missingness
From baseline to week 44 the estimated difference in mean weight change between placebo and topiramate increased by all four methods. From week 44 to week 60 the difference increased in the complete case analysis and in LOCF, but decreased in MI and BOCF (Figure 3). The complete case analysis estimated the greatest difference in weight loss (-9.5 kg on week 60) and BOCF the smallest (-1.5 kg on week 60). MI and LOCF were similar with MI resulting in a slightly greater difference from the beginning and throughout most of the trial and a slightly smaller difference in the end (-6.4 kg on week 60) when compared to LOCF (-6.8 kg on week 60).

These differences were a result of an overall weight loss in the topiramate group (figure 4) and weight gain in the placebo group (Figure 5). The change within the pooled topiramate group was again biggest in the complete case analysis (-5.9 kg on week 60) and smallest in the BOCF (-0.9 kg on week 60). Also, MI and LOCF estimated a similar weight change in the beginning of the trial, but in the end MI (-3.4 kg on week 60) showed a much smaller weight loss than LOCF (-5.5 kg on week 60) (Figure 4). The change within the placebo group was similar in the complete case analysis and BOCF in the beginning, however in the end the complete case analysis estimated the biggest weight gain (3.7 kg on week 60) and BOCF the smallest (0.5 kg on week 60). In the placebo group MI and LOCF were similar in the beginning, but in the end MI (3.0 kg on week 60) showed a greater weight gain than LOCF (1.3 kg on week 60) (Figure 5).

We got similar results for the analyses of the difference in weight change between the groups and the weight change within the groups when analysing the percentage change from baseline.
The trial's primary outcome measure at 60 weeks
Regardless the methods of analysis participants in the topiramate groups on average lost more weight than participants in the placebo group.

The primary outcome measure at end of trial according to the trial protocol was the percentage change from enrolment to 60 weeks (week –8 to week 60) and for placebo compared to topiramate 96 mg the mean difference was -10.5% (SE = 2.2) in the complete case analysis, -6.1% (SE = 0.7) using LOCF, -5.5% (SE = 1.1) using MI and -1.7% (SE = 0.4) using BOCF. For placebo compared to topiramate 192 mg/day the mean difference was -10.0% (SE = 1.9), -7.5% (SE = 0.8), -7.3% (SE = 1.0) and -1.5% (SE = 0.4) using complete case analysis, LOCF, MI and BOCF respectively (Table 1).

When we re-analysed the percentage change from enrolment to week 44 of the subset of participants our results were similar, but not identical to those published previously (webtable 2). The mean difference between placebo and topiramate 96 mg was -7.4% in the complete case analysis, -6.2% using LOCF, -6.0% using MI and -5.6% using BOCF. For placebo compared to topiramate 192 mg the mean difference was -8.0%, -7.5%, -7.6% and -6.4% using complete case analysis, LOCF, MI and BOCF respectively.

Discussion
We compared 4 methods of analysing the effect of topiramate in a weight loss and maintenance trial. We found that in the beginning of the trial LOCF and MI estimated similar body weight changes, but over time and with high attrition they estimated different body weight changes. This however, did not have a big effect on the difference in body weight change between topiramate and placebo, which was similar throughout the trial. Complete case analysis estimated the greatest difference and BOCF the smallest. We also estimated the weight change as original planned in the trial protocol and as it was done in a previous publication of the trial (10), and our results regardless of method confirmed that in this weight maintenance trial topiramate produces a greater weight-loss than placebo. Other placebo-controlled trials have shown that topiramate reduces weight (13).

In several simulation studies MI has been shown to provide both a confidence interval with better coverage (realistic precision) and a less biased estimate of the intervention effect than both complete case analysis and single imputations using single imputations (14)(15)(16). In these studies the method has been to simulate missingness from a complete dataset and further with different imputation techniques trying to amend the missingness calculating an estimate of the intervention effect from the imputed dataset. Accordingly, having the correct estimate of the intervention effect from the complete dataset, these studies have been able to evaluate how close the result of an imputation method comes to the correct estimate from the original complete dataset. The conclusion of such simulation studies have been overwhelmingly in favour of MI.

In trials on anti-obesity drugs it has previously been shown that LOCF estimate similar effect sizes (2), but overestimates the precision compared to MI (8) and therefore describing LOCF as a conservative analysis is wrong.
Our results can be interpreted in relation to the per-protocol (PP) hypothesis that assumes that all participants adhere to treatment, which is an unrealistic scenario in trials on anti-obesity drugs, or the intention-to-treat (ITT) hypothesis that assumes that some participants do not adhere to treatment. Thus the PP analysis estimates the weight loss as if the participants adhere to the treatment and the ITT analysis estimates the weight loss of the intention to give the treatment regardless of adherence (17).

If a drug truly reduces the weight, then BOCF will result in a conservative estimate (smaller weight loss) compared to a PP analysis, but maybe a realistic estimate compared to a ITT analysis (6), as participants tend to regain some of their lost weight within a after ended treatment (5).

The interpretation of the LOCF analyses is difficult due to the course of weight change during a weight loss trial. Most of the weight loss occurs early during treatment, then levels out and some is regained in the end of a trial (18). Compared with a PP analysis, it may be reasonable to assume that LOCF underestimates the weight loss in the short term and overestimates it in the long term. On the other hand, compared with the ITT hypothesis, it is likely that LOCF overestimates the weight loss, as BOCF seems more realistic.

