FACULTY OF HEALTH AND MEDICAL SCIENCES
UNIVERSITY OF COPENHAGEN

THE NORDIC COCHRANE CENTRE

PhD thesis

Jeppe Bennekou Schroll

Unpublished data, particularly in relation to harms, in clinical trials

Academic advisor: Peter C. Gøtzsche

Submitted: 05/10/14

Institutnavn:	Det Nordiske Cochrane Center
Name of department:	The Nordic Cochrane Centre
Author:	Jeppe Bennekou Schroll
Titel og evt. undertitel:	Upublicerede data i kliniske forsøg, med særligt fokus på skadevirkninger
Title / Subtitle:	Unpublished data, particularly in relation to harms, in clinical trials
Subject description:	This thesis investigates how adverse events are handled in clinical trials and looks into whether unpublished data might be a feasible source of data on harms. Different sources of unpublished data are explored, for instance public summary reports from the US and European drug agencies as well as clinical study reports written by the drug sponsors.
Academic advisor:	Peter C. Gøtzsche
Submitted:	05. october 2014

Contents

Preface and acknowledgements *4*
Preface 4
Acknowledgements 4
Financial support 5
English summary *6*
Danish summary *8*
Introduction and objectives *10*
Harms 10
Unpublished data 12
Objectives 13
Description of the research project *14*
Paper 1. Coding 14
Paper 2. Searching Unpublished data 14
Paper 3. Deaths in trials 15
Paper 4. FDA and EMA reports 15
Paper 5. Harms in orlistat trials 15
Summary of the results *16*
Paper 1. Coding 16
Paper 2. Searching Unpublished data 17
Paper 3. Deaths in trials 18
Paper 4. FDA and EMA reports 20
Paper 5. Harms in orlistat trials 20
Conclusion *24*
Word list *26*
References *27*
Paper 1. Coding *31*
Paper 2. Searching Unpublished data *38*
Paper 3. Deaths in trials *49*
Paper 4. FDA and EMA reports *51*
Paper 5. Harms in orlistat trials *57*

Preface and acknowledgements

Preface

This thesis is based on work done during my employment at the Nordic Cochrane Centre from 2011 to 2014. During my employment I attended a three-month fellowship at the San Francisco branch of the US Cochrane Centre. The thesis is a synopsis based on the following 5 articles:

(1) Schroll JB, Maund E, Gøtzsche PC. Challenges in coding adverse events in clinical trials: a systematic review. PLoS One. 2012;7(7):e41174.

(2) Schroll JB, Bero L, Gøtzsche PC. Searching for unpublished data for Cochrane reviews: cross sectional study. BMJ. 2013 Apr 23;346:f2231.

(3) Schroll J. Deaths in trials should always be reported. BMJ. 2013 Jul 4;347:f4219.

(4) Schroll J, Abdel-Sattar, M, Bero L. The Food and Drug Administration reports provided more data but were more difficult to use than the European Medicines Agency reports. J Clin Epidemiol. 2014 Aug 19.

(5) Schroll JB, Penninga E, Gøtzsche PC. Assessment of harms in protocols, clinical study reports and published papers of trials of orlistat. Submitted.

Acknowledgements

I would like to thank my main supervisor, Peter C. Gøtzsche, who has taught me a lot about scientific rigor as well as scientific writing and has given me the liberty to pursue interesting projects along the way. Special thanks to my good colleague Andreas Lundh, with whom I shared an office, and who has helped with many of the challenges life as a Ph.D. student has to offer. Thanks to all my colleagues at the Nordic Cochrane Centre. A special thanks to Karsten Juhl Jørgensen, Britta Tendal and Asbjørn Hróbjartsson for sparring with me. My American supervisor Lisa Bero was an inspiration with her good humour and hard work. I would also like to thank my co-authors Bianca Hemmingsen, Emma Maund, Maher Abdel-Sattar and Elisabeth Penninga. Finally thanks to my parents, my sister and wife and two wonderful boys who have inspired and supported me.

Financial support

Thanks to the Danish Council for Independent Research, The Nordic Cochrane Centre and the Cochrane Methods Innovation Fund for funding my research.

English summary

When doctors decide whether or not to prescribe a drug, they should balance the expected beneficial effects against the harmful ones. Several studies have shown that the beneficial effects reported in the scientific literature are generally overestimated, which makes the decision difficult. Whether reported harmful effects are similarly unreliable has not been investigated equally thoroughly.

The aim of this thesis was to study how harms are coded and subsequently reported in randomised clinical trials and to investigate whether unpublished data can improve the overview of harms. This was done in 5 projects reported in 5 papers.

When patients experience adverse events in an industry sponsored drug trial, they report this information which is then categorised according to a dictionary by the trial sponsor, a process called 'coding'. In our first project, we systematically reviewed papers about coding. It is an important process, as variation in coding may lead to dilution of the signals of harm, potentially overlooking important adverse events. Only one study had examined how coding differed between coders and it found important differences.

In our second project, we surveyed all Cochrane authors on how they sought unpublished information about harms and benefits. Although adverse events were often poorly reported in published papers, only 8% of those Cochrane authors who got unpublished data, got data on harms. Companies were generally not likely to provide data.

In the third project I sought unpublished data according to the principles we learned in the second project. We were conducting a Cochrane review on sulfonylurea drugs where we initially asked the manufacturer of repaglinide for data but we were told that no trials beside the published ones existed. However, only three trials were publicly available. As it seemed unlikely that a drug would get approved with so little data, I searched the FDA website and found that the manufacturer had conducted two additional trials. After several contacts to the manufacturer, they sent us the remaining trials along with internal reports of the three published trials. I found important discrepancies between published and unpublished data.

Regulatory agencies can be an important source of unpublished trial data, so we compared what was available at the US and the European drug agencies in a fourth project by accessing the data on their

websites. We found that the US drug agency provided more data, especially on harms, and that the European drug agency provided information about drugs that have been rejected or withdrawn. We therefore recommended that the websites of both agencies should be searched.

Clinical study reports are comprehensive documents that may run over thousands of pages for each trial. The manufacturer sends the reports to the drug regulators as part of their application for marketing approval. In our fifth project, we compared data from these reports with published data for the slimming pill orlistat. We found that much data had been left out in the published papers and that the authors had used quite unusual, unreported methods that reduced the possibility of identifying harms. The duration of each adverse event was not analysed even though there was a huge difference between placebo and orlistat.

Harms were poorly reported and the most reliable data came from clinical study reports. Despite the large amount of data we found that it was feasible to analyse them, which emphasises the need for these reports to become publicly available.

Danish summary

Når læger afgører om et lægemiddel skal ordineres, afvejer de forventede gavnlige virkninger mod forventede skadelige virkninger. Flere undersøgelser har vist, at de gavnlige virkninger rapporteret i den videnskabelige litteratur generelt er overvurderet, hvilket gør beslutningen vanskelig. Om de rapportede skadelige virkninger ligeledes er upålidelige, er ikke undersøgt lige så grundigt.

Formålet med denne afhandling var at undersøge, hvordan skadevirkninger blev behandlet og efterfølgende rapporteret i randomiserede kliniske studier, samt at undersøge om upublicerede data kan bidrage med gavnlig viden om skadevirkninger. Dette blev gjort i 5 projekter rapporteret i 5 artikler.

Når patienterne oplever skadevirkninger i et industrisponsoreret lægemiddelforsøg, rapporterer de denne information til forskeren, hvorefter firmaansatte kategoriserer hver skadevirkning efter en ordbog, en proces, der kaldes "kodning". I vores første projekt gennemgik vi systematisk alle artikler om kodning. Variation i kodning kan føre til udvanding af signalerne om skadevirkninger, hvilket potentielt kan føre til, at vigtige skadevirkninger overses. Kun ét studie havde undersøgt, hvordan kodning varierede mellem personer og der var vigtige forskelle.

I vores andet projekt adspurgte vi alle Cochrane-forfattere om, hvordan de søgte efter upublicerede data om skadevirkninger og gavnlige virkninger. Selvom skadevirkninger ofte blev dårligt rapporteret i offentliggjorte artikler, var det kun 8% af de Cochrane forfattere, der fik upubliserede data, der fik oplysninger om skadevirkninger. Lægemiddelvirksomheder leverede generelt ikke data.

I det tredje projekt brugte jeg de principper, vi lærte i det andet projekt til at lede efter upublicerede data. I forbindelse med udafbejdelse af af en Cochrane-gennemgang af diabetes lægemiddelgruppen sulfonylurea, bad vi producenten af repaglinid om data, men fik at vide, at ingen forsøg udover de publicerede eksisterede. Kun tre forsøg var imidlertid offentligt tilgængelige. Det virkede usandsynligt, at et lægemiddel ville blive godkendt med så få data, så jeg søgte på FDAs hjemmeside og fandt ud af, at fabrikanten havde foretaget yderligere to forsøg. Efter flere kontakter til producenten, sendte de os de resterende forsøg sammen med interne rapporter for de tre publicerede forsøg. Jeg fandt vigtige uoverensstemmelser mellem publicerede og upublicerede data.

Da lægemiddelmyndighederne kan være en vigtig kilde til upublicerede data, sammenlignede vi, hvad der var tilgængeligt på hjemmesiden hos den amerikanske og den europæiske lægemiddelstyrelse i et fjerde projekt. Vi fandt, at den amerikanske lægemiddelstyrelse havde flere data, særligt om skadevirkninger, mens den europæiske lægemiddelstyrelse havde information om lægemidler, der var blevet afvist eller trukket tilbage. Vi anbefaler derfor, at begge hjemmesider søges af forskere og beslutningstagere, der vil have et fuldt overblik.

Kliniske studierapporter er omfattende dokumenter, der kan være på over tusind sider for hvert forsøg, og som producenten sender til lægemiddelstyrelsen som led i deres ansøgning om markedsføringstilladelse. I vores femte projekt, sammenlignede vi data fra disse rapporter med offentliggjorte artikler for slankemidlet orlistat. Vi fandt, at store mængder data var blevet udeladt af de offentliggjorte artikler, samt at firmaet havde brugt usædvandlige metoder, der reducerede muligheden for at identificere skadevirkninger og som heller ikke blev beskrevet fyldestgørende. Varigheden af de enkelte skadevirkninger blev ikke analyseret på trods af en stor forskel mellem orlistat og placebo.

Skadevirkninger er dårligt rapporteret, og de mest pålidelige data kommer fra de kliniske studierapporter. På trods af den store datamængde er det meningsfuldt at analysere, og disse rapporter bør være offentligt tilgængelige.

Introduction and objectives

Harms

The decision to prescribe a drug is based on the balance between the drug's benefits and harms. Randomised clinical trials are the most reliable source for both effects, but they tend to focus on benefits and to underreport harms. Samples of randomised controlled trials have shown that sometimes adverse event are not mentioned at all, and more frequently, adverse events are poorly reported.(6) Details on how the patients were asked were not reported at all, which is problematic since it has a great impact on the number and characteristics of the collected adverse events.(7) In a study on hypertension, 16% of the patients reported adverse events spontaneously, 24% responded to a general enquiry, and 62% stated adverse events when presented with a specific questionnaire.(8)

"Harms" are the totality of negative consequences of an intervention.(9) The term "side effects" does not necessarily imply harm and will therefore not be used in this thesis. "Adverse drug reaction" implies that causality between the intervention and the event is established and since this is not always the case at all stages of a clinical trial, we will preferably use the term "adverse event".

An international consensus group has developed a guideline that defines what information regarding adverse events should be reported in journal papers.(9) The guideline recommends that it should be clarified how harms were collected and that each adverse event should be defined. A survey of trials showed that the guideline was only followed adequately in around 20% of the trials.(10) In a survey of harms reporting in 192 drug trials on 7 diverse topics, the severity of clinical adverse effects and laboratory-determined toxicity was adequately defined in only 39% and 29% of trial reports, respectively.(11) Only 46% of the trials gave specific reasons for discontinuation of study treatment due to toxicity.(11) In systematic reviews of harms, the search strategies were generally inadequate and not reported thoroughly,(12) and although it is standard to assess the risk of bias in trials included in reviews and state sources of funding, both were done in less than half of the meta-analyses of adverse events.(13)

Given the poor quality of reporting of harms it was no surprise that two recent studies found that only a fraction of adverse events were tallied in the published papers when compared with clinical

study reports (CSRs).(14,15) CSRs are detailed reports which usually take up hundreds or thousands of pages for each trial. They are written by drug companies and send to the drug regulatory agencies as part of their application for market approval. One of the reasons for the few published adverse events was that different filters were used in the publications. For instance, only adverse events occurring in more than 5% of the patients were tabulated. Another study found that less than half of all serious adverse events reported by the drug manufacturer in summaries on their website were reported in published papers.(16) There is also evidence that specific adverse events have been misclassified, for instance in the case of paroxetine where suicide attempts were classified as "accidental injury".(17)

Before harms are possibly reported in a published paper a complex process has taken place. A patient in a trial has a symptom and at the next visit (which might be months ahead), the patient may or may not describe the experience to the investigator, partly dependent on whether the investigator asks proper questions. The investigator translates the information into a biomedical entity and might filter some of it. If the investigator decides to make a note of the event, this information will later be transformed by a medical coder in the sponsoring company. Coders use a predefined list of possible adverse events organised in a hierarchy when they code the narrative description of an adverse event.(18) With the most commonly used system, each adverse event can be coded as several terms, which may lead to inconsistency and failure in identifying harms.(18) At the end of the trial, such data are categorised and summarised, and adverse events are lumped into broad categories for practical reasons. At each of these steps, decisions are made that might impact the overall impression of harms and lead to important harms being missed, e.g. "gastrointestinal events" may include cases of mild nausea as well as life-threatening bleeding ulcers.

Most drug trials are funded by the pharmaceutical industry and most data on harms are never published. Harms data is much more heterogeneous than data for benefit which makes it difficult to condense them and hence difficult to publish them comprehensibly in a journal article. Systematic reviews have traditionally relied on published papers so usually only few adverse events are pooled. Drug regulators, on the other hand, get access to a lot of the harms data when the companies submit their application for marketing approval. However, the drug regulators only publish short summaries.

Unpublished data

Most doctors get information about a drug's benefits and harms through peer reviewed medical journals or books, traditionally considered "published data". The papers in these journals are a condensed version of everything that has happened in a trial and a lot of additional data exist on file, in archives of companies and regulators. If the published papers were a fair representation of their unpublished counterpart, it would not be problematic that a big portion of data is hidden. However, the published papers are not a fair presentation since trials with positive outcomes are much more likely to be published.(19–21) The trials that are actually published are more likely to report outcomes that were positive than the ones that were not.(22) By considering only the published data we would overestimate the benefit of an intervention and it is possible that also the estimation of harm would be distorted.