Our MI analysis is more compatible with a PP analysis than an ITT analysis and should be interpreted as such (17). We had weight data for some of the withdrawn participants from a 2 to 3 weeks tapering off period in the MI analysis, however during that period they were still on treatment, but on a smaller dose, and therefore the result is still a good estimation of the weight loss as if all participants had continued the treatment. If we had had complete data on some of the participants that dropped out or did not adhere to the treatment, we could have used their data for MI, which then would have reflected an ITT analysis (17).

The complete case analysis can be an unbiased PP analysis, but only when the participants in the analysis can be regarded as a random sample of the study population (when the missing mechanism is MCAR). In our dataset the missing mechanism was not MCAR. It seemed that those who had missing data weighted more than those without and thus complete case analysis overestimated the weight loss. MCAR is only rarely the case in clinical trials (4).

Our results should also be interpreted in relation to how results from weight loss trials are being reported in scientific articles in general. Often, only the weight change within the treatment groups is stated, and sometimes also the difference in weight change between the groups is stated. We have reported both and found that MI and LOCF resulted in similar differences between topiramate and placebo, but that the weight change within the groups was smaller in the MI analysis than in LOCF and this difference increased over time and with increasing missing data. The most likely explanation for this is that LOCF, but not MI, ignores the course of weight change (described above). Figure 4 and 5 both show that the graphs for MI is more or less a translation of the complete case analysis, suggesting that the bias LOCF introduce compared to MI is more or less the same in both treatment groups, and therefore MI and LOCF estimates similar difference between the groups. Further, MI introduced greater uncertainty of the intervention effect estimate than both the single imputations LOCF and BOCF. Obviously LOCF is biased towards the
alternative hypothesis being true and BOCF is biased towards the null hypothesis being true and both methods have unduly high estimates of the precision (table 1 & figure 3-5).

**Limitation**

We do not know the true treatment effect of topiramate or the weight of the missing participants and thus it is impossible to validate our findings. Also we do not know the exact mechanism of missingness. If the mechanism is MNAR all imputation methods are likely to be biased. However, MI may provide less biased results in this situation as well (16). Most participants dropped out if the trial because of trial termination and this is not a common reason for attrition in obesity trials. However the reason for termination was safety and tolerability issues with the drug. We did not include safety or harm in our imputation model and thus missingness may be at risk of depending on data that are ‘unobserved’ by the MI procedure. However, the harms may still be correlated to the variables in our model and figure 1 showed that missingness could be predicted by the weight of the participants. Further, MI has been found to be a reliable imputation method in weight loss trials when only relying on measurements over time (2).

**Suggestions for future research**

The major reason for missing data when participants’ quit the treatment is their lack of interest for having their weight measured and for obvious reasons trialists cannot force the patient to show up for a visit. Therefore we need weight loss trials that include incentives for participants to be followed-up and other logistic methods for measuring the weight and side effects of those participants who decide to quit treatment.

Trialists are aware of this problem and most trial protocols specify that investigators shall do their best to have a last measurement at the end of the specified follow up period, but do not provide the incentives or methods. The protocol for the trial in this study stated:

“Participants withdrawn from the study prior to completion of treatment period will be encouraged […] to attend study visits with assessments equivalent to those performed at Visit 23 (Week 60)”

A similar statement was in the protocol for the RIO-North America trial, a weight loss study of rimonabant (19):

“For patients with premature treatment discontinuation or patients considered lost to follow-up, the [case report form] must be filled in up to the last visit performed. The Investigator should make every effort to re-contact and to identify the reason why the patient failed to attend the visit and to determine his/her health status and to retrieve study medication.”

When the trial was published it was criticised in an editorial for only measuring those participants who completed the trial (53%) rather than all patients who were not lost to follow-up (93%) (18). It is indeed possible to have a high follow-up as seen in a recently published weight loss trial of free meals and an intensive weight loss program. After 24 month the investigators had weight data on 92% of the study participants (20).
Sometimes some of the participants that withdrew from a trial are measured and analysed, but unfortunately not published. In a placebo controlled trial of long-term treatment with sibutramine the authors only published an unadjusted ITT analysis (LOCF) (n=464) (21) that showed a greater weight reduction with the drug compared to an adjusted analysis of all available patients (including those who withdrew) (n=205) which was available in a trial report submitted to the Danish Medicines Agency.

We believe that it is possible to get complete weight data for some and a final planned weight measurement for many of the withdrawn participants. When imputation is undertaken reliable techniques, such as MI, should be used and to make a proper ITT analysis imputation should also be based on weight data collected after withdrawal. We can only then provide a meaningful answer to the question of how much weight one can expect to loose on average when using a particular intervention.

**Conclusion**
The different imputation methods gave different results, but all showed that topiramate reduced weight compared to placebo. In anti obesity trials imputation is obligatory due to the amount and type of missing data and because the complete case analysis is biased. However, the ITT analysis using LOCF, which in general is considered a conservative analysis, was in this trial similar to the PP analysis that used the more reliable MI method and therefore we suggest that post withdrawal weight data must be obtained to make a proper ITT analysis and MI should replace the standard imputation strategies LOCF and BOCF because they overestimate the precision.


19. Pi-Sunyer FX, Aronne LJ, Heshmati HM, Devin J, Rosenstock J. Effect of rimonabant, a cannabinoid-1 receptor blocker, on weight and cardiometabolic risk factors in overweight or obese patients: RIO-North America: a randomized controlled trial. JAMA 2006;295:761-775.


Figure 1. Mean weight of patients attending and missing the following visit. Some of the missing patients return and therefore week 28 and 32 both include 448 patients.

Figure 2. The most frequently occurring patterns of missingness.
Figure 3. Analysis of difference in weight change between placebo and topiramate pooled (96 and 192 mg/day) from baseline to week 60 using different methods.
Figure 4. Analysis of weight loss over time in topiramate pooled (96 or 192 mg/day) group using different imputation methods

- **Completers**
- **LOCF**
- **MI (all data)**
- **BOCF**

% subjects missing
Figure 5. Analysis of weight loss over time in the placebo group using different imputation methods.

Analysis of weight loss over time in the placebo group using different imputation methods.
Table 1. Percentage weight change from enrolment (-8 week) to 60 weeks

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Placebo</th>
<th>Topiramate 96 mg</th>
<th>Topiramate 192 mg</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SE</td>
<td>Mean</td>
</tr>
<tr>
<td>Completers</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Change, %</td>
<td>(n=28)</td>
<td>(n=31)</td>
<td>(n=27)</td>
</tr>
<tr>
<td></td>
<td>4,2</td>
<td>1,13</td>
<td>-6,3</td>
</tr>
<tr>
<td></td>
<td>-6,3</td>
<td>1,82</td>
<td>-5,8</td>
</tr>
<tr>
<td>LOCF</td>
<td>(n=187)</td>
<td>(n=190)</td>
<td>(n=184)</td>
</tr>
<tr>
<td>Change, %</td>
<td>1,3</td>
<td>0,52</td>
<td>-4,8</td>
</tr>
<tr>
<td></td>
<td>-4,8</td>
<td>0,53</td>
<td>-6,2</td>
</tr>
<tr>
<td>MI</td>
<td>(n=187)</td>
<td>(n=190)</td>
<td>(n=184)</td>
</tr>
<tr>
<td>Change, %</td>
<td>3,0</td>
<td>0,69</td>
<td>-2,5</td>
</tr>
<tr>
<td></td>
<td>-2,5</td>
<td>0,79</td>
<td>-4,3</td>
</tr>
<tr>
<td>BOCF</td>
<td>(n=187)</td>
<td>(n=190)</td>
<td>(n=184)</td>
</tr>
<tr>
<td>Change, %</td>
<td>0,6</td>
<td>0,20</td>
<td>-1,0</td>
</tr>
<tr>
<td></td>
<td>0,6</td>
<td>0,20</td>
<td>-0,9</td>
</tr>
</tbody>
</table>

SE: standard error
Each difference (topiramate - placebo) had a P-value < 0.001 (Unpaired T-test)
Web material
Web figure 1. Convergence chart showing the mean and standard deviation of the imputed values of log-transformed weight measurements at each iteration of the multiple imputation method for each of the 5 requested imputations.

The purpose of this plot is to look for patterns in the lines. There should not be any, and these should look suitably “random” which is the case for the mean, but note that the standard deviation increases up to iteration number 450 and the reaches a plateau.
# Webtable 1. Baseline values for variables included in the model for multiple imputation.

<table>
<thead>
<tr>
<th></th>
<th>Placebo</th>
<th>Topiramate 96 mg</th>
<th>Topiramate 192 mg</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of subjects</td>
<td>187</td>
<td>190</td>
<td>184</td>
<td>561</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male, no. (%)</td>
<td>39 (21)</td>
<td>53 (28)</td>
<td>44 (24)</td>
<td>136 (24)</td>
</tr>
<tr>
<td>Female, no. (%)</td>
<td>148 (79)</td>
<td>137 (72)</td>
<td>140 (76)</td>
<td>425 (76)</td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White, no. (%)</td>
<td>182 (97)</td>
<td>188 (99)</td>
<td>181 (98)</td>
<td>551 (98)</td>
</tr>
<tr>
<td>Black, no. (%)</td>
<td>4 (2)</td>
<td>1 (0.5)</td>
<td>2 (1)</td>
<td>3 (1)</td>
</tr>
<tr>
<td>Asian, no. (%)</td>
<td>1 (1)</td>
<td>1 (0.5)</td>
<td>1 (1)</td>
<td>7 (1)</td>
</tr>
<tr>
<td><strong>Mean baseline values</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age, years (SD)</td>
<td>44 (11)</td>
<td>44 (10)</td>
<td>44 (11)</td>
<td>44 (11)</td>
</tr>
<tr>
<td>Weight, kg (SD)</td>
<td>96.6 (14.7)</td>
<td>98.8 (13.8)</td>
<td>99.3 (15.1)</td>
<td>98.2 (14.5)</td>
</tr>
<tr>
<td>Height, cm (SD)</td>
<td>167 (9)</td>
<td>168 (9)</td>
<td>168 (9)</td>
<td>168 (9)</td>
</tr>
<tr>
<td>Waist size, cm (SD)</td>
<td>105 (12)</td>
<td>106 (11)</td>
<td>107 (11)</td>
<td>106 (11)</td>
</tr>
<tr>
<td>Plasma glucose, mmol/L (SD)*</td>
<td>6.6 (1.8)</td>
<td>6.6 (1.8)</td>
<td>6.7 (1.8)</td>
<td>6.6 (1.8)</td>
</tr>
<tr>
<td>Triglycerides, mmol/L (SD)</td>
<td>1.2 (0.5)</td>
<td>1.1 (0.6)</td>
<td>1.3 (0.8)</td>
<td>1.2 (0.6)</td>
</tr>
<tr>
<td>HDL cholesterol, mmol/L (SD)</td>
<td>1.2 (0.3)</td>
<td>1.2 (0.3)</td>
<td>1.2 (0.3)</td>
<td>1.2 (0.3)</td>
</tr>
<tr>
<td>HDL/LDL cholesterol, ratio, (SD)</td>
<td>2.8 (1.0)</td>
<td>2.9 (1.0)</td>
<td>2.8 (0.9)</td>
<td>2.8 (1.0)</td>
</tr>
<tr>
<td>Insulin, mcU/ml (SD)</td>
<td>10.2 (5.0)</td>
<td>10.7 (8.4)</td>
<td>11.2 (6.2)</td>
<td>10.7 (6.7)</td>
</tr>
<tr>
<td>Hemoglobin, g/L (SD)</td>
<td>143 (13)</td>
<td>144 (12)</td>
<td>143 (13)</td>
<td>143.3 (12.3)</td>
</tr>
<tr>
<td>Hemoglobin 1Ac, percent (SD)</td>
<td>5.5 (0.5)</td>
<td>5.5 (0.5)</td>
<td>5.5 (0.5)</td>
<td>5.5 (0.5)</td>
</tr>
</tbody>
</table>