In figure 1 each filing cabinet – and corresponding report represents – different trials that are published to a different extent. Trial A represents the common situation where only parts of summary data are published. If positive outcomes are more likely to be reported it is a case of selective outcome reporting. Often data on harms are less well reported. Trial B is the less common situation where the published paper is actually a fair representation of what happened in the trial. Trial C represents the few cases where drug regulators or companies have supplied clinical study reports to independent researchers. Both the summary report and the filing cabinet is in the "public domain". Notice that there are still unpublished data from these "best case scenarios" which could be case report forms or individual patient data. Trial D is the all too common situation where a study is not published at all. If the reason for lack of publishing is non-significant results, this would be a case of publication bias. Notice that the amount of data in the unpublished domain is huge compared to the public domain.

Trial registration is mandatory for many trials. The purpose of registration on a website, as for example ClinicalTrials.gov, is to try to prevent publication bias and selective outcome reporting. However, a study found that only 22% of the trials had reported results on the website a year after completion, which is the deadline determined by law.(23) Registration of trials has made it easier to document publication bias but it has not solved the problem. Simple reminders sent by independent researchers actually increased the reporting(24) so more encouragement could decrease the problem.

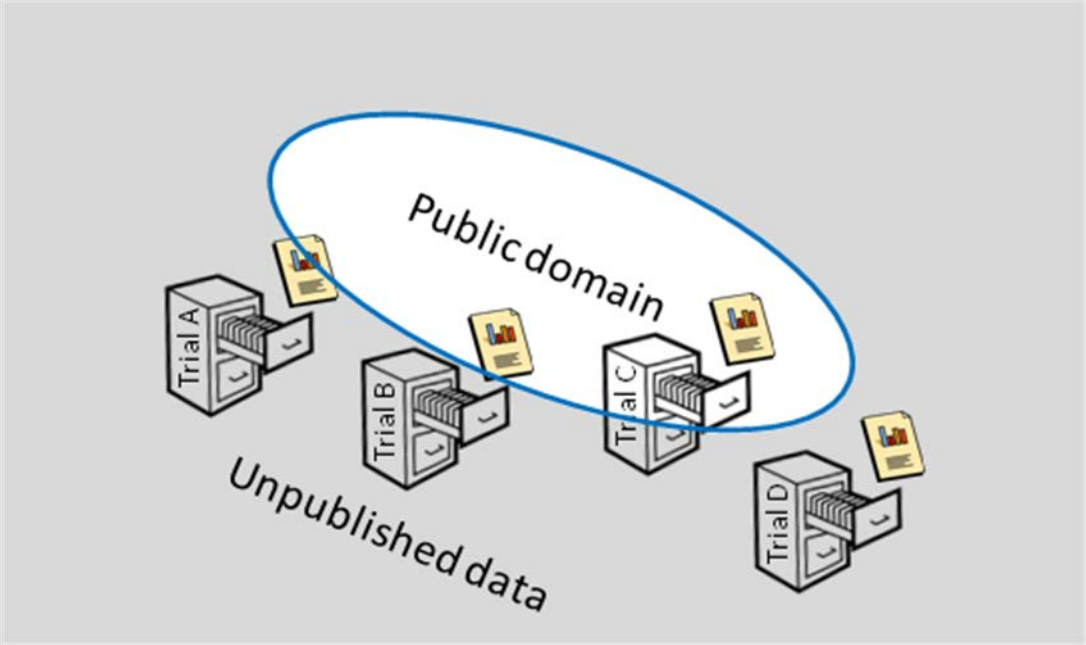


Figure 1. Examples of trial data in the public domain, for instance in medical journals. The filing cabinet represents the total amount of data produced in relation to a clinical trial (protocol, clinical study report and individual patient data, case report forms). The coloured report represents a comprehensive summary of the trial which is usually only a fraction of the total amount of pages in the filing cabinet.

By definition unpublished data are not easy to obtain and often involves a considerable amount of contacts to investigators or companies.(25,26) Regulatory agencies can be a more generous source. We have acquired access to the clinical study reports of all the placebo-controlled trials submitted to the European Medicines Agency (EMA) for marketing approval for the slimming pill, orlistat, and their corresponding trial protocols.(27) These documents include individual patient data on harms. Slimming pills are interesting because they are historically associated with many adverse events. Today all slimming pills except orlistat have been withdrawn from the European market. We used these unique data for an in-depth exploration of the problems related to the reporting of harms in drug trials.

Objectives

The objective of this thesis is to investigate if adverse events are handled in an unbiased manner and to determine if extracting unpublished harms data is feasible. The objective can be broken down into:

1) Study coding of adverse events in clinical trials and determine how coding effects the reporting of adverse events in the published paper.

2) Study how authors of Cochrane reviews search for and use unpublished data.

3) In relation to a Cochrane review develop a strategy for searching for unpublished data.

4) Compare the available amount of unpublished data at the European and US drug regulator.

5) Compare how adverse events are gathered, analysed and reported in protocols, clinical study reports and published papers of orlistat.

Description of the research project

Paper 1. Coding

For our first paper we conducted a systematic review of coding of adverse events. We were primarily interested in coding in clinical trials and studies of inter- or intra-observer variation. We developed a comprehensive search strategy by following the recommendations in the Cochrane Handbook(28) and searched MEDLINE, Embase and the Cochrane Central Register of Controlled Trials. Two independent researchers screened all titles identified by the search. Disagreement was resolved by discussion. Due to the expected heterogeneous nature of the papers no meta-analysis was planned but a narrative approach was chosen. The review was reported in accordance with the PRISMA guidelines.(29)

Paper 2. Searching unpublished data

We conducted a cross sectional study of all Cochrane contact authors. We asked them about their experiences with searching for unpublished data in relation to their review. All Cochrane contact authors received an invitation to an online questionnaire by e-mail. The questionnaire was designed to be as brief and clear as possible. In order to secure high participation rates up to 3 reminders were sent if the invitee did not respond. Before the contact authors were invited we pilot tested the questionnaire on 10 researchers, collected their comments and adjusted the questionnaire accordingly. We used the online questionnaire SurveyMoneky which gave us the possibility to send individual reminders and keep track of how many of our invitations had been answered. The invitation letter and the questionnaire was designed to achieve high participation rates which was

based on a systematic review.(30) The cross sectional review was reported in accordance with the STROBE guidelines.(31)

Paper 3. Deaths in trials

We conducted a Cochrane review(32) of sulfonylurea drugs for the treatment of diabetes mellitus type 2. The search, data extraction and analysis were in accordance with Cochrane methodology(33). We found only few trials that reported on patient relevant outcomes and therefore I decided to investigate if there were any unpublished data. The search process was built on the guidance of the Cochrane Handbook(28) as well as the experiences we gained by questioning the Cochrane authors about unpublished data.(2,25) This included searching clinicaltrials.gov, the company's website and the European and US drug agencies for reports involving each of the sulfonylurea drugs. In the reports we looked for additional trials beside the one we had identified through the literature search.

Paper 4. FDA and EMA reports

For the fourth paper we decided to investigate how much data was available at the regulators' websites. Both the European Medicines Agency (EMA) and the American Food and Drug Administration (FDA) give public access to summaries and reports created in the approval process, often considered "unpublished". The FDA website is known to be very difficult to navigate.(34) We identified all medicines approved between January 2011 and December 2012 and paired the available documents with the ones on the alternative agency's website. Orphan drugs and biological drugs were not included, as their approval process is quite different. Two researchers then extracted information about the included trials in the application for marketing approval. We extracted information that could potentially identify each trial, for instance trial ID, investigator names, dates and also information about benefits and harms. We planned to compare how often information was reported for each agency and compare the two agencies with each other. In the end we wanted to be able to give guidance to authors of systematic reviews on where they should look for unpublished data.

Paper 5. Harms in orlistat trials

Paper 5 is the main paper of this thesis and it investigates how adverse events were handled and reported in the case of orlistat. We had access to 7 study reports which consisted of 8,716 pages and

included 3,776 patients. From the protocols and clinical study reports we extracted all information about adverse events. Through PubMed searches we identified publications based on the 7 included trials and we compared all three sources (protocols, clinical study reports and publications) with each other. We focused on the reported harms as well as the methodology for handling harms. Previously clinical study reports have been considered confidential and our methodology was inspired by the few previous publications in this field.(14,35) Two researchers independently extracted information from the documents and disagreements were resolved by discussion. For one trial we did an exploratory analysis based on the individual adverse event data from the clinical study reports. Since the reports were in a "scanned format" we had to use text recognition software on more than a hundred pages of tables to be able to analyse the data. It was then exported to an Excel spreadsheet and comparisons between the analyses in the summaries of the clinical study reports were compared to our new dataset.

Summary of the results

Paper 1. Coding

Our search returned 520 hits and 12 articles were included in our review.(1) Only one study reported something that could be considered an interobserver comparison. Two coders had coded investigators' verbatim descriptions of adverse events according to the medical dictionary MedDRA and in 12% of the cases the two coders disagreed.(36) All codes were then evaluated by medical professionals and they found that 8% of the codes were not medically accurate. Other included studies compared how well different dictionaries compared to product labelling and found some discrepancies. Other studies raised concerns about the dictionary MedDRA which is today considered standard. Because of MedDRAs ever increasing number of entries adverse events might be split in a way that is not biologically sound and which might hinder detection of harms due to decreased statistical power. Advanced lumping techniques have been developed, especially for pharmacovigilance, but are rarely used in clinical trials.

In the orlistat trials only "treatment emergent adverse events" (TEAEs) were reported. TEAE is defined as a new condition or worsening of an existing condition after initiation of the intervention which at first glance seems to make sense. However, worsening or even whether something is "new" is a matter of definition. Many drug trials will have a run-in period from a few weeks up to

half a year where "baseline" adverse events are gathered. If a patient has a recurrence of an event from the run-in period it will be ignored as it is considered not to be related to treatment. Because of the hierarchical structure of the medical dictionaries comparisons of two events can be done on many levels. For instance, is "tension headache" the same as "facial pain"? A study on a fictional data set found more than 26 ways of defining TEAEs which returned between 2 and 7 adverse events depending on the chosen definition.(37)

After our study was published, coding of suicidality in clinical study reports of duloxetine has been investigated by our research group.(38) They found that summaries in clinical study reports might not represent the underlying patient level data because of limitations in medical dictionaries. Overall this important step in adverse event handling is poorly investigated. Lack of blinding and poor reliability might bias the summary of adverse events.

Paper 2. Searching for unpublished data

We invited 5915 contact authors of Cochrane reviews and protocols to participate in our survey.(2) We received 2184 replies (37%) of which 1889 were complete. We were not able to track if all e-mail addresses were active so the response rate of authors actually receiving the e-mail might have been a little higher. Of the respondents only 24% stated that they did not search for unpublished data for their review, most frequently because they did not expect reply. Among the authors that searched for unpublished data 44% never received data, most commonly because they never received a reply. In 74% of the cases obtained data came from trialists and in only 6% did the data come from the manufacturer. In 75% of the cases 1-3 contacts were enough and in 53% of the cases the data was received within a month of the request. However, 41% found that the time consumption was the most challenging part of searching for unpublished data. 21% found that poor readability and organisation of the data was the greatest challenge. Through interviews the respondents elaborated on the challenges of using unpublished data.(39)

Authors most frequently received summary data (51%) but also individual patient data were commonly received (21%). Surprisingly data on harms was rarely obtained (8%). Many reviews are focused on benefits, and harms data is usually heterogeneous and might therefore be difficult to request in a simple manner.

A third of the respondents got information about previously completely unpublished trials. The authors that got data from manufactures on average used more contacts, waited longer for the data and were less frequently supplied with individual patient data than authors getting data from non-manufacturers.

We suspect that the non-responders to our survey were less likely to search for unpublished data. Therefore the actual percentage of authors that searched for unpublished data might be a bit lower. Unfortunately up to half of the researchers never received data. It seems that manufacturers are especially difficult to approach. They rarely provide data even though they conduct most of drug trials. Many contacts were involved in contacting manufactures and the low frequency of obtaining individual patient data might deter authors. It might not be worth the time to contact manufacturers especially if clinical study reports can be requested from drug regulators. Another concern is that drug companies might only supply data in the situations where they supported their drug. Drug regulators were also a very seldom source of data, even though the frequency was rising.

It was encouraging to see that many actually got access to individual patient data, the highest detail of data, which is often the most protected. On the other hand it was concerning to discover that information about harms were rarely received. Harms are seldom the primary focus of reviews but harms have to be treated equally to benefits throughout all stages of review development. Another aspect could be that harms data is often heterogeneous and difficult to manage in a conventional meta-analysis with a fixed number of outcomes. Current recommendation in the Cochrane Handbook is to select a few important harms outcomes.(40) Even though this is a pragmatic approach it disregards most of the data with the potential consequence of overlooking important harms. Traditional meta-analyses are not built for hierarchical ordered datasets but future research should look into how pooling of harms can be handled.

Our study results have limited generalisability due to the low response rate.

Paper 3. Deaths in trials

As we were reaching the end of our work on the Cochrane review where we had screened more than 7000 records(32), I realised that we had almost no data on patient relevant outcomes such as mortality and cardiovascular morbidity. This was surprising because the included drug class, sulphonylurea drugs used for the treatment of diabetes mellitus type 2, had been approved in recent

years. In the data extraction phase we had assumed no deaths had occurred if there was no information on deaths in the published paper. During the early phases of the review the main author had contacted the companies that developed drugs in this class but had received the answer that there were no unpublished data.

Unconvinced I searched clinicaltrials.gov but I found no relevant studies because the drugs were approved prior to mandatory registration. I then turned to the US and European drug regulators for summaries of the drugs. The main sulphonylurea drugs were approved too long time ago to be available on the agencies website. However, I managed to find a summary of repaglinide on the FDA website that described five 1-year trials. This drug was also included in our review but we had only found 3 published papers. We contacted Novo Nordisk again, the manufacturer of the drug, but were now told that they did not share data that were not in public domain. Among the group of authors we discussed the situation and tried to contact Novo Nordisk again. They had now changed their mind and gave us 5 short summary reports. Two of the trials had not previously been published. What was worse was that the summary reports revealed that patients had died during two of the already published trials. These deaths were not mentioned in the published articles and we had falsely assumed that no one had died.

Two trials (one unpublished) had a higher rate of cardiovascular events in the repaglinide group – a known risk of old sulphonylurea drugs – but still concluded that the drug was safe. Fewer hypoglycaemic events were reported in the published paper than in the internal reports.

Selective reporting of outcomes have been reported previously(22,41) and reporting of harms have been shown to be inadequate. However, omission of deaths from published papers makes it hard to trust anything, as trialists and companies have an ethical obligation to report important harms.

If authors of systematic reviews assumed 0 deaths when none were reported systematic reviews might underestimate the risk of death by many drugs.