*Oral glucose intolerance test

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Placebo</th>
<th>Topiramate 96 mg</th>
<th>Topiramate 192 mg</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean SE</td>
<td>Mean SE CI95%</td>
<td>Mean SE CI95%</td>
</tr>
<tr>
<td>Completers</td>
<td>(n=83)</td>
<td>(n=71)</td>
<td>(n=81)</td>
</tr>
<tr>
<td>Change, %</td>
<td>-9.6 0.78</td>
<td>-17.0 0.81</td>
<td>-17.6 0.98</td>
</tr>
<tr>
<td>Difference,% (topiramate - placebo)</td>
<td>6.0 1.3 5.6 10.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LOCF</td>
<td>(n=99)</td>
<td>(n=96)</td>
<td>(n=105)</td>
</tr>
<tr>
<td>Change, %</td>
<td>-9.1 0.69</td>
<td>-15.3 0.71</td>
<td>-16.6 0.84</td>
</tr>
<tr>
<td>Difference,% (topiramate - placebo)</td>
<td>7.5 1.10 5.4 9.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MI</td>
<td>(n=99)</td>
<td>(n=96)</td>
<td>(n=105)</td>
</tr>
<tr>
<td>Change, %</td>
<td>-8.8 0.77</td>
<td>-14.7 0.86</td>
<td>-16.4 0.93</td>
</tr>
<tr>
<td>Difference,% (topiramate - placebo)</td>
<td>7.6 1.23 5.2 10.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOCF</td>
<td>(n=99)</td>
<td>(n=96)</td>
<td>(n=105)</td>
</tr>
<tr>
<td>Change, %</td>
<td>-9.6 0.65</td>
<td>-15.1 0.69</td>
<td>-16.0 0.81</td>
</tr>
<tr>
<td>Difference,% (topiramate - placebo)</td>
<td>6.4 1.05 4.4 8.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

SE: standard error
CI95%: 95% confidence interval
Each difference (topiramate - placebo) had a P-value < 0.001 (Unpaired T-test)
In this project, we compared 4 methods of analysing the effect of a weight-reducing drug in a weight maintenance trial. We found that, over time and with increasing attrition, the difference in body weight change between the 4 methods also increased. When attrition was high LOCF and MI estimated different body weight changes within the treatment groups, but similar differences between the groups. Complete case analysis estimated the greatest difference and baseline carried forward (BOCF) the smallest. As expected, LOCF and BOCF both had higher precision than MI and the complete case analysis.

A major limitation to our study is that we do not know the true treatment effect of topiramate or the body weight of the missing participants and it is therefore not possible to validate our findings against a gold standard - the correct result. Also, we do not know the exact mechanism of missingness. Missingness can be classified as missing completely at random (MCAR), missing at random (MAR), or missing not at random (MNAR). When missingness is MCAR the available data can be regarded as a random sample of the ‘full’ dataset and only in this case the complete case analysis exclusively on participants with all data observed gives an unbiased result. However, in most studies MCAR is rarely the case. When missingness is MAR, missingness depends on the observed data and in this case MI can make reliable imputation of the missing data using the distribution of the observed. Finally, when missingness is MNAR, missingness also depends on unobserved data. For obvious reasons unobserved data cannot be included in the imputation method and therefore all methods are likely to be biased. Nevertheless, MI may provide less biased results in this situation as well, compared to the other methods (64) and has been found to be a reliable imputation method in trials of anti-obesity drugs (24).

In general, MI is considered one of the most reliable methods for imputing missing data. It has been available for more than 20 years and in standard statistical programs for about 10 years. But MI is rarely used in randomised trials and has only recently been proposed for trials on anti-obesity drugs (24). There has been and still is a steadfast tradition of using LOCF in these trials, but policy is changing. The European Medicines Agency has from 2011 implemented a guideline for handling missing data in clinical trials (65) that the drug industry has to follow. The guideline does not find LOCF, (but BOCF) appropriate for chronic conditions such as obesity that will return to baseline when the treatment is stopped and it describes MI as a more proper imputation technique. Therefore, I assume that we will see more trials using BOCF and MI in the future. With these changes, drug companies will be more motivated for minimising attrition and missing data because increasing attrition and missing data will result in a treatment effect that approaches zero when BOCF is used.
Likewise, there will be a larger proportion of withdrawn patients with complete data as their data are required for a proper ITT analysis.

A Cochrane review on long-term use of the anti-obesity drugs sibutramine, orlistat and rimonabant found that all trials reported ITT-analyses using LOCF. The authors of the review preferred this analysis to a complete case analysis, as it is more conservative, but acknowledge that the bias introduced with LOCF can go in both directions, and due to high attrition rates they conclude that more methodological rigorous trials are needed (66). An average attrition rate of 37% at 1 year in trials on anti-obesity drugs (67) is high, but not compared to the 'real world'. A survey of orlistat and sibutramine users in British Columbia, Canada, found that the persistence rates with the drugs were 6% and 8%, respectively, after one year. Therefore the high attrition rates in trials on anti-obesity drugs may not be the biggest problem as they reflect the 'real world', whereas failure to follow up on participants that discontinue the treatment and the use of LOCF comprise major methodological problems. It shouldn't be difficult to obtain a final body weight even for patients who dropped out of the study.