The lack of accordance between the results and the conclusion in one of the published reports is unfortunately not a new phenomenon. In a sample of all randomised controlled trials with non-significant results from December 2006, researchers found that around half of all abstracts had "spin" in their conclusion.(42) "Spin" was defined as strategies to highlight benefit despite the non-significant result.

Paper 4. FDA and EMA reports

We identified 27 pairs of summary reports from the FDA and EMA that had been approved between January 2011 and December 2012 at one of the agencies.(4) The most common drug classes were antineoplastic drugs (n=6) and anti-infective drugs (n=5). The documents that were available at the FDA consisted of more pages on average than the document from EMA (219 vs. 88). Most documents from both FDA and EMA contained a table of contents but in the FDA's documents the page numbers did not match, which made the document difficult to navigate. Most documents were searchable. Reports from both agencies contained enough information about pivotal trials to include those in a meta-analysis. The FDA documents were redacted primarily to protect proprietary interests whereas the reason and amount of redaction in in the EMA documents were unclear. None of the reports reported clinicaltrials.gov ID's or comprehensive investigator names which would have made identification of publications easier. The FDA provided more data on harms than the EMA (all important harms reported in 93% vs. 26% of the reports). For instance comprehensive tables of all adverse events that occurred in the trials and tables of serious adverse events. The EMA provided summaries for withdrawn and rejected drugs as well as a reason for this particular decision whereas the FDA did not provide this. We therefore recommend that both agencies' websites should be searched if one wants a comprehensive picture of a drug's benefits and harms. Data from FDA has previously been shown to alter meta-analyses based only on published data.(43) Sufficient data for conducting meta-analyses was readily available from both agencies. Harms data is more comprehensible but their documents are harder to search and do not provide data on withdrawn or rejected drugs. FDA even redacted information about indications in the summaries that were not granted. The FDA documents gave more insight in the processing of applications as letters from the agency to the applicant were available. They could be regarding a concern of harm or a request for additional analyses. Guidance on how to access the FDA website has been developed.(34)

Paper 5. Harms in orlistat trials

In the last paper we examined the 7 orlistat protocols and compared them to the CSRs and published papers. The 7 trials were double-blinded placebo-controlled trials that lasted between 52 and 104 weeks. They were conducted in the USA and Europe between 1992 and 1996 and each

treatment arm included between 114 and 244 patients. The patients' body mass index was between 28 and 43 and each patient had between 9 and 17 visits with research staff during the study period. Most trials started with a 4 week run-in period where all patients received placebo and all adverse events were ignored. If an event reoccurred after the run-in period it would only be considered an adverse event if the severity had increased, a so called TEAE. However, it was not specified when events were considered a re-occurrence and as we discussed in paper 1 this can have great impact on the perceived harms.

For protocols 1-3 an appendix specified that investigators were discouraged to use "diarrhoea" as an adverse event. Instead, investigators were urged to use a list of prespecified adverse events including "increased defaecation", "liquid stools", "soft stools", "fatty/oily evacuations" and "oily spotting". The rationale for these coding recommendations was that the sponsor did not consider diarrhoea well defined but found that using the term could lead to "misunderstandings". The remaining protocols did not contain this appendix but we suspect that it was still used as no adverse event was coded as "diarrhoea".

At each visit the investigator had to check a box if the patients had "adverse experiences" but it was not specified how the investigator should ask the patient about adverse events and the planned analysis was vaguely described as "descriptive statistics". The primary outcomes for "quality of life" was also vaguely described as the composition of the subscale was not clear.

When we compared the protocols with the CSRs we discovered that the CSRs had added a star to several gastrointestinal adverse events in the appendix that gave guidance on how to code gastrointestinal adverse events. The unstarred events should – very unconventional – only be considered adverse events if they were "bothersome". The events were "fatty/oily stool", "liquid stools", "increased defaecation", "stools soft", "decreased defaecation" and "pellets". The approach seems especially problematic for "liquid stools" as this was recommended as an alternative to "diarrhoea" in the protocol. It was not explained why this star was added all of a sudden. It is very problematic to change the protocol without explanation in a direction that could potentially favour the drug by ignoring the patients' complaints.

The methods section of the CSRs specified which coding dictionary would be used and gave more details on how adverse events would be presented. The primary outcomes for quality of life had

changed to the subscales "overweight distress", "depression" and "satisfaction with treatment". We could not find any explanation as to why the outcomes had been changed.

In the results section of the CSRs we noticed that narrative summaries tended to downplay the differences. For instance, most summaries emphasised that most adverse events were mild or moderate in intensity without highlighting that there were considerably more events in the orlistat group. It was also emphasised that most patients had only 1 to 2 gastrointestinal adverse events without comparing the duration of the events.

The CSRs contained 71 to 270 times more pages than their corresponding publications. Even though the medical dictionary used was mentioned in 5 publications none of the publications specified that only "bothersome" events were coded for several gastrointestinal events. The reported adverse events were limited by severe restrictions that were not planned in the protocol or in methods of the CSRs. For instance, events that were considered "unrelated" were censored in 3 of the 7 publications. In 4 publications we were able to compare the number of events reported in the papers with the number of events in the CSRs and only between 3% and 33% of the events were reported. One paper lumped several adverse events in a new category that had not been planned in the protocol or CSR.

In our exploratory analysis we imported 3446 adverse events from tables and found that patients in the orlistat group experienced 2008 versus 1438 in the placebo group. This was in contrast to the summaries in the CSR that reported 1198 (60% of the events in the tables) events for the orlistat group and 930 from the placebo group. The reason for the lower numbers in the CSR was that if an event – for instance headache – occurred multiple times in one patient – for instance on day 100, 200 and 300 – it would only be counted once. Neither the protocol nor the CSR explained this practice.

Duration of adverse events was recorded in the case report form but never analysed. We compared the average duration of adverse events in the two groups and found an average duration in the orlistat group of 22.5 days (95% CI: 20.0-25.1) compared to 14.7 days (12.9-16.6) in the placebo group. When this was multiplied with the average number of adverse events in each arm we found that a patient in the orlistat group on average experienced 288 days of adverse events as compared to only 141 days in the placebo group. This difference is in sharp contrast to the statement in the CSRs summaries that said: "… adverse event profile differences between orlistat- and placebo-

treated patients were small or non-existent, except for adverse events of the gastrointestinal system".

Even though it is not obligatory to present an analysis of the duration of adverse events (44) it is hard to believe that the sponsor has not made the analysis given that the data was recorded and the protocol was so vague as to which analyses should be conducted. Description of adverse events took up very few pages in the protocol but ended up taking most pages in the clinical study report.

In our exploratory analysis we also found that many patients in the placebo group were withdrawn due to hyperglycaemia. It was specified in the protocol that patients should be withdrawn if HbA1c was above a specific threshold but it was not specified that this should be considered an adverse event, which it was.

Protocols should meticulously describe how harms are intended to be recorded and analysed. What applies to benefit should of course also apply to harm. In 2013 an attempt was made to standardise protocols by researchers from academic institutions and time will tell if the companies adapt these recommendations.(45)

Clinical study reports have only been investigated in relatively few studies (14,15,35) because access to these reports is relatively new.(27) They are considered important sources of information(46) and the EMA handed out more than 300 documents and over 1.6 million pages during the first two years of their new openness policy (47) which has later been paused because two pharmaceutical companies sued the agency. It is a bit surprising that this has not led to more research results but an explanation could be the huge amount of data that has been a tough nut to crack for small research groups. More research into how to handle large amounts of data is needed. At the moment it is uncertain to what extent CSRs will be publically available.

Even though we have highlighted many limitations in the way adverse events are recorded and analysed we haven't even considered the limitations of the randomised controlled trial in itself. Clinical trials do not always represent realistic situations. Often the sickest patients are excluded and patients have much more frequent blood tests and physician visits than would be feasible in a real life situation. Therefore the harms recorded in this artificial situation is probably and underestimation of the harms that patients in real life situations would experience.

Our study was explorative in nature and our focus was to find problems in the handling of adverse events. The problems we have encountered could have been counteracted by excellent initiatives that we could have missed.

Conclusion

Adverse events are coded into a dictionary for practical reasons but the high diversity of entities might decrease the chance of detecting adverse events that represent the same biological phenomenon. There are many different definitions on when an adverse event is "new" which makes the process susceptible to bias and the most common dictionary in not properly validated.

Even though unpublished data is often obtained when researchers conduct systematic reviews and ask for them we found that harms are infrequently gathered. We encourage authors of systematic reviews to inquire about harms data.

In the case of repaglinide we found many inconsistencies between internal summary reports and published papers and we were able to establish that even deaths can go unpublished.

Both the EMA and the FDA website provided sufficient data so that the described trials could be included in a meta-analysis. The FDA had more information about harms whereas the EMA provided more information about withdrawn and rejected drugs. We encourage researchers to search both websites as they complement each other but they cannot be a substitute for individual patient data or CSRs.

The analysis plans for orlistat were vague which makes reporting of harms prone to post hoc decisions and bias. This was confirmed when we compared the protocols to the publications where some adverse event were only considered if they were "bothersome". Methods for handling harms were very poorly described in publications.

We found that duration for all adverse events were recorded but not analysed. In our analysis adding this parameter meant that an average patient receiving orlistat would experience double as many days with adverse events. This was in contrast to the CSRs that concluded that besides gastrointestinal adverse events there was no difference in the adverse event profile. Currently duration is not mandatory to analyse, but we have shown that it perhaps should be, as omitting it may bias the perception of harms.

Finally we have shown that it is feasible to analyse data from clinical study reports and we were able to digitalise the scanned document and perform analyses. Research into when these extended analyses are necessary could help prioritise the resources when performing systematic reviews.

Word list

Adverse drug reaction	Harmful effect where causality to drug is confirmed.
Adverse events	Harmful effect but where causality to drug is not confirmed.
Case report form	A form that an investigator of a trial uses to record all information of a given patient. Can take of several hundred pages. Includes individual patient data on benefit and harm. Are usually stored by the drug company and sent to the drug regulatory on request.
Clinical study reports (CSR)	Comprehensive reports of up to several thousand pages per clinical trial written by the drug company and sent to the regulatory drug agency as part of the application for marketing approval. They usually do not include individual patient information on benefit but might on harms.
Coding	A process where a narrative description of an adverse event is categorised according to medical dictionary which is organised in a hierarchy.
EMA	European Medicines Agency.
FDA	U.S. Food and Drug Administration.
Harms	The totality of harmful effects.
Run-in	An early phase of a clinical trial (before the intervention) where all the patients receive placebo.
Side effect	Unintended effect of drug. Can be harmful or beneficial.
Summary reports	Reports conducted by the FDA, EMA or other regulatory agencies as part of the processing of application for marketing approval. The EMA calls the reports EPARs and FDA calls them Drug Approval Packages. They will sometimes contain unpublished data but usually on an aggregated form.
Treatment emergent adverse events (TEAE)	A new condition or worsening of an existing condition after initiation of the intervention. In some trials only TEAEs are reported.
Unpublished data	All data that are not published in journals or books.

References

1. Schroll JB, Maund E, Gøtzsche PC. Challenges in coding adverse events in clinical trials: a systematic review. PloS One. 2012;7(7):e41174.

2. Schroll JB, Bero L, Gøtzsche PC. Searching for unpublished data for Cochrane reviews: cross sectional study. BMJ. 2013;346:f2231.

3. Schroll J. Deaths in trials should always be reported. BMJ. 2013 Jul 4;347(jul04 1):f4219–f4219.

4. Schroll JB, Abdel-Sattar M, Bero L. The Food and Drug Administration reports provided more data but were more difficult to use than the European Medicines Agency reports. J Clin Epidemiol. 2014 Aug 19;

5. Schroll JB, Penninga E, Gøtzsche PC. Assessment of harms in protocols, clinical study reports and published papers of trials of orlistat. Submitted.

6. Loke YK, Derry S. Reporting of adverse drug reactions in randomised controlled trials - a systematic survey. BMC Clin Pharmacol. 2001;1:3.

7. Wallin J, Sjövall J. Detection of adverse drug reactions in a clinical trial using two types of questioning. Clin Ther. 1981;3(6):450–2.

8. Olsen H, Klemetsrud T, Stokke HP, Tretli S, Westheim A. Adverse drug reactions in current antihypertensive therapy: a general practice survey of 2586 patients in Norway. Blood Press. 1999;8(2):94–101.

9. Ioannidis JPA, Evans SJW, Gøtzsche PC, O'Neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. Ann Intern Med. 2004 Nov 16;141(10):781–8.

10. De Vries TW, van Roon EN. Low quality of reporting adverse drug reactions in paediatric randomised controlled trials. Arch Dis Child. 2010 Dec;95(12):1023–6.

11. Ioannidis JP, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. JAMA J Am Med Assoc. 2001 Jan 24;285(4):437–43.

12. Golder S, Loke Y, McIntosh HM. Poor reporting and inadequate searches were apparent in systematic reviews of adverse effects. J Clin Epidemiol. 2008 May;61(5):440–8.

13. Golder S, Loke Y, McIntosh HM. Room for improvement? A survey of the methods used in systematic reviews of adverse effects. BMC Med Res Methodol. 2006;6:3.

14. Maund E, Tendal B, Hrobjartsson A, Jorgensen KJ, Lundh A, Schroll J, et al. Benefits and harms in clinical trials of duloxetine for treatment of major depressive disorder: comparison of clinical study reports, trial registries, and publications. BMJ. 2014 Jun 4;348(jun04 2):g3510–g3510.

15. Wieseler B, Wolfram N, McGauran N, Kerekes MF, VervÃ¶lgyi V, Kohlepp P, et al. Completeness of Reporting of Patient-Relevant Clinical Trial Outcomes: Comparison of Unpublished Clinical Study Reports with Publicly Available Data. Ghersi D, editor. PLoS Med. 2013 Oct 8;10(10):e1001526.

16. Hughes S, Cohen D, Jaggi R. Differences in reporting serious adverse events in industry sponsored clinical trial registries and journal articles on antidepressant and antipsychotic drugs: a cross-sectional study. BMJ Open. 2014 Jul 1;4(7):e005535.

17. Lenzer J. Secret US report surfaces on antidepressants in children. BMJ. 2004 Aug 7;329(7461):307.

18. Brown EG. Using MedDRA. Drug Saf. 2004;27(8):591–602.

19. Scherer RW, Langenberg P, von Elm E. Full publication of results initially presented in abstracts. Cochrane Database Syst Rev. 2007;(2):MR000005.

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

21. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. N Engl J Med. 2008 Jan 17;358(3):252–60.

22. Chan A-W, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA J Am Med Assoc. 2004 May 26;291(20):2457–65.

23. Prayle AP, Hurley MN, Smyth AR. Compliance with mandatory reporting of clinical trial results on ClinicalTrials.gov: cross sectional study. BMJ. 2012 Jan 3;344(jan03 1):d7373–d7373.

24. Maruani A, Boutron I, Baron G, Ravaud P. Impact of sending email reminders of the legal requirement for posting results on ClinicalTrials.gov: cohort embedded pragmatic randomized controlled trial. BMJ. 2014 Sep 19;349(sep19 2):g5579–g5579.

25. Wolfe N, Gøtzsche PC, Bero L. Strategies for obtaining unpublished drug trial data: a qualitative interview study. Syst Rev. 2013;2:31.

26. Young T, Hopewell S. Methods for obtaining unpublished data. Cochrane Database Syst Rev. 2011;(11):MR000027.

27. Gøtzsche PC, Jørgensen AW. Opening up data at the European Medicines Agency. BMJ. 2011;342:d2686.

28. Lefebvre C, Manheimer E, Glanville J. Chapter 6: Searching for studies. In: Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions [Internet]. Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011.; Available from: www.cochrane-handbook.org

29. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ. 2009;339:b2700.

30. Edwards PJ, Roberts I, Clarke MJ, DiGuiseppi C, Wentz R, Kwan I, et al. Methods to increase response to postal and electronic questionnaires. Cochrane Database of Systematic Reviews [Internet]. John Wiley & Sons, Ltd; 1996 [cited 2014 Jul 21]. Available from: http://onlinelibrary.wiley.com/doi/10.1002/14651858.MR000008.pub4/abstract

31. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. The Lancet. 2007 Oct 26;370(9596):1453–7.

32. Hemmingsen B, Schroll JB, Lund SS, Wetterslev J, Gluud C, Vaag A, et al. Sulphonylurea monotherapy for patients with type 2 diabetes mellitus. Cochrane Database Syst Rev. 2013;4:CD009008.

33. Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions [Internet]. The Cochrane Collaboration. 2011. Available from: www.cochrane-handbook.org

34. Turner EH. How to access and process FDA drug approval packages for use in research. BMJ. 2013;347:f5992.

35. Jefferson T, Jones MA, Doshi P, Del Mar CB, Hama R, Thompson MJ, et al. Neuraminidase inhibitors for preventing and treating influenza in healthy adults and children. Cochrane Database Syst Rev. 2014;4:CD008965.

36. Tonéatti C, Saïdi Y, Meiffrédy V, Tangre P, Harel M, Eliette V, et al. Experience using MedDRA for global events coding in HIV clinical trials. Contemp Clin Trials. 2006 Feb;27(1):13–22.

37. Nilsson ME, Koke SC. Defining Treatment-Emergent Adverse Events with the Medical Dictionary for Regulatory Activities (MedDRA)*. Drug Inf J. 2001;35(4):1289–99.

38. Maund E, Tendal B, Hrobjartsson A, Lundh A, Gotzsche PC. Coding of adverse events of suicidality in clinical study reports of duloxetine for the treatment of major depressive disorder: descriptive study. BMJ. 2014 Jun 4;348(jun04 2):g3555–g3555.

39. Wolfe N, Gøtzsche PC, Bero L. Strategies for obtaining unpublished drug trial data: a qualitative interview study. Syst Rev. 2013;2:31.

40. Loke YK, Price D, Herxheimer A. Chapter 14: Adverse effects. In: Higgins J, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions [Internet]. Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011.; Available from: www.cochrane-handbook.org

41. Rising K, Bacchetti P, Bero L. Reporting bias in drug trials submitted to the Food and Drug Administration: review of publication and presentation. PLoS Med. 2008 Nov 25;5(11):e217; discussion e217.

42. Boutron I, Dutton S, Ravaud P, Altman DG. REporting and interpretation of randomized controlled trials with statistically nonsignificant results for primary outcomes. JAMA. 2010 May 26;303(20):2058–64.

43. Hart B, Lundh A, Bero L. Effect of reporting bias on meta-analyses of drug trials: reanalysis of meta-analyses. BMJ. 2012;344:d7202.

44. Structure and content of clinical study reports. E3. [Internet]. International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use; 1995 [cited 2014 Jun 26]. Available from: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E3/E3_Guideline.pdf

45. Chan A-W, Tetzlaff JM, Gøtzsche PC, Altman DG, Mann H, Berlin JA, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. BMJ. 2013;346:e7586.

46. Doshi P, Jefferson T. Clinical study reports of randomised controlled trials: an exploratory review of previously confidential industry reports. BMJ Open. 2013 Jan 1;3(2):e002496.

47. Doshi P, Jefferson T. The first 2 years of the European Medicines Agency's policy on access to documents: secret no longer. JAMA Intern Med. 2013 Mar 11;173(5):380–2.

OPEN ACCESS Freely available online

PLoS one

Challenges in Coding Adverse Events in Clinical Trials: A Systematic Review

Jeppe Bennekou Schroll*, Emma Maund, Peter C. Gøtzsche

Nordic Cochrane Centre, Rigshospitalet, Copenhagen, Denmark

Abstract

Background: Misclassification of adverse events in clinical trials can sometimes have serious consequences. Therefore, each of the many steps involved, from a patient's adverse experience to presentation in tables in publications, should be as standardised as possible, minimising the scope for interpretation. Adverse events are categorised by a predefined dictionary, e.g. MedDRA, which is updated biannually with many new categories. The objective of this paper is to study interobserver variation and other challenges of coding.

Methods: Systematic review using PRISMA. We searched PubMed, EMBASE and The Cochrane Library. All studies were screened for eligibility by two authors.

Results: Our search returned 520 unique studies of which 12 were included. Only one study investigated interobserver variation. It reported that 12% of the codes were evaluated differently by two coders. Independent physicians found that 8% of all the codes deviated from the original description. Other studies found that product summaries could be greatly affected by the choice of dictionary. With the introduction of MedDRA, it seems to have become harder to identify adverse events statistically because each code is divided in subgroups. To account for this, lumping techniques have been developed but are rarely used, and guidance on when to use them is vague. An additional challenge is that adverse events are censored if they already occurred in the run-in period of a trial. As there are more than 26 ways of determining whether an event has already occurred, this can lead to bias, particularly because data analysis is rarely performed blindly.

Conclusion: There is a lack of evidence that coding of adverse events is a reliable, unbiased and reproducible process. The increase in categories has made detecting adverse events harder, potentially compromising safety. It is crucial that readers of medical publications are aware of these challenges. Comprehensive interobserver studies are needed.

Citation: Schroll JB, Maund E, Gøtzsche PC (2012) Challenges in Coding Adverse Events in Clinical Trials: A Systematic Review. PLoS ONE 7(7): e41174. doi:10.1371/journal.pone.0041174

Editor: Imti Choonara, Nottingham University, United Kingdom

Received April 13, 2012; **Accepted** June 18, 2012; **Published** July 20, 2012

Funding: These authors have no support or funding to report.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: js@cochrane.dk

Introduction

The decision to prescribe a drug is based on the balance between the drug's benefits and harms. All drugs have unwanted effects and reliable information about these effects is important. Throughout a clinical trial, adverse events, including harms of the drug, are monitored and recorded for the purposes of patient safety, regulatory requirements, and developing a safety profile of the drug. The process of condensing thousands of pages of data on adverse events from clinical trials to tables in regulatory submissions and summaries in papers and product labeling is complex and involves many assumptions and choices. Readers of medical journals need to be aware of these issues in order to appraise published study reports critically.

Before harms are reported (or not reported) in a published paper, many decisions have been made. A patient in a trial may experience 'something'. In some studies, patients can contact investigators by phone; in other studies, the symptoms may not be recorded before the next visit (which might be weeks ahead). The patient may or may not describe the experience to the investigator, partly dependent on the method of elicitation used by the investigator (e.g. open ended questions, symptom checklists). Information about adverse events can also be gathered from medical records and laboratory values. The investigator interprets the information in a biomedical framework and might filter some of it, especially if he believes the event is not drug related [1]. If the investigator decides to record the event, he will do so in the patient's case report form (CRF). This information will later be transformed by a medical coder employed by the trial sponsor. Coders use a medical dictionary, which is a predefined list of possible adverse events organized in a hierarchy, to code the narrative description of an adverse event [2].

Pharmaceutical companies have historically used many different dictionaries, such as WHO's Adverse Reaction Terminology (WHO-ART), the Thesaurus of Adverse Reaction Terms (COSTART), or the International Classification of Diseases (ICD 9 and ICD 10), to categorize adverse events, frequently customizing a dictionary for a specific trial. In 1994, the pharmaceutical industry, together with regulatory agencies, developed a standard dictionary named the Medical Dictionary for Regulatory Activities (MedDRA). Initially, the purpose was to

allow standardized electronic submissions [3]. MedDRA is a five level hierarchy with Lowest Level Terms at the bottom, followed by Preferred Terms, and with System Organ Class (SOC) at the top (Figure 1). Events are initially coded with Lowest level terms and they consist of thousands of synonyms and alternative spelling of Preferred Terms. Preferred Terms are unique medical entities. Companies are not allowed to add new terms but can suggest new terms – or alternate placing in the hierarchy – which will then be considered for the biannual update. To ensure an adverse event is only counted once in the standard summary tables, each Preferred Term can have only one primary SOC but several secondary ones to aid data retrieval [3]. It is mandatory for pharmaceutical companies to use MedDRA when applying for approval in the EU and Japan. In the US it is the terminology of choice [4].

With MedDRA, each adverse event can be coded as several different terms, 'insomnia' could for instance be coded as 11 different preferred terms [2]. This may lead to inconsistency and failure in identifying harms [2]. At the end of the trial, data are categorized and summarized, and adverse events are lumped into broad categories for practical reasons. At each of these steps, decisions are made that might impact the overall impression of harms and might lead to important harms being missed, e.g. "gastrointestinal events" may include cases of nausea as well as bleeding ulcers.

Mislabeling of adverse events can skew the interpretation of a drug's harms. The antidepressant paroxetine was tested in adolescents in an infamous trial that initially declared that the drug was "generally well tolerated" [5]. The paroxetine group, however, had an overrepresentation of "emotional lability". After scrutiny by the FDA and independent experts, it turned out that this term was only used when patients had "suicidal tendencies". Other cases of suicidal tendencies had been coded as aggression or "exacerbation of depression" [6].

With paroxetine, the miscoding appeared to be deliberately misleading, but it illustrates some fundamental problems with coding. Small deviations from the ideal of objective coding can lead to significantly changed conclusions and are usually impossible for the reader to discover. Development of new drugs that make a difference to old drugs in terms of benefit is increasingly difficult, and many new drugs are therefore being marketed as having less harms than their predecessor. Hence, readers of the scientific literature should be particularly focused on harms, and whether they have been reported reliably.

Our objective was to conduct a systematic review of studies on intra- and interobserver variation and other potential problems related to interpretation and translation of adverse events (as reported by clinicians) into coding terms for use in clinical study reports (for regulatory approval) and in publications (for marketing).

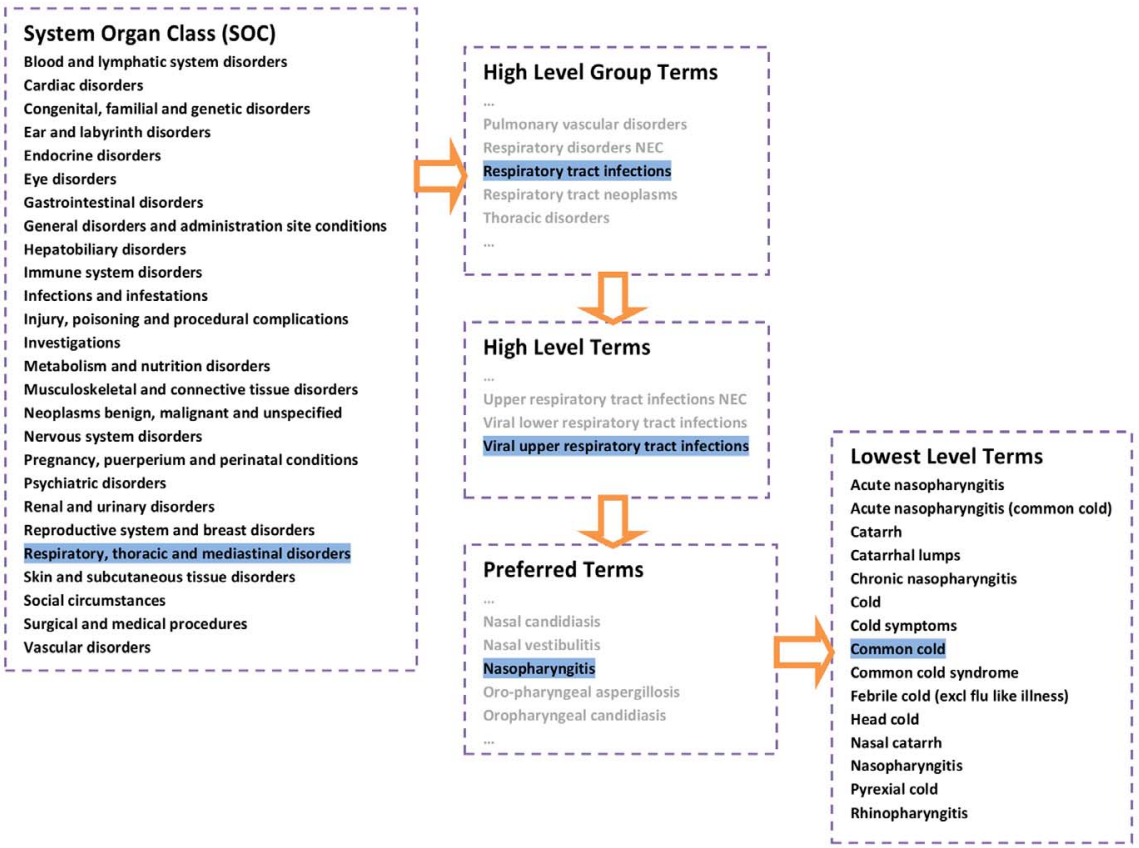


Figure 1. The MedDRA 5-level hierarchy demonstrated by using 'common cold' as an example.
doi:10.1371/journal.pone.0041174.g001

Methods

We searched PubMed, The Cochrane Library (CENTRAL and Methods) and EMBASE on the 28th of October 2011 and updated the search the 9th of March 2012. The search string was a combination of synonyms of adverse events and interobserver studies (see details in Text S1). Search terms also included the names of common dictionaries used for medical coding. We had no language or other restrictions for the searches. We also went through the reference lists of the included studies, visited medical dictionary websites and contacted principal authors for information about additional studies. Our protocol is available on request.