Authors of systematic reviews and meta-analyses have methods for imputing missing outcome data in trials but they are of limited quality and the best methods require individual patient data (68). Individual patient data are only rarely available. Therefore, the solution to the problem with missing data remains with those who design and conduct the trial. It is indeed possible to have low rates of missing data, as seen in a recently published weight loss trial of free meal plus motivated structured weight loss program compared with usual care. After 24 months, the investigators had weight data on 92% of the study participants (69). Most weight loss drug trials also include a diet or behavioural guidance, which can be used to minimise missing weight data. Further, many trial protocols describe that outcome data should be obtained from withdrawn patients.
The aim of the fifth paper was to provide other researchers with arguments for increased transparency in drug regulation and to inform them about EMA’s widened access to documents such as trial reports and protocols.
Opening up data at the European Medicines Agency

Widespread selective reporting of research results means we don’t know the true benefits and harms of prescribed drugs. Peter Gøtzsche and Anders Jørgensen describe their efforts to get access to unpublished trial reports from the European Medicines Agency

Peter C Gøtzsche professor, Anders W Jørgensen PhD student

Nordic Cochrane Centre, Rigshospitalet and University of Copenhagen, Dept 3343, Blegdamsvej 9, DK-2100 Copenhagen Ø, Denmark

Doctors cannot choose the best treatments for their patients despite the existence of hundreds of thousands of randomised trials. The main reason is that research results are being reported selectively. Comparisons of published drug trials with unpublished data available at drug regulatory agencies have shown that the benefits of drugs have been much over-rated and the harms under-rated. Comparisons of trial protocols with published papers have also shown widespread selective reporting of favourable results.

Selective reporting can have disastrous consequences. Rofecoxib (Vioxx) has probably caused about 100 000 unnecessary heart attacks in the United States alone, and class 1 antiarrhythmic drugs probably caused the premature death of about 50 000 Americans each year in the 1980s. An early trial found nine deaths among patients taking the antiarrhythmic drug and only one among those taking placebo, but it was never published because the company abandoned the drug for commercial reasons.

Allowing researchers access to unpublished trial reports submitted to drug regulatory agencies is important for public health. Such reports are very detailed and provide more reliable data than published papers, but it has been virtually impossible to get access to them. We eventually succeeded in getting access to reports held by the European Medicines Agency (EMA) after three years of trying. Our case has set an important precedent, and we summarise here the process and the arguments.

Our application for access

On 29 June 2007 we applied for access to the clinical study reports and corresponding protocols for 15 placebo controlled trials of two anti-obesity drugs, rimonabant and orlistat. The manufacturers had submitted the reports to the EMA to obtain marketing approval in the European Union. We explained that we wanted to explore the robustness of the results by adjusting for the many missing data on weight loss and to study selective publication by comparing protocols and unpublished results with those in published reports.

The information was important for patients because anti-obesity pills are controversial. The effect on weight loss in the published trials is small, and the harms are substantial. People have died from cardiac and pulmonary complications or have experienced psychiatric disturbances, including suicidal events, and most of the drugs have been deregistered for safety reasons.

A basic principle in the European Union is to allow its citizens the widest possible access to the documents its agencies possess. But there are exemptions, and the EMA refuses access if disclosure would threaten commercial interests unless there is an over-riding public interest.

We argued in our first letter to the EMA that secrecy was not in the best interests of the patients because biased reporting of drug trials is common. Furthermore, we hadn’t found any information that could compromise commercial interests in 44 trial protocols of industry initiated trials we had reviewed previously.

Without any comment on our arguments, the EMA replied that the documents could not be released because it would undermine commercial interests. We appealed to the EMA’s executive director, Thomas Lönngren, and asked him to explain why the EMA considered that the commercial interests of the drug industry should over-ride the welfare of patients. We argued that the EMA’s attitude increased the risk of patients dying because their doctors prescribed drugs for them without knowing what the true benefits and harms were. He sent us a similar letter to the EMA’s first letter, ignoring our request for clarification, and told us we could lodge a complaint with the European ombudsman, which we did.

Over the following three years the EMA put forward several arguments to avoid disclosing the documents: protection of commercial interests, no over-riding public interest, the administrative burden involved, or the worthlessness of the data to us after the EMA had redacted them. It also did not respond to the ombudsman’s letters before his rather generous deadlines had run out.
Box 1: Basic principles on citizens’ access to EU documents

“Any citizen of the Union, and any natural or legal person residing or having its registered office in a Member State, has a right of access to documents of the institutions, subject to the principles, conditions and limits defined in this Regulation.”

“Openness enables citizens to participate more closely in the decision-making process and guarantees that the administration enjoys greater legitimacy and is more effective and more accountable to the citizen in a democratic system. Openness contributes to strengthening the principles of democracy and respect for fundamental rights as laid down in Article 6 of the EU Treaty and in the Charter of Fundamental Rights of the European Union.”

Box 2: The path to the data

The delays on our part amounted to 130 days (11% of the time); we awaited replies for 1028 days.