All abstracts and titles were screened for inclusion by two independent observers (EM, JBS) . Any differences were resolved by discussion. When eligibility could not be determined based on title and abstract alone, the full text article was retrieved. Eligible studies were interobserver studies of coding in clinical trials. Other studies addressing challenges in coding of adverse events in clinical trials were also included. Review articles were excluded.

We adhered to the PRISMA guidelines for reporting systematic reviews [7], see Checklist S1 for details. Because of expected heterogeneity in the results, our review was planned to be qualitative.

Results

Our search returned 520 unique citations. We retrieved the full text for 61 articles and included 9 of these. The papers we excluded were reviews (n = 13), papers with no data (n = 9), not describing adverse events (n = 7) and papers not referring to clinical trials (n = 7) or otherwise not relevant (n = 16). See figure 2 for details. Three additional papers were included from the references of located papers. Only one of the included papers was an interobserver study of coding. All included papers are described in table 1.

The only interobserver study of coding was done by Toneatti et al. who performed a pilot project where two experienced coders used MedDRA for the first time. They coded 260 events independently and a medical committee later determined whether the coding was accurate [8]. In 12% of the cases, the coding resulted in two different Preferred Terms. When the comparison was made higher up in the MedDRA hierarchy, the difference was smaller, around 5%, indicating that the different Preferred Terms were related to some extent. The 12% difference can, however, be important because statistical analysis of adverse events is often done at this level in the hierarchy. The medical committee determined that in 8% of the cases, which the coders perceived as easy to code, the coding was nevertheless inaccurate. The study was extended and 1640 events were coded. The inaccuracy rate was around 10% in the larger sample but interobserver variation was not reported.

A study by Brown et al. from 1996, the early days of MedDRA, found that from existing product labeling 90% of the terms could be exact or acceptably matched in MedDRA [9]. The next year the Brown et al. compared how accurate adverse events from clinical trials could be coded in MedDRA versus COSTART. This study also found that 90% of the matches were exact or acceptable with MedDRA but only 62% with COSTART [10]. The authors pointed out that the entire COSTART dictionary was imported in MedDRA.

White et al. looked at 204 post marketing surveillance events [11]. When the same verbatim text was coded with MedDRA and WHO-ART 32 pairs (16%) were rated as medically different. In 13 cases, the WHO-ART code was included in the product label and the MedDRA code was not.

In a paper from 2002 Brown was concerned about the increasing amount of terms in MedDRA. He showed that 315 WHO-ART terms could be mapped to 943 MedDRA terms. In 2004 Brown compared adverse events reported in the Physician's Desk Reference from 10 randomly selected drugs with corresponding MedDRA terms. He found that some adverse events (e.g. infection and pain) corresponded to hundreds of terms in MedDRA [2].

The constant updating of MedDRA has also been a source of concern. Toneatti et al. also examined the impact of updating from version 5.0 to 6.1. Out of 436 unique Lowest Level Terms, 38 (9%) changed either the Preferred Term or the SOC related to them, or both [8].

Each Preferred Term is associated with one primary SOC. This SOC is predefined by MedDRA and users are not allowed to change this or anything else in the MedDRA hierarchy. The most appropriate primary SOC for an adverse event might, however, differ from study to study. In an HIV trial, 23% of primary SOCs were altered when using a predetermined strategy to choose the most appropriate primary SOC [12]. It demonstrates the subjectivity of the hierarchy.

There is often doubt about how an adverse event should be coded and therefore it is necessary to develop "coding guidelines" for each trial. A 45-page manual has been developed by an expert group to address more general issues, which means that coding can no longer be performed by a physician without special training . If a diagnosis and several symptoms – that are included in the diagnosis – are reported, several strategies can be used in coding this data. One strategy is to code both symptoms and diagnosis, another is to code the diagnosis and leave out the symptoms that are included in this diagnosis. It is recommended that coders do not make diagnoses based on reported symptoms.

The manual offers specific guidance on how to handle suicide and self harm. It explicitly states that an intentional overdose should be coded as an overdose, and not as a suicide attempt. "Cut her own wrist" should be coded as "self inflicted laceration" and only as a suicide attempt if the verbatim clearly states that the purpose was suicide. The unfortunate consequence of these recommendations is that suicide attempts become much harder to detect in pharmaceutical trials.

Infections can either be coded by the microorganism or the anatomic location of infection. The current recommendation is that chlamydial respiratory infection should be coded as "Chlamydial infection" [13]. "Chlamydial infection" will then represent respiratory and urogenital infections, even though it is clinically relevant to distinguish between these illnesses. When creating a rigid system that exclusively categorise events, it will always be possible to find examples that, in a given context, should have been categorized differently.

Another important factor that will effect whether an event is coded or not, is the definition of "treatment emergent adverse event". It is usually defined as any *new* adverse event or worsening of an existing condition after initiation of therapy [14]. Even though the definition seems quite clear, Nilsson et al. identified 26 different ways of defining treatment emergent adverse event. Depending on the selected strategy, the authors' test data returned from 2 to 7 adverse events [14]. One of the reasons for the many definitions is determination of initial severity. If the patient had several appointments before they actually got the active drug (run-in period), and they reported 'headache' but with varying severity during these visits it is unclear which severity should be used. It is very important because all following headaches in the actual trial with the same severity would not be considered an adverse event and would therefore not be coded [14]. The most important factor that influences the number of adverse events is the way that

Table 1. Description of included studies.

Author/Year	Aim	Study design	Main findings
Brown 1996 [9]	To determine MedDRA's adequacy in representing medical terms used in UK data sheets	A product from each of the main drug classes in the British National Formulary was scrutinised for medical terms which were then coded using MedDRA. Matches were classed for accuracy	Identical or acceptable matches for 90% of the side effects
Brown 1997 [10]	To compare MedDRA to the COSTART for specificity of coding clinical trial data and for the effects of coding on the analysis and presentation of safety data from the trial	Verbatim descriptions of adverse events from a phase II trial were coded by MedDRA and COSTART and the association was assessed for accuracy. The incidence of adverse events using the different dictionaries was compared.	Using MedDRA resulted in more exact matches than using COSTART (90% vs 62%). With MedDRA 267 codes were used, with COSTART only 169. The two terminologies gave different breakdowns of adverse events
Brown 2002 [20]	To explore the numerical and conceptual relationships between WHO-ART and the MedDRA and their ability to detect signals	A sample of approximately one sixth of all WHO-ART preferred terms was taken. MedDRA was searched for each of these terms to find the best match	315 WHO-ART terms were identified and were matched with 943 MedDRA preferred terms
Brown 2004 [2]	To identify common adverse events in clinical trials by looking at product labeling and comparing this to MedDRA terms	Adverse events from 10 randomly selected drugs in the Physician's Desk Reference were compared with MedDRA terms	Some terms in the product labels were associated with hundreds of MedDRA terms. E.g. "infection" (several hundreds) and "pain" (168 items)
Fescharek 2004 [21]	To investigate MedDRA's impact on retrievel strategies, analysis and presentation of coded data	Comparison of trial data coded in WHO-ART with the same data recoded in MedDRA	In WHO-ART 214 different terms were used; whereas in MedDRA 312 different terms were used. They were grouped quite differently
Journot 2008 [12]	To be able to use the MedDRA hierarchy for data analysis by redefining the hierarchy to fit trial objectives	The authors developed a new general 5-step strategy to select a SOC (system organ class) for an adverse event as trial primary SOC, consistent with trial-specific objectives. This was applied to clinical trial data and compared to the original MedDRA hierarchy	Altogether, 23% of MedDRA primary SOCs were modified
Nilsson 2001 [14]	To analyse the impact of defining "treatment emergent adverse events"	Since only treatment emergent adverse events are reported in trials the authors identified in how many ways this could be defined and the consequences on test data	At least 26 different strategies for censoring adverse events exist. Depending on the chosen strategy the same data resulted in 2 to 7 adverse events.
Toneatti 2005 [8]	To assess the feasibility of coding with MedDRA. To develop an approach for MedDRA implementation within an institutional research unit that contributes to an efficient, concise and reproducible event coding	1) Two blinded coders used MedDRA to code 260 verbatim descriptions of adverse events from a clinical trial and reported difficulties in coding. Variability between the two coders was measured and accuracy was determined by a medical coding committee. 2) MedDRA 6.1 was applied to both the list of frequent adverse events and a trial coded with MedDRA 5.0	1) 32 adverse events (12%) were coded differently by the two coders; 13% of the adverse events were assessed to be "non-accurate". 2) When changing to a new MedDRA version, 38 (9%) adverse events changed.
White 1998 [11]	To obtain a preliminary assessment of the impact of MedDRA on the frequency of expedited adverse event reports based on current (non-MEDDRA) labeling	Verbatim adverse event reports (surveillance) for two different marketed drugs were coded with WHO-ART and MEDDRA and it was determined whether the code was mentioned in the product label. A rating scale was used to quantify the differences	Twenty-seven terms (13%) had some syntactic differences although these were not considered medically significant. Thirty-two terms (16%) were rated as medically significantly different but did not affect the label. Ten terms (5%) were rated as both medically different and resulted in a labeling discrepancy
Zhao-Wong 2006 [17]	The purpose was to obtain more user input on issues related to the feasibility study and MedDRA terminology in general	A survey of MedDRA users performed by the MSSO, the organization maintaining MedDRA	Received 12 responses out of 29 invited. The majority of MedDRA users relied on primary paths for both re-porting and analysis. The usage of secondary links was limited
MedDRA Term Selection 2011 [13]	To aid medical coders in choosing codes consistently	Not a study but a manual	Describes many situations where there might be doubt on how to code a reported adverse event and suggests a solution
MedDRA Data Retrieval 2011 [16]	To aid investigators in presenting adverse events	Not a study but a manual	Describes how adverse events can be presented by the hierarchy and how to use standard and custom searches to lump related adverse events together

doi:10.1371/journal.pone.0041174.t001

adverse events are compared. If a patient had "headache" at visit 1 and "head pain" at visit 2, does that represent the same event or are they different? Obviously, more details would be preferred but the two terms would normally be coded and then compared on a predefined level in the hierarchy of MedDRA. It will obviously make a big difference whether you compare verbatim text, Lowest Level Term or Preferred Term. If verbatim text is chosen, then "headache" and "head pain" would be considered two different adverse events [14]. Comparing on Preferred Term level would probably mean that the events were considered identical. The

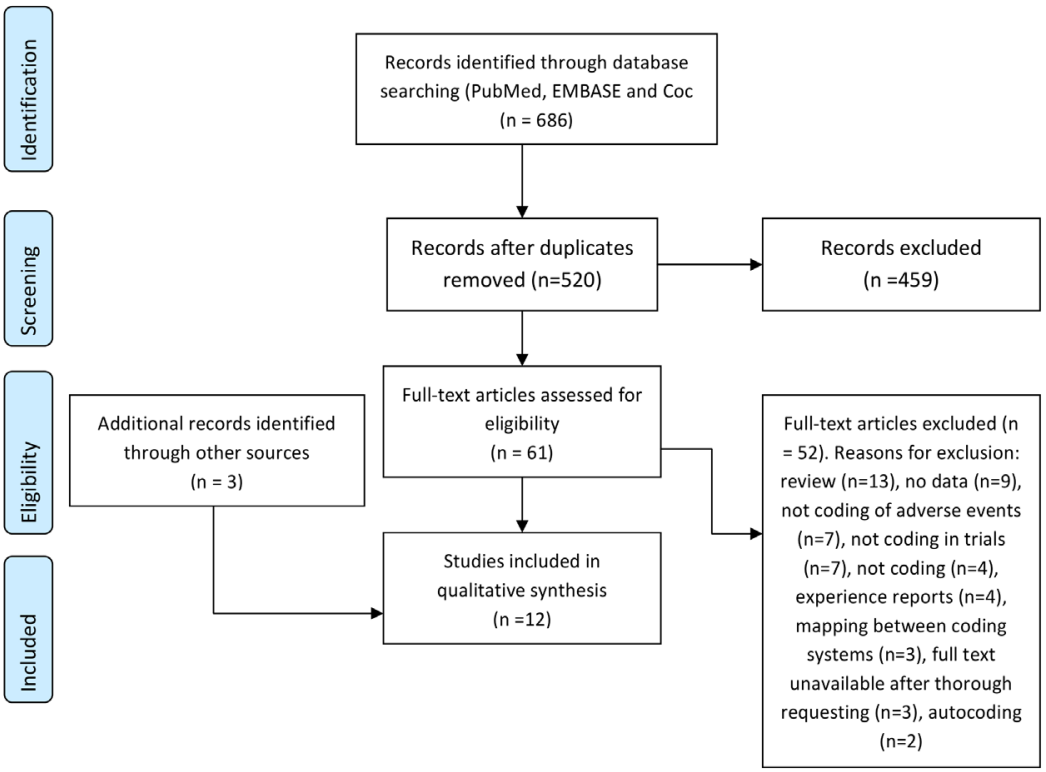


Figure 2. Flow chart of the process of identifying studies.
doi:10.1371/journal.pone.0041174.g002

drawback is that you might overlook two medically different events that are lumped together. Obviously, more adverse events will be censored if the Preferred Term level is used. Brown et al. (1997) also looked at the impact of coding on the way harms were presented. Several new adverse events were detected using MedDRA because of splitting of existing groups. This meant that the list of the 10 most common adverse events changed substantially when MedDRA was used. They concluded that "use of a different terminology can alter the apparent safety profile of a drug" [10]. The same conclusion was reached by Fascharek et al. after coding the same trial data in WHO-ART and MedDRA .

In a survey of only 12 MedDRA users it was established that the usage of "secondary links" is limited [15]. "Secondary links" are searches and secondary SOCs that will make it possible to lump related adverse events together thereby increasing statistical power. MedDRA has more than 18,000 Preferred Terms and, as we have described above, there is a risk of signal dilution compared to previous dictionaries with less terms. Even the developers of MedDRA acknowledges that the hierarchy cannot be relied on to retrieve exhausting information about adverse events [16]. Several authors have prompted for regulatory guidance on MedDRA implementation [17]. The expert group states that simple summaries might not always be sufficient, and that you may have to explore the safety data in greater detail [16].

Discussion

The only interobserver study of adverse event coding we found showed that 12% of the adverse events at Preferred Term level were coded differently by two coders. This could be quite significant for some trials but obviously it depends on what symptoms were coded differently and how. Important interpretation is done by the medical coder, and 8% of the coding was declared as medically inaccurate when rated by experts. This study has not led to further investigations of the subject, which is surprising.

The constant development of more terms in MedDRA might intuitively lead to less interobserver variation because there will be more exact matches to the verbatim text. Conversely, it might also lead to increased variation because it becomes difficult to code nonspecific terms, but this has not been studied.

If there is great uncertainty on how adverse events are coded it will lead to non-differential misclassification. This will underestimate the relationship and may result in failure to detect important adverse events.

With MedDRA it is possible to match the investigator's verbatim descriptions more closely because of the increasing amount of terms. The drawback is that it becomes harder to statistically detect adverse events that are related but do not present themselves in the same way in each patient, i.e. signal dilution, because events are split into subcategories. Advanced searching and data analysis in MedDRA, where related categories and Preferred Terms are lumped together, have been developed to try and counteract this problem, but a survey showed that that these tools are not used [15]. The recommendations by the expert group on when to explore adverse events are vague and it is even recommended to design the analysis post hoc [16], which carries a

risk of bias. If these problems could be solved, it would lead to more transparent handling of adverse events. Better guidelines would need to be developed by regulatory authorities.

Because of "background noise", summary tables at SOC level are usually only efficient in finding adverse events that occur frequently in the treatment group and rarely in the placebo group. For example, if a trial runs over 2 years, most patients in both the treatment and the placebo group might have experienced headaches. With such background noise, it will be almost impossible to detect any other neurological diseases or symptoms at the SOC level.

Another problem with summary tables at SOC level is that sometimes related adverse events are not even in the same category. In MedDRA there are disorders that are defined by laboratory tests . For example, "Hepatic function abnormal" belongs to the SOC "Hepatobiliary disorders" whereas "liver function test abnormal" belongs to "Investigations". In a summary table, these identical adverse events would be presented in two different categories.

The CONSORT group recommends that coding strategies should be reported [18]. They also recommend that adverse events should be defined. Unfortunately, MedDRA doesn't hold any formal definitions of adverse events. In the protocol, one can of course define important expected adverse events, but the consequence is that the investigator and the coder will have to look in two different systems. The usage of definitions is usually limited. Lack of definitions is an important limitation and makes comparison of harms between different trials problematic.

The many ways to define treatment emergent adverse events, and hence censor adverse events, can result in bias because data analysis is often done unblinded [19] and the most favourable strategy might be chosen.

In package inserts, common and serious adverse events are reported. With MedDRA, we get a greater variety of adverse events but each one becomes less frequent. As the package inserts are mainly based on frequency, we would expect the total number of adverse events to go down using MedDRA. This makes it difficult to compare adverse events historically and newer package inserts should therefore be interpreted cautiously.

The increased specificity of MedDRA terms might be partly responsible for the common failure to detect important adverse events before drug approval, leading to many patients being harmed by dangerous drugs. In post-marketing surveillance studies, sensitive techniques for detecting adverse events have been developed, e.g. data mining and lumping of hundreds of related terms to counteract the problems of splitting adverse events. However, as observational studies can only detect strong signals reliably, we should have more emphasis on detecting adverse events in the clinical trials, perhaps by using some of the same techniques.

The organization behind MedDRA advertises that it is clinically validated but defines this as "developed and maintained by medical experts" [4]. This is not a guarantee that coding in MedDRA is reproducible nor is it a guarantee that adverse events are identified as well as, or better than, previous dictionaries. It is essential that the many different ways to define and handle adverse events becomes standardized or at least documented. To decrease interobserver variation coders and investigators should be meticulous and well trained. We recommend that a thorough interobserver study of coding should be performed elucidating both the magnitude and the nature of the problems with variability in coding. Brown 1997 [10] investigated differences in the accuracy of coding and incidence of adverse events in a clinical trial of an unspecified neuroleptic drug using COSTART and MedDRA. Since MedDRA has changed significantly over the past 15 years, the ability of MedDRA to identify known adverse events compared with older dictionaries should be re-evaluated using trial data for a known drug.

Limitations

Because of the constant development of MedDRA, the results of the interobserver study and other studies we have included might no longer apply. Our study might be subject to publication bias since MedDRA is predominately used by the pharmaceutical industry, which might have experimented with MedDRA during its implementation without publishing their results.

Conclusion

Important differences in coding between two coders exist but the consequences have been poorly elucidated. The implementation of MedDRA has led to a more specific coding system but it has made signal detection much more difficult and has had great consequences for product labeling. Strategies to improve detection of adverse events have been developed but are rarely used. It is very surprising that so little research has been performed in this important area for public health. This needs to be remedied.

Supporting Information

Checklist S1 PRISMA checklist.
(DOC)

Text S1 Search strategy for PubMed, Cochrane and EMBASE.
(DOCX)

Author Contributions

Conceived and designed the experiments: JBS PCG. Performed the experiments: JBS EM. Analyzed the data: JBS EM PCG. Wrote the paper: JBS EM PCG.

References

1. Corrigan OP (2002) A risky business: the detection of adverse drug reactions in clinical trials and post-marketing exercises. Soc Sci Med 55: 497–507.
2. Brown E (2004) Using MedDRA: implications for risk management. Drug Saf 27: 591–602. 27810 [pii].
3. Wood K (1994) The medical dictionary for drug regulatory affairs (MEDDRA) project. Pharmacoepidemiology and drug safety 3: 7–13.
4. Mozzicato P (2009) MedDRA. An Overview of the Medical Dictionary for Regulatory Activities. Pharm Med 23: 65–75.
5. Kondro W, Sibbald B (2004) Drug company experts advised staff to withhold data about SSRI use in children. CMAJ 170: 783.
6. Furukawa TA (2004) All clinical trials must be reported in detail and made publicly available. BMJ 329: 626. 10.1136/bmj.329.7466.626-a [doi];329/7466/626-a [pii].
7. Moher D, Liberati A, Tetzlaff J, Altman DG (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 339: b2535.
8. Toneatti C, Saidi Y, Meiffredy V, Tangre P, Harel M, et al. (2006) Experience using MedDRA for global events coding in HIV clinical trials. Contemp Clin Trials 27: 13–22. S1551-7144(05)00142-4 [pii];10.1016/j.cct.2005.09.009 [doi].
9. Brown EG, Clark E (1996) Evaluation of MEDDRA in representing medicinal product data sheet information. Pharmaceutical Medicine 10: 111–118.
10. Brown DR, Brown EG, Moulvad TB (1997) A comparison of two medical terminologies in coding and analysing clinical trial safety data. International Journal of Pharmaceutical Medicine 11: 85–89.
11. White CA (1998) A preliminary assesment of the impact of MedDRA on adverse event reports and product labeling. Drug Information Journal 32: 347–362.
12. Journot V, Tabuteau S, Collin F, Molina JM, Chene G, et al. (2008) About the necessity to manage events coded with MedDRA prior to statistical analysis: proposal of a strategy with application to a randomized clinical trial, ANRS 099 ALIZE. Contemp Clin Trials 29: 95–101. S1551-7144(07)00082-1 [pii];10.1016/j.cct.2007.05.007 [doi].

13. ICH Secretariat (2011 October) MedDRA Term Selection: Points to Consider Document, release 4.2. Available: http://www.meddramsso.com/files_acrobat/ptc/9491-1410_TermSelPTC_R4_2_sep2011.pdf. Accessed 2011 Sept 11.
14. Nilsson ME, Koke SC (2001) Defining treatment-emergent adverse events with the medical dictionary for regulatory activities (MedDRA). Drug Information Journal 35: 1289–1299.
15. Zhao-Wong AC, Rump E, Moraleda T, Mozzicato P, Revelle P (2006) Proposed terminology changes to facilitate the analysis of MedDRA-coded data. Drug Information Journal 40: 291–303.
16. ICH Secretariat (2011 October) MedDRA® Data retrieval and presentation: Points to consider, release 3.2. http://www.meddramsso.com/files_acrobat/ptc/9610-1410_DatRetPTC_R3_2_sep2011.pdf. Accessed 2011 Dec 22.
17. Brown E (2003) Methods and pitfalls in searching drug safety databases utilising the Medical Dictionary for Regulatory Activities (MedDRA). Drug Saf 26: 145–158. 260302 [pii].
18. Ioannidis JP, Evans SJ, Gotzsche PC, O'Neill RT, Altman DG, et al. (2004) Better reporting of harms in randomized trials: an extension of the CONSORT statement. Ann Intern Med 141: 781–788. 141/10/781 [pii].
19. Gotzsche PC (1996) Blinding during data analysis and writing of manuscripts. Control Clin Trials 17: 285–290. 0197-2456(95)00263-4 [pii].
20. Brown EG (2002) Effects of coding dictionary on signal generation: a consideration of use of MedDRA compared with WHO-ART. Drug Saf 25: 445–452. 250609 [pii].
21. Fescharek R, Kubler J, Elsasser U, Frank M, Guthlein P (2004) Medical dictionary for regulatory activities (MedDRA): Data retrieval and presentation. International Journal of Pharmaceutical Medicine 18: 259–269.

BMJ

BMJ 2013;346:f2231 doi: 10.1136/bmj.f2231 (Published 23 April 2013)

RESEARCH

Searching for unpublished data for Cochrane reviews: cross sectional study

Jeppe Bennekou Schroll *PhD student*[1], Lisa Bero *professor*[2], Peter C Gøtzsche *professor*[1]

[1]The Nordic Cochrane Centre, Rigshospitalet, Dept 7810, Blegdamsvej 9, DK-2100 Copenhagen Ø, Denmark; [2]Clinical Pharmacy and Health Policy Studies, University of California San Francisco, San Francisco, CA, USA

Abstract

Objective To describe the experiences of authors of Cochrane reviews in searching for, getting access to, and using unpublished data.

Design Cross sectional study.

Setting Cochrane reviews.

Participants 2184 corresponding authors of Cochrane reviews as of May 2012.

Main outcome measure Frequencies of responses to open ended and closed questions in an online survey.

Results Of 5915 authors contacted by email, 2184 replied (36.9% response rate). Of those, 1656 (75.8%) had searched for unpublished data. In 913 cases (55.1% of 1656), new data were obtained and we received details about these data for 794 data sources. The most common data source was "trialists/investigators," accounting for 73.9% (n=587) of the 794 data sources. Most of the data were used in the review (82.0%, 651/794) and in 53.4% (424/794) of cases data were provided in less than a month. Summary data were most common, provided by 50.8% (403/794) of the data sources, whereas 20.5% (163/794) provided individual patient data. In only 6.3% (50/794) of cases were data reported to have been obtained from the manufacturers, and this group waited longer and had to make more contacts to get the data. The data from manufacturers were less likely to be for individual patients and less likely to be used in the review. Data from regulatory agencies accounted for 3.0% (24/794) of the obtained data.

Conclusions Most authors of Cochrane reviews who searched for unpublished data received useful information, primarily from trialists. Our response rate was low and the authors who did not respond were probably less likely to have searched for unpublished data. Manufacturers and regulatory agencies were uncommon sources of unpublished data.

Introduction

Selective reporting of trials is common.[1] Thus despite the existence of hundreds of thousands of published randomised trials and thousands of updated Cochrane reviews, the true benefits and harms of many interventions are still unknown.

Recent studies have reported successes in obtaining details, including results, of unpublished clinical trials from licensing authorities and health technology agencies.[2-4] These sources have the potential to reduce reporting biases in reviews of drug interventions. The inclusion of unpublished or inadequately reported data in meta-analyses generally leads to more reliable effect estimates.[5] However, only a little over 10% of the Cochrane reviews from 2000-06 included unpublished trials.[6]

Unpublished data include complete trials that have never been published as well as specific outcomes that are not reported in published trials. For this study we considered data published even if published only in conference abstracts, research reports, and dissertations.

The Cochrane handbook suggests searching for unpublished data from the following sources: local experts, pharmaceutical companies, national and international trial registers (for example, clinicaltrials.gov), company trial registers, subject specific trial registers, and trial results registers.[7] Regulatory agencies are not mentioned in the handbook and no guidance is given on how to obtain data or protocols from such agencies. It is also unclear how the different sources should be prioritised—that is, which sources are most likely to supply useful data.

Many review authors have obtained unpublished trial protocols, reports, additional summary data, or individual patient data from a variety of sources. We provided an overview of the experiences of Cochrane review authors in searching for, getting access to, and using unpublished information from trials.

Methods

We conducted an online survey of corresponding authors of Cochrane reviews and protocols. The survey contained closed and open ended questions.

Correspondence to: J B Schroll js@cochrane.dk

Extra material supplied by the author (see http://www.bmj.com/content/346/bmj.f2231?tab=related#webextra)

Survey questionnaire

BMJ 2013;346:f2231 doi: 10.1136/bmj.f2231 (Published 23 April 2013)

We gathered information on how trial characteristics and data were obtained, types of data (for example, whole trials, missing outcomes, and additional analyses), difficulties encountered, and how the data were used. Our previous experience suggested that unpublished data are obtained in a wide variety of—and sometimes unexpected—ways, indicating that open ended and qualitative questions would provide useful information that we could not collect using only structured questions.

Sample selection

We retrieved a list of all corresponding authors of Cochrane reviews and protocols through the Cochrane Collaboration Information Management System (Archie). This information was imported into an online survey application (SurveyMonkey) and we invited all authors by email to participate. If the invitees did not respond within 10 days, we sent a reminder. A second reminder was sent after 20 days, and a final one after 30 days. Respondents who only partially filled in the survey also received a reminder. We collected data from the 21 May to 8 August 2012.

Statistical analysis

We reported frequencies of responses for each question response choice. Partial responses were also included. During data collection, but before we analysed the data, we hypothesised that drug manufacturers might differ from the other data sources. We used a χ^2 test to compare the proportions of the recorded characteristics between manufacturers and non-manufacturers. We dichotomised scales with more than two categories.

Questionnaire

The survey was tested by 10 pilot testers. Their comments were implemented in the final version.

The respondents were asked to answer the questions in relation to a review in which they had been directly involved. If the respondents had been involved in several reviews, they were encouraged to choose one that included searching for unpublished data and that had resulted in experiences that could possibly benefit other review authors. Respondents who did not search for or obtain unpublished data were asked to give a reason. Respondents who did search for and obtain data were asked to provide a citation for their work and to state their primary source of unpublished data. They could choose between manufacturers, regulatory agencies, investigators, commercial and non-commercial trial registers, funders, ethics committees, and others. For their primary source they were asked to provide a name, year of query, number of attempts at getting the data, delay until the data were obtained, method of communication, reasons for thinking that data might be available, details on the data obtained, and whether the data were used in their review. Respondents could also provide information on secondary sources of unpublished data. Finally, they were asked whether they investigated a drug intervention, what the biggest difficulties were in obtaining unpublished data, and if they had any additional comments. The survey contained 82 questions but took less than five minutes to complete, as not all questions were relevant for each specific case (see supplementary file).

Results

We sent the questionnaire to 5915 corresponding authors of Cochrane reviews and protocols; 2184 replied (response rate 36.9%), 1889 of whom completed all questions in the survey (figure⇓). Most of the dropout occurred when the respondents were asked to provide a citation to the work they authored (n=194).