29 Jun 2007: We asked the EMA to provide access to the clinical study reports and their corresponding protocols on rimonabant and sibutramine

20 Aug 2007: The EMA replied that the documents could not be released because they came under the exception of commercial interests

24 Aug 2007: We explained that the EMA’s lack of transparency violated basic principles in the EU treaty and that it leads to suboptimal treatment of patients

17 Sept 2007: With no comment on our arguments, the EMA referred again to commercial interests and noted we could institute court proceedings against the EMA or complain to the European ombudsman

8 Oct 2007: We appealed to the ombudsman, noting that the published literature on drugs is flawed and arguing that protocols and study reports did not disclose anything that could undermine commercial interests

30 Jan 2008: The EMA replied to two letters from the ombudsman, referred to protection of commercial interests and mentioned that it could not identify any over-riding public interest that could justify disclosure of the requested documents

26 Feb 2008: We told the ombudsman that the EMA had failed to explain why commercial interests would be undermined

28 Apr 2008: The EMA replied to the ombudsman that it needed to protect the data against unfair commercial use; that evaluating the balance between benefits and risks of medicines is the EMA’s job; and that redaction of personal data would cause disproportionate effort

17 Jun 2008: In our reply to the ombudsman, we argued against this and noted that if commercial success depends on withholding data that are important for rational decision making by doctors and patients, there is something fundamentally wrong with our priorities in healthcare

22 Jan 2009: The ombudsman proposes a friendly solution to the EMA and asks it to grant us access to the documents or provide a convincing explanation why such access cannot be granted

26 Feb 2009: The EMA restates the commercial interests; claims that we have not given evidence of an over-riding public interest; and refers to the workload involved in redacting the documents

10 Mar 2009: The ombudsman again proposes a friendly solution to the EMA and asks it to clarify its reasoning

7 Apr 2009: The EMA repeats its previous arguments.

19 May 2009: We again counter the EMA’s arguments: the EMA has provided no evidence that the documents are commercially sensitive; many patients had been harmed by selective publication of trial data on COX 2 inhibitors; and redacting the documents should be quick and easy

31 Aug 2009: We tell the ombudsman that we have received trial data from the Danish Medical Agency on a third anti-obesity drug, sibutramine

6 Oct 2009: The ombudsman goes to the EMA to inspect the documents we had requested

19 May 2010: The ombudsman issues a draft recommendation that the EMA should grant us access to the documents or provide a convincing explanation as to why not

7 Jun 2010: In a press release the ombudsman accuses the EMA of maladministration because of its refusal to grant access

31 Aug 2010: The EMA informs the ombudsman that it will provide access

1 Feb 2011: We receive the data

Protection of commercial interests

Protection of commercial interests was the EMA’s over-riding argument. It would undermine the protection of commercial interests to allow us access, it said, as the documents represented the full details of the clinical development programme and the most substantial part of the applicant’s investment. Competitors could use them as a basis for developing the same or a similar drug and gather valuable information on the long term clinical development strategy of the company to their own economic advantage.

We explained that the clinical study reports and protocols are based on well known principles that can be applied to any drug trial; that the clinical study reports describe the clinical effects of drugs; and that nothing in the EMA’s guidelines for preparation of such reports indicates that any information included in them can be considered a trade secret. The trial protocols are always sent to the clinical investigators, and it is unlikely that companies would have left in any information that could be of commercial value (such as a description of the drug synthesis). We also noted that the clinical study reports and trial protocols represent the last phase of drug development, which
Over-riding public interest in disclosure

Even if commercial interests were undermined by disclosure, access would still have to be granted if there was an over-riding public interest. The EMA argued that it could not identify any over-riding public interest and remarked that the evaluation of safety and efficacy of drugs is its responsibility—the EMA constantly monitors drugs and updates its assessment reports and requires changes in product information as appropriate.

We considered this insufficient. Monitoring adverse effects reported by doctors to drug agencies would not have revealed that rofecoxib causes heart attacks. Few such events are reported, and heart attacks are common in people with arthritis.

Postmarketing passive surveillance systems can therefore usually not detect whether a drug leads to more heart attacks than expected; randomised trials are needed for this.

We provided more evidence of the detrimental effects of selective publication but to no avail. The EMA continued to claim that we had not documented the existence of an over-riding public interest. We noted that we could not prove this in this specific case because we were denied access to the data, but we drew attention to the fact that the total number of patients in the main clinical studies of orlistat differed according to the source of the information: published reports, the EMA’s website, and the website of the US Food and Drug Administration.

The ombudsman indicated that we had established an over-riding public interest, but he did not take a definitive stance on whether an over-riding public interest existed because this question needed answering only if disclosure undermined commercial interests. He asked the EMA to justify its position that there wasn’t an over-riding public interest, but the EMA avoided replying by saying that we had not given evidence of the existence of such an interest. We believe that we had.

Furthermore, the EMA’s argument was irrelevant. A suspect asked for his alibi on the day of the crime doesn’t get off the hook by asking for someone else’s alibi.

Administrative burden

According to the EMA, the redaction of (unspecified) “personal data” would cause the EMA a disproportionate effort that would divert attention from its core business, as it would mean redacting 300 000-400 000 pages. This was surprising. The Danish Drug Agency had not seen the workload as a problem when it granted us access to the reports for the anti-obesity drug sibutramine, which was locally approved in Denmark. The 56 study reports we received comprised 14 309 pages in total, and we requested only 15 study reports from the EMA (the pivotal studies described in the European Public Assessment Reports (EPARs) on rimonabant and orlistat). The ombudsman declared that the EMA had overestimated the administrative burden involved.

Worthlessness of data after redaction

The EMA argued that, “as a result of the redaction exercise, the documents will be deprived of all the relevant information and the remaining parts of them will be worthless for the interest of the complainant.”

From what we know of clinical trial reports and protocols it struck us as odd that they would contain so much personal data that the documents became worthless. The ombudsman noted that the requested documents do not identify patients by name but by their identification and test centre numbers, and he concluded that the only personal data are those identifying the study authors and principal investigators and to redact this information would be quick and easy.

The EMA also remarked that a possible future release of the assessment reports of the EMA’s Committee for Medicinal Products for Human Use and the (co)rapporteur assessment reports “could satisfy the request of the complainants.” These reports were not available and they would have been worthless to us because they are merely summaries used for regulatory decisions.

Maladministration

The EMA was completely resistant to our arguments and those from the ombudsman. However, after the ombudsman accused the EMA of maladministration in a press release on 7 June 2010,16 three years after our request, the EMA reversed its stance. The EMA now gave the impression that it had favoured disclosure all the time, agreed with the ombudsman’s reasoning, and noted that the same principles would be applied for future requests for access but that it would consider the need to redact part of the documents.

The EMA’s last letter was unclear: “The Agency will do its utmost to implement its decision as quickly as possible, in any case within the next 3 months at the latest. The Agency will keep the European Ombudsman promptly informed of the exact implementation date.”

It was not clear whether the three months was the deadline for sending the reports to us, for implementing its new policy, or both. We received the data we requested from the EMA on 1 February 2011, which in some cases included individual patient data in anonymised format, identified by individual and test centre numbers.

Concluding remarks

According to the EMA’s responses to the ombudsman, the EMA put protecting the profits of the drug companies ahead of protecting the lives and welfare of patients. Moreover the EMA’s position is inconsistent because it resisted requests to give access to trial data on adult patients while providing access to data on paediatric trials, in accordance with EU legislation.16 The Declaration of Helsinki gives authors the duty to make publicly available the results of their research on humans.17 The declaration also says that, “Medical research involving human subjects must . . . be based on a thorough knowledge of the
scientific literature.” If the knowledge base is incomplete, patients may suffer and cannot give fully informed consent and research resources are wasted. The EMA should be promoting access to full information that will aid rational decision making, not impede it.

Our case sets an important precedent. On 30 November 2010 the EMA declared it would widen public access to documents, including trial reports and protocols. We recommend that the FDA and other drug regulatory agencies should follow suit. Access should be prompt—for example, within three months of the regulator’s decision—and documents should be provided in a useful format. Drug agencies should get rid of the huge paper mountains and require electronic submissions from the drug companies, including the raw data, which should also be made publically available.

Contributors and sources: PCG wrote the first draft, both authors commented on it and are guarantors.

Funding: Sygekassernes Helsefond and Rigshospitalet, University of Copenhagen.

Competing interests: All authors have completed the unified competing interest form at www.icmje.org/coiDisclosure.pdf (available on request from the corresponding author) and declare no support from any organisation for the submitted work; no financial relationships with any organisation that might have an interest in the submitted work in the previous three years; and no other relationships or activities that could appear to have influenced the submitted work.

All documents in this case (133 pages) are available at www.cochrane.dk/research/EMA, together with a comprehensive 26 page report of the case including 54 references.

Provenance and peer review: Not commissioned; externally peer reviewed.


Accepted: 7 March 2011

Cite this as: BMJ 2011;342:d2686

Reprints: http://journals.bmj.com/cgi/reprintform
Subscribe: http://resources.bmj.com/bmj/subscribers/how-to-subscribe
DISCUSSION OF PAPER 5

In the fifth project, we described how we managed to get access to unpublished trial reports and protocols at the European Medicines Agency (EMA). In June 2007, we applied for access to the clinical study reports and their corresponding protocols of the placebo controlled trials of four anti-obesity pills, rimonabant, orlistat, sibutramine and amfepramone. The first two were submitted by the manufacturers to EMA for obtaining marketing approval in the European Union and the last two to the Danish Medicines Agency (DMA). Shortly after our application, the DMA told us that amfepramone was approved in the 1960’s and that DMA did not have the requested documents on this drug. However, we still hoped to get access to the three other drugs, as anti-obesity drugs are controversial. The effect on weight loss in the published trials is small (66) and the harms are substantial. People have died from cardiac and pulmonary complications or have experienced psychiatric disturbances including suicide attempts, and most of the drugs have been pulled off the market for safety reasons.

In August 2009, the DMA provided us with the trial reports on sibutramine, but not the protocols, as they were not part of the submission to DMA. In February 2011, we received both trial reports and protocols of orlistat and rimonabant from EMA.

EMA and DMA had a very different view on our application and handled it differently. EMA rejected our request immediately whereas DMA was very positive. DMA granted us access to the documents 12 month after our application was submitted, but EMA did that 36 month after our submission. In both cases, a third party got involved. Abbott, the producer of sibutramine, appealed DMA’s decision on releasing the documents to the Danish Ministry of Health and we appealed EMA’s decision to the European Ombudsman.

In both cases, the primary argument for not allowing us access to the documents was that it would undermine the protection of commercial interests, as the documents represented the full details of the clinical development programme. Competitors could use them as a basis for developing similar drugs and gather valuable information on the development strategy of the company. We explained to EMA that the documents are based on well-known principles that can be applied to any drug trial; that the clinical study reports describe the clinical effects of drugs; and that nothing in EMA’s guidelines for preparation of such reports indicates that any information included in them can be considered a trade secret. We also noted that unpublished trial data are generally less positive than published ones.

The manufacturer of sibutramine was assisted by a professor from the University of Copenhagen who came up with arguments against the release of the documents.
Apparently, he had knowledge about the information in the documents and argued that Neurosearch could use the information in their development of a similar product, tesofensine. This professor was a member of the tesofensine advisory board. He was also the primary investigator of a large tesofensine trial and the primary investigator of a large sibutramine trial, and it is common that drug companies allow doctors to be members of advisory boards, consultants and investigators for competing companies. I really find this a strange clash of competing interests and wonder what precautions the companies take to protect their interests.