Percentages can add to more than 100%, as several of the response options were not mutually exclusive. Of the 2184 respondents, 528 (24.2%) did not search for unpublished data. The reasons, "not expecting success," "not expecting reliable data," and "too time consuming," each accounted for around 20% of the replies (table 1⇓). The most common reason given was "other" (n=265, 52.4%) of which the majority specified that the review was still in an early phase and that the search had yet to be performed (n=177). Box 1 lists other reasons, which can be categorised into the following groups: only wanted to include published data and therefore deliberately chose not to search for unpublished data (11 authors), found published data and therefore did not think it was necessary to search for unpublished data (n=5), did not know how to search for unpublished data (n=26), thought searching for unpublished data was the responsibility of the trial search coordinator in the Cochrane review group in question (n=5), tried to search but failed (n=12), and simply stated it was not relevant (n=14).

Among the 1656 authors who did search for unpublished data, 730 (44.1%) never obtained any, 913 obtained data, and 13 did not reply to this question. The most common reason for not obtaining data was never receiving a response (66.2% of 717; an additional 13 did not specify a reason, table 1). The second most common reason was that the contacted person did not have the data (39.3% of 717). By analysing the comments in the "other" category, we found additional common reasons for not obtaining data: no unpublished studies were found, investigators were reluctant to release data until the study was published, commercial confidentiality, promised data but never delivered, and author's contact information could not be found. In some instances, authors only wanted to deliver data if they became coauthors of the review. See box 2 for more examples.

A total of 676 respondents gave details on 794 sources that provided data. The most common data source was trialists, accounting for 73.9% of the 794 (table 2⇓). Only 6.3% of the data came from manufacturers, 3.0% from regulatory agencies, and 6.3% from non-commercial trial registers. The "other" category accounted for 8.3% (n=66), where the most common sources were dissertations and conferences (which these authors regarded as unpublished, contrary to our definition of published data). Journal editors, Cochrane review groups, the World Health Organization, librarians, consumer support groups, and Google searches also contributed. The respondents did not contact any sources not already listed in the Cochrane handbook.[6] The most common regulatory agencies that provided data were the Food and Drug Administration (n=11) and the European Medicines Agency (n=4).

Source details

The most common way to approach sources of information was by email (table 1). Using websites was the most common approach specified by respondents that chose the "other" category. In 75.2% of the 794 cases, 1-3 contacts were enough, but in 6.4% of the cases (n=51) more than 10 contacts were necessary to get the data (table 1).

Unpublished data were provided in less than a month in 53.4% (n=424) of the cases (table 1), but in 9.1% (n=72) of the cases, the authors had to wait for more than six months. The most common reason why authors contacted a specific source was that they knew a trial had been conducted (61.1% of the 794, table 1). The idea to contact a specific data source only came from the Cochrane handbook in 4.2% of the 794 cases. Authors

BMJ 2013;346:f2231 doi: 10.1136/bmj.f2231 (Published 23 April 2013)

Box 1: Quotations to highlight reasons for not searching for unpublished data

- One of our inclusion criteria is that the data must be published
- We are still in the process of data extraction of papers, if we do get very little publications we might think about unpublished data
- Was not aware there was any unpublished data in my topic area
- Did not know it was possible to get
- Haven't tried yet. Assumed this was done as part of the searching process done by the Cochrane group searcher
- Where I have previously asked for unpublished data, authorship has been requested

Box 2: Quotations from "other" category for reasons why data were never obtained

- Got a response to say they would look for data, but then no further response
- Drug company responded that data were confidential
- Drug company stated it could not be used for research, only for formulary decision making
- Said they were preparing for future publication
- Respondent said they did not think the information I requested was relevant/helpful to the review question (this was a drug sponsored trial for which I requested subgroup data)
- Respondent said it was unnecessary for my clinical question
- Big pharma said they didn't regard the question of sufficient clinical value to warrant disclosing the data
- It was too long since the original studies were published. Some authors were uncontactable and we had answers from authors who had thrown away the data we needed

quite often specified that they learnt about unpublished data at conferences, either through personal contacts or abstracts. Trial registers were also used to identify unpublished studies or missing outcome data. Published papers with poor reporting could also be used to find missing outcome data by contacting the authors of the papers. Other sources were whistleblowers, peer reviewers who drew attention to unpublished trials, and meta-analyses of unpublished trials, sometimes done by the manufacturers—for example, pooling premarket studies to increase power. One respondent routinely contacted all manufacturers of a drug and another respondent always approached corresponding authors to confirm the validity of data extraction and to query unpublished data.

In 44.3% of the cases where data were obtained (295/666), the authors investigated a drug intervention. The time involved in searching for unpublished data was the most challenging element (41.0% of 666, table 1). Poor organisation and readability of the data was challenging for 20.9% and 9.8% of the 666 respondents who obtained data, respectively. Thirty seven per cent had no problems and 16.4% specified challenges not covered by the standard answers. The most common reasons were that the authors did not receive data or did not receive a reply (see examples in box 3).

Data obtained

The most common outcome data authors obtained were unpublished summary data from already published trials: this was supplied by 50.8% of the 794 data sources that provided data (table 3⇓). Missing data on outcomes (28.5%) and individual patients (20.5%) was also common. Data on harms were rare (8.4%). A total of 17.5% of the 794 cases had obtained "other" data, which were mostly data on methodological quality (randomisation, blinding, etc). Respondents also acquired subgroup analyses, theses, information about ongoing trials, and reports of protocol modifications that had not been reported. However, some data were partial, redacted, and subject to confidentiality agreements. Most used the acquired data in their review (82.0% of 794, table 1). The most common reason for not using the data was that they were in an unusable form (6.3%). Eight per cent chose "other" and the majority specified they had not used the data because their review was ongoing.

In around a third (267 of 794) of the cases the authors got information from previously unpublished trials, and in around two thirds (562 of 794, not mutually exclusive) they got additional data from already published trials. It could be suspected that the strategy for accessing unpublished trials compared with unpublished data differed. We performed a post hoc subgroup analysis and found no difference between the number of contacts needed before the author received data.

Drug and device manufacturers

The authors who obtained data from drug and device manufacturers were more likely to have to contact them 10 or more times compared with authors obtaining data from other sources (24% *v* 5%, P<0.001, table 4⇓). They also more frequently waited for more than one month (74% *v* 45%, P<0.001), and more frequently the contact was in person or by telephone (36% *v* 13%, P<0.001). Manufacturers less frequently supplied individual patient data than other sources (12% *v* 26%, P=0.02). Data from non-manufacturers were more often used and it was more common that the respondents reported that there were no difficulties compared with manufacturers. However, these differences were not significant (P=0.07 in both instances).

Discussion

A large proportion (around three quarters) of Cochrane review authors searched for unpublished data. A large fraction of those who did not search for unpublished data did so because their work was still ongoing, but another large fraction abstained from searching because they did not expect success. Searching for unpublished data from already published trials is problematic because authors may be difficult to locate and rarely respond.[8] Around 20% of authors refrained from searching unpublished data because they did not expect them to be reliable.

In our survey, 55.1% of those who searched for data obtained them and most (82.0%) used these data in their review. This suggests that the methodological rigor (or quality) of the data are adequate even though some of the data were of a nature where risk of bias assessment was pointless (additional point estimates, standard deviations, etc). Other studies have also

BMJ 2013;346:f2231 doi: 10.1136/bmj.f2231 (Published 23 April 2013)

Box 3: Quotations about main challenges in incorporating unpublished data in reviews

Data for one trial were provided in an old database format that was very difficult to access and navigate

Delineating what was useable from what was not, especially as we were not replying on study design as a filter—this made it a nightmare

We just weren't sure what we had been sent was right there were discrepancies between published report and data provided. When I asked the author for clarification they did not respond

In the one case that I received IPD [individual patient data], I didn't use this because the amount of data was overwhelming and would have taken too much of my time to decipher

evaluated the reliability of published and unpublished trials without finding differences.[9 10]

The respondents' last common concern was that searching for unpublished data was time consuming. This is not necessarily the case. One study[8] found that when the source was contacted by email, the reply arrived within a median of one day. In our survey, more than half of the authors had received their data within a month. But even though it might be time consuming, completely omitting searching for unpublished or inadequately reported data is a risky strategy, as such data will generally be less positive than published data.[1 2 4 5 11] Several respondents refrained from searching unpublished data because they found published studies, but this strategy cannot be recommended as, on average, it leads to biased reviews.

Some respondents thought that the trial search coordinator in Cochrane review groups searched for unpublished data, which may not be the case. Trial registers should always be consulted and this could be done by the search team. However, querying authors for missing outcomes or missing data and additional studies can only be done by the authors of the review, who have in-depth knowledge of the literature. Lastly, some respondents abstained from searching because of previous demands for authorship. This is, hopefully, rarely the case and should not discourage authors from searching for unpublished trials.

Almost half of the respondents who sought unpublished data obtained none. The most common reasons were that they never received a reply or were told that no data were available. Another 54 were told it was too much trouble to deliver the data. In a few cases, confidentiality and lack of interest in helping were the obstacles. We have experienced a drug company that only wanted to deliver unpublished data to a Cochrane review if they saw the draft manuscript. This was obviously unacceptable as the delivery of data should not depend on what the drug company or any other data source thinks about the preliminary results.

It was surprising that only 6.3% of authors got data from drug and device manufacturers. When our respondents tried to obtain data from manufactures they experienced longer waits, received fewer individual patient data, needed to make more requests, encountered more difficulties, and were less likely to be able to use the data. Owing to the low response rate in this study these associations should be interpreted with caution. It is nevertheless of concern that only 6.3% of authors received data from manufacturers as a large proportion of research funded by drug manufacturers remains unpublished.[12] The respondents who were successful more often contacted manufacturers by telephone or verbally (36%) than they did the non-manufacturers (13%). From respondents' comments we learnt that authors often knew that manufacturers had data because one of their own authors had been involved in the trials. On at least two occasions, respondents were told by drug manufacturers that their clinical question was not sufficiently relevant for the data to be released (see box 2). It has been well documented that manufacturers often refuse to share data.[13]

We had expected that research ethics committees and funders would rarely be a source of information, but it was unexpected that company owned trial registers and non-commercial trial registers in particular were also rarely a source of information (0.9% and 6.3%, respectively). The company owned trial registers might not contain relevant information, and non-commercial trial registers should be used more.

The authors often became aware of unpublished data through colleagues and websites. Only 4.2% got the idea from the Cochrane handbook to ask a specific source for data, despite the handbook containing a detailed section about searching for unpublished data.

Regulatory agencies are uncommon sources of data even though the FDA website has contained a lot of valuable data for decades, and even though the EMA opened up its archives in 2010.[14] Among the respondents who searched for data in 2011 and 2012 only 5% got data from regulatory agencies compared with 3% for our entire population.

Among the 24 authors who obtained data from regulatory agencies, only seven got full reports and only one unique review incorporated the data in the review. Some authors might not be aware of the amount of accessible data at regulatory agencies. We therefore suggest that the Cochrane handbook should mention regulatory agencies as a source of unpublished data and provide specific guidance on how to search the websites of the FDA and EMA as they are difficult to access.

Almost 21% of authors got individual patient data, primarily from trialists. Authors should be encouraged to request this type of data, and it can probably be done without compromising the response rate.[15]

How to obtain data

The respondents in our survey most commonly sent emails to corresponding authors, agencies, and companies, and this has also given the best response rates previously.[15] A combined approach with both email and letter might be even better.[15] Asking specific compared with open ended questions improves response rates.[15] This was also the experience of several of our respondents.

Guidance on how vigorously authors should search for unpublished data is needed. Our results suggest that authors should routinely ask trialists for more data when conducting a review. Searching a trial register (clinicaltrials.gov or similar) is also a good idea and is not time consuming. For data on drugs and devices, we suggest that the authors contact the regulatory agencies first, as it is time consuming and generally disappointing to go to the manufacturers.

Limitations of this study

The response rate in our study was low. A substantial part of our emails may have reached inactive email boxes or been caught by spam filters. We sent our survey to busy authors, and previous research has shown that time is a barrier and some people routinely bin surveys.[16] Our respondents were probably

BMJ 2013;346:f2231 doi: 10.1136/bmj.f2231 (Published 23 April 2013)

more likely to search for unpublished data than the authors who did not reply. Our sample might therefore not be representative of all authors of Cochrane reviews.

Conclusion

Most authors who searched for unpublished data received useful data, primarily from trialists. Manufacturers and regulatory agencies were seldom sources of unpublished data.

Contributors: PCG and LB conceived the study. The protocol was drafted by JBS; LB and PCG contributed. JBS created the online survey, sent out the invitations, and tabulated the data. All authors analysed the data. JBS drafted the manuscript; PCG and LB contributed. All authors had full access to all the data in the study. JBS is guarantor and takes responsibility for the integrity of the data and the accuracy of the data analysis. All authors have approved the final manuscript.

Funding: This study was funded by the Cochrane Collaboration Methods Innovation Fund. The funder had no influence on the study design, interpretation of data, or the decision to publish the results.

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

Ethical approval: This study was certified as exempt from human subjects review by the University of California human research protection programme (reference No 037504).

Data sharing: Anonymised datasets are available on request from the corresponding author at js@cochrane.dk.

1 Scherer RW, Langenberg P, von Elm E. Full publication of results initially presented in abstracts. *Cochrane Database Syst Rev* 2007;(2):MR000005.
2 Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med* 2008;358:252-60.
3 Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med* 2008;5:e45.
4 Rising K, Bacchetti P, Bero L. Reporting bias in drug trials submitted to the Food and Drug Administration: a review of publication and presentation. *PLoS Med* 2008;5:e217.
5 Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;291:2457-65.
6 Van Driel ML, De Sutter A, De Maeseneer J, Christiaens T. Searching for unpublished trials in Cochrane reviews may not be worth the effort. *J Clin Epidemiol* 2009;62:838-44.e3.
7 Lefebvre C, Manheimer E, Glanville J. Chapter 6: searching for studies. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.
8 Gibson CA, Bailey BW, Carper MJ, Lecheminant JD, Kirk EP, Huang G, et al. Author contacts for retrieval of data for a meta-analysis on exercise and diet restriction. *Int J Technol Assess Health Care* 2006;22:267-70.
9 MacLean CH, Morton SC, Ofman JJ, Roth EA, Shekelle PG. How useful are unpublished data from the Food and Drug Administration in meta-analysis? *J Clin Epidemiol* 2003;56:44-51.
10 Hart B, Lundh A, Bero L. The effect of reporting bias on meta-analyses of drug trials: re-analysis of meta-analyses. *BMJ* 2011;343:d7202.
11 Hopewell S, Loudon K, Clarke MJ, Oxman AD, Dickersin K. Publication bias in clinical trials due to statistical significance or direction of trial results. *Cochrane Database Syst Rev* 2009;(1):MR000006.
12 Sterne JAC, Egger M, Moher D. Chapter 10: addressing reporting biases. In: Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of 10. 23 Interventions. Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.
13 Stafford N. German agency asks Pfizer to disclose all reboxetine data. *BMJ* 2009;338:b2521.
14 Gotzsche PC, Jorgensen AW. Opening up data at the European Medicines Agency. *BMJ* 2011;342:d2686.
15 Young T, Hopewell S. Methods for obtaining unpublished data. *Cochrane Database Syst Rev* 2011;(11):MR000027.
16 Kaner EF, Haighton CA, McAvoy BR. 'So much post, so busy with practice—so, no time!': a telephone survey of general practitioners' reasons for not participating in postal questionnaire surveys. *Br J Gen Pract* 1998;48:1067-9.