We received the documents from both EMA and DMA, but the outcome was different. We only received the core trial reports from DMA without appendices, as these were not available. EMA provided us with the full trial reports including appendices. The DMA had redacted a substantially amount of ‘personal data’. This was patient numbers and descriptions of side effects. The reason for this was that DMA claimed that the patient numbers could be used to identify the patients and the descriptions of side effects were similar to health records. Also, investigator names in the documents from DMA were redacted, but only because they were regarded as commercially confidential. The European Ombudsman concluded that the only “personal data” are those identifying the study authors and principal investigators. EMA only redacted the curriculum vitae of the investigators and therefore, we received documents with names and addresses of investigators and individual patient data listings, including detailed descriptions of side effects before any coding had taken place.

Further DMA sent the documents in binders whereas EMA sent pdf files. The latter makes electronic searches possible and navigation through the many pages easier. However, it can be improved. The preferable way must be a database with individual anonymised patient data that can be imported to standard statistical software. Independent researchers would then be able to make their own analyses and conclusions.
Systematic reviews, randomised trials and abstracts are key resources for evidence-based medicine, but they should always be read with caution because important information may have been unreported for various reasons.

In the two first studies, we found that industry-supported meta-analyses are less transparent, describe fewer limitations in the included trials and have more favourable conclusions than other meta-analyses. Ignorance due to lack of expertise or unwillingness due to commercial interests seem to be the best possible explanations for the low transparency and methodological quality in industry-supported meta-analyses, whereas there is much evidence in support of commercial interests being the main reason for biased conclusions. Unpublished data and selective outcome reporting is probably the greatest threat to systematic reviews and meta-analyses.

In the third study of abstracts on rofecoxib, we found that the basic principle of balanced reporting of benefits and harms was generally violated, although the harms were equally predictable as the benefits from the mechanism of action of the drug. Before the withdrawal of rofecoxib, abstracts mostly reported on benefits and were in favour of rofecoxib. Only after the drug had been withdrawn, the harms came in focus, and only due to court trials unpublished data were revealed that showed that the harms were evident early on, long before the drug was withdrawn.

In the fourth study, we found that various imputation methods give different results. In most trials, imputation is obligatory because the complete case analysis is biased. The general perception that LOCF is a conservative analysis is wrong. It overestimates the precision and sometimes also the treatment effect. Therefore the more reliable MI should be used, but to get the most reliable result, trial investigators should do their utmost to collect outcome data on all trial participants and also report the results.

In the fifth paper, we conclude that there is something fundamentally wrong with our priorities in health care if commercial success of the drug industry is dependent on withholding data that are needed for rational decision-making for doctors and patients.
Healthcare professionals, politicians and patients should realize that commercial interests play an important role in how healthcare information is assembled and disseminated. They should also prioritise independent and transparent research, which, for example is one of the aims of the Cochrane Collaboration.

All research should be read with caution. Systematic reviews and meta-analyses should report details of concealment of allocation, blinding, inclusion and exclusion criteria for trials, search strategies, estimated effects, and review authors should assess the risk of bias in each included trial and should also have their protocols published to allow readers to judge the reliability of the reviews. This has been specified in the PRISMA statement for reporting systematic reviews and meta-analyses (44).

Plans for handling missing data in trials should not only be pre-specified in the protocol, but should also be based on evidence. Authors of trial protocols should discuss whether imputation is needed and if so also justify the method. The SPIRIT initiative (Standard Protocols Items for Randomised Trials) is currently preparing a guideline for writing trial protocols, which needs to discuss imputations.

Our access to clinical trial protocols and study reports sets an important precedent and on 30 November 2010, EMA declared it would widen public access to documents. We recommend FDA and other drug agencies to follow suit. The access should be quick (e.g. 3 months) and in a useful format. Drug agencies should get rid of the huge paper loads and require electronic submissions from the drug companies, including the raw data, which should also be made publicly available.
RECOMMENDATIONS FOR FUTURE RESEARCH

Research on systematic reviews should focus on the implications of having access to unpublished trial reports. One study suggested it is not worth the effort to search for unpublished trials because they have poor or unclear methodological quality (70), but other studies clearly show that including unpublished trial reports in meta-analyses will impact on their results (51, 71, 72).

The increased focus on harms and the attempts to improve the quality of abstracts will probably result in a more balanced reporting of benefits and harms. One way to study this is to compare the reporting of benefits and harms in abstracts on new drugs with abstracts on old drugs.

The increased access to unpublished study reports and detailed descriptions of harms gives the possibility of assessing how harms are coded. The medical industry is required to use the Medical Dictionary for Regulatory Activities (MEDDRA) to classify harms in randomised trials. MEDDRA is being developed continually, but only little research seems to have been done on inter-coder variation. A study of coding harms in an HIV trial found a 12% difference between 2 coders’ decision on the preferred term for harms (73). A study on the reliability of reading electrocardiograms found much disagreement between cardiologists. Four cardiologists read 537 ECGs and 20% of were found compatible with coronary heart disease by at least 1 observer, but all readers only agreed on 4% (74). Therefore it is likely that the inter-coder variation will increase with more coders, but it needs to be studied.
REFERENCES


43. Collier A, Heilig L, Schilling L, Williams H, Dellavalle RP. Cochrane Skin Group systematic reviews are more methodologically rigorous than other systematic reviews in dermatology. Br J Dermatol 2006;155:1230-1235.