Accepted: 26 March 2013

Cite this as: *BMJ* 2013;346:f2231

BMJ 2013;346:f2231 doi: 10.1136/bmj.f2231 (Published 23 April 2013)

What is already known on this topic

Unpublished data are less positive than published data

Omitting unpublished data in meta-analyses can bias the results

What this study adds

Authors of Cochrane reviews often search for unpublished data (75.8% in our sample) and around half of the authors who did, succeeded

Drug and device manufacturers infrequently provide data

Drug regulatory agencies should be used more

Tables

Table 1| **Responses by Cochrane authors to questions in survey**

Reasons	No (%)
Why didn't you try to get access to unpublished data?:	n=506*
Other or not relevant	265 (52.4)
Did not expect success	116 (22.9)
Too time consuming	108 (21.3)
Did not expect data to be reliable	102 (20.2)
Did not know it could be important	42 (8.3)
What were the main reasons why you did not obtain unpublished data?:	n=717*
Never received a response	475 (66.2)
Information requested was not available	282 (39.3)
Other	155 (21.6)
Respondent said it was too much trouble	54 (7.5)
How did you approach the source of data?:	n=794*†
Email	666 (83.9)
Telephone/in person	116 (14.6)
Letter/fax	114 (14.4)
Other	101 (12.7)
How many times did you make contact?:	n=794†
1-3	597 (75.2)
4-6	117 (14.7)
7-9	29 (3.7)
≥10	51 (6.4)
How long did it take before you got the data?‡:	n=794†
<1 week	101 (12.7)
1 week to <1 month	323 (40.7)
1 month to <6 months	298 (37.5)
≥6 months	72 (9.1)
How did you know the potential data source might have data?:	n=794*†
They conducted trials	485 (61.1)
Other	156 (19.6)
Colleagues	128 (16.1)
Through websites	90 (11.3)
I had no idea	86 (10.8)
Earlier published attempts at accessing data (for example, data from EMA or FDA)	34 (4.3)
Cochrane handbook	33 (4.2)
Court proceedings	7 (0.9)
What were the main challenges in incorporating the unpublished data in your review?:	n=666*
Time	273 (41.0)

BMJ 2013;346:f2231 doi: 10.1136/bmj.f2231 (Published 23 April 2013)

Table 1 (continued)

Reasons	No (%)
There were no difficulties	248 (37.2)
Poor organisation of the material obtained	139 (20.9)
Other	109 (16.4)
Poor readability of the material obtained	65 (9.8)
Sheer mass of obtained material	45 (6.8)
Expenses	24 (3.6)
Did you use the obtained data in your review?":	n=794†
Yes	651 (82.0)
Other	64 (8.1)
No, data not usable	50 (6.3)
No, trial quality not sufficient	13 (1.6)
No, trial excluded owing to other reasons	16 (2.0)

EMA=European Medicines Agency; FDA=Food and Drug Administration.

*Percentages may total >100 because responses are not mutually exclusive.

†Relates to number of data sources described by total of 676 respondents. Some respondents contacted more than one data source.

‡From first contact to when data were delivered.

BMJ 2013;346:f2231 doi: 10.1136/bmj.f2231 (Published 23 April 2013)

Table 2| **Source of data from which respondents obtained unpublished data**

Sources of data	No (%)
Trialists/investigators	587 (73.9)
Other	66 (8.3)
Manufacturers	50 (6.3)
Non-commercial trial register (for example, clinicaltrials.gov)	50 (6.3)
Drug and device regulatory agencies	24 (3.0)
Company owned trial register	7 (0.9)
Funders	7 (0.9)
Research ethics committees/institutional review boards	3 (0.4)
No of sources*	794 (100.0)

*676 respondents gave details on 794 sources.

BMJ 2013;346:f2231 doi: 10.1136/bmj.f2231 (Published 23 April 2013)

Table 3| **Common sources of outcome data**

Sources of outcome data	No (%)*
Missing data:	
Summary data (mean, standard deviation, sample size, etc)	403 (50.8)
Missing outcomes (for example, quality of life)	226 (28.5)
Individual patient data/raw data	163 (20.5)
Alternate analysis (for example, intention to treat)	96 (12.1)
Data on harms	67 (8.4)
Clinical study reports (regulatory authorities, full report)	45 (5.7)
Unpublished trials:	
Outcomes in summary format only	135 (17.0)
Study report without individual patient data	115 (14.5)
Individual patient data/raw data	63 (7.9)
Protocols	95 (12.0)
Contact information for trialists	77 (9.7)
Correspondence, approval letters, reviewer comments (for example, from regulatory agency)	5 (0.6)
Other (please specify)	139 (17.5)

*Percentage totals >100 because responses are not mutually exclusive.

BMJ 2013;346:f2231 doi: 10.1136/bmj.f2231 (Published 23 April 2013)

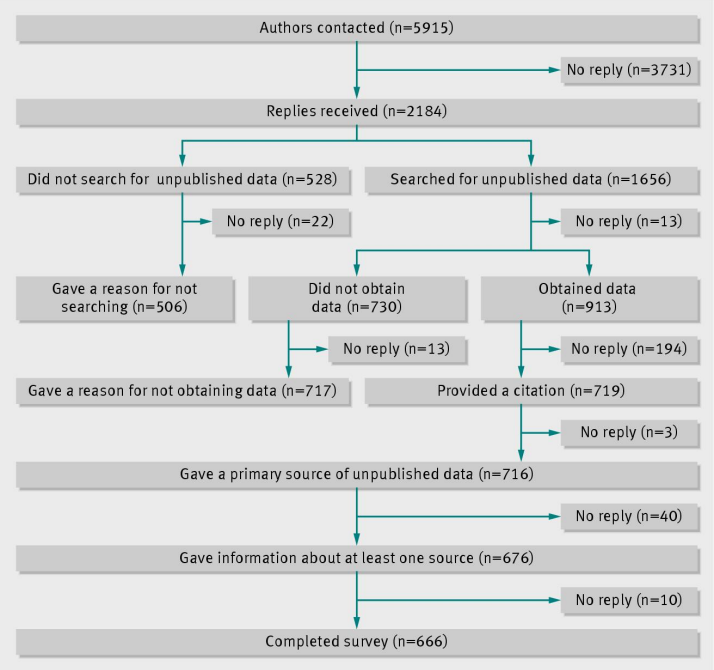
Table 4| Subgroup analysis comparing manufacturer compared with non-manufacturer data sources. Values are numbers (percentages) unless stated otherwise

Variables	Manufacturer (n=50)	Non-manufacturer (n=744)	P value*
Investigated a drug	42 (84)	254 (34)	<0.001
10 or more contacts needed†	12 (24)	39 (5)	<0.001
Waited a month or more†	37 (74)	333 (45)	<0.001
Contacted by telephone or verbally	18 (36)	98 (13)	<0.001
Used the data	36 (72)	515 (83)	0.07
Got individual patient data	6 (12)	196 (26)	0.02
No difficulties encountered	10 (20)	238 (32)	0.07

Each respondent could describe several sources (manufacturers, trialists, etc).
*χ^2 test.
†Data dichotomised from original four categories.

BMJ 2013;346:f2231 doi: 10.1136/bmj.f2231 (Published 23 April 2013)

Figure



Flowchart

BMJ

BMJ 2013;347:f4219 doi: 10.1136/bmj.f4219 (Published 4 July 2013)

VIEWS & REVIEWS

PERSONAL VIEW

Deaths in trials should always be reported

In Novo Nordisk's internal reports of trials of the diabetes drug repaglinide, **Jeppe Schroll** finds deaths that were not reported in published trials, potentially underplaying harms in subsequent analyses

Jeppe Schroll *PhD student, Nordic Cochrane Centre, Rigshospitalet, Dept 7811, 2100 Copenhagen Ø, Denmark*

Researchers generally do not publish what they planned to report in their protocols,[1] and important differences can also exist between internal trial reports and published papers.[2] It has been suspected that even deaths are sometimes omitted,[3] but there is little direct evidence of this.

In the final stages of conducting a Cochrane review about sulfonylurea treatment for patients with type 2 diabetes,[4] I realised that we had included only a few trials that considered outcomes important to patients. This was surprising given that we included several drugs that had been approved in the past 30 years, when clinical evaluation in trials was required. I searched for protocols on http://clinicaltrials.gov, but these trials were conducted before registration was mandatory.

I turned to the website of the US Food and Drug Administration (FDA) to look for reviews that might give clues about unpublished trials. Useful FDA reviews had compared the diabetes drugs repaglinide and nateglinide with sulfonylurea drugs. The sulfonylurea drugs themselves had been approved before the FDA started to publish reviews online. The FDA's repaglinide review[5] described five one-year trials. Having found only three of these trials, I contacted Novo Nordisk, and was told that the company did not share data that were not already published. After a second request, Novo Nordisk agreed to share data and sent five internal reports of six pages each.

The internal report of trial 050 described two deaths in the repaglinide arm: "a possible relation to the trial product could not be excluded by the investigator," it read. Despite the investigator's concern, the two deaths were not reported in the published paper,[6] which said: "The safety profile of repaglinide is similar to that of glyburide [glibenclamide], and there was no difference in adverse events." Serious adverse events—including the two deaths—were outlined in the internal report but had been omitted in the published paper.[6]

In trial 048, one death was reported in the internal report, this time in the comparator arm, but not in the published paper.[7] In trial 049, the internal report did not describe any deaths, but the published paper reported three deaths in the repaglinide arm and one in the sulfonylurea arm.[8] The published paper for trial 049 also reported 19 cardiovascular events (5%) in the repaglinide arm compared with only four (2%) in the sulfonylurea comparator—but nonetheless, the conclusion was that repaglinide was well tolerated and safe.[8] One of the never published trials (trial 046) had similar outcomes, with 25 cardiovascular events (14%) in the repaglinide arm compared with four (5%) in the sulfonylurea arm. However, the difference between groups was downplayed: the internal report concluded that the "frequency of adverse event[s] was similar." This conclusion was reached even after tolbutamide—an earlier, similar sulfonylurea drug—had been shown to increase the number of cardiovascular deaths.[9]

I also noted discrepancies in the number of patients with hypoglycaemia. In a published paper on trial 050, 26 (9%) patients in the repaglinide arm and 13 (9%) in the glibenclamide arm "experienced hypoglycemia."[6] But in the internal report, many more patients had a hypoglycaemic reaction (44 (16%) in the repaglinide arm; 20 (14%) in the sulfonylurea arm). Severe hypoglycaemia was defined in the published trial report, but the numbers of patients were not reported.[6] In the internal report, four participants in the repaglinide arm and one patient in the sulfonylurea arm had severe hypoglycaemia.

I asked Novo Nordisk about this discrepancy. It did not consider the data unpublished because pooled data from the five trials, including deaths, were published in a review by an independent researcher.[10] The company said that the reason for the few

js@cochrane.dk

BMJ 2013;347:f4219 doi: 10.1136/bmj.f4219 (Published 4 July 2013)

published hypoglycaemic events in trial 050 was that it only reported hypoglycaemic events in the maintenance period—but the company gave no explanation for this choice.

The review[10] pooled deaths (six deaths for repaglinide), but these differed from that stated in the internal and published reports (nine for repaglinide). The review also erroneously calculated the death rate as 0.1%, when in fact it was 0.5%. It is published as a supplement, and the abstract does not explain that it contains new data. The review's methods are not described, and only by reading the acknowledgments do you find out that the author had access to Novo Nordisk's internal reports. Only one of the three published trials are cited, and the "independent" researcher received grants from Novo Nordisk.[11]

The approval of repaglinide was based on these five trials,[5] which were designed to show equivalence with various sulfonylurea drugs for the unvalidated surrogate marker HbA_{1c} (glycated haemoglobin). However, we now know that a drug could have a positive effect on HbA_{1c} while it increases the risk of cardiovascular disease.[12] Only three of five trials on repaglinide were published, and there is also a lack of published trials for other diabetes drugs.[12]

There have been other cases where important adverse outcomes have been omitted from published papers. Cardiovascular events were left out of a paper about rofecoxib (Vioxx),[13] which has resulted in the loss of many lives. Additional cardiovascular events were also found in a study of rosiglitazone when the case reports were scrutinised by the FDA.[14]

The repaglinide trials show that mortality can be omitted in published papers even though the number of deaths was recorded and even though the investigators thought that it might be related to the drug. It should never be assumed that no deaths occurred just because none was reported, which is an especially important caveat for researchers conducting meta-analyses. Companies with a financial interest in downplaying harms are liable not to give an unbiased presentation of the results, which is why we need access to raw data.[15] Deaths and serious adverse events should always be reported—not as a pooled analysis in a substandard secondary publication with important errors, but in the original paper.

Competing interests: I have read and understood the BMJ Group policy on declaration of interests and have no relevant interests to declare.

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

1 Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;291:2457-65.
2 Rising K, Bacchetti P, Bero L. Reporting bias in drug trials submitted to the Food and Drug Administration: A review of publication and presentation. *PLoS Med* 2008;5:e217.
3 Williamson PR, Gamble C. Identification and impact of outcome selection bias in meta-analysis. *Stat Med* 2005;24:1547-61.
4 Hemmingsen B, Schroll JB, Lund SS, Wetterslev J, Gluud C, Vaag A, et al. Sulphonylurea monotherapy for patients with type 2 diabetes mellitus. *Cochrane Database Syst Rev* [forthcoming].
5 Center for Drug Evaluation and Research. Approval package for Prandin tablets 20-741/S009. 2000. www.accessdata.fda.gov/drugsatfda_docs/nda/2000/020741_S009_PRANDIN_TABS_AP.pdf.
6 Wolffenbuttel BH, Landgraf R. A 1-year multicenter randomized double-blind comparison of repaglinide and glyburide for the treatment of type 2 diabetes. Dutch and German Repaglinide Study Group. *Diabetes Care* 1999;22:463-7.
7 Madsbad S, Kilhovd B, Lagert I, Mustajoki P, Dejgaard A. Comparison between repaglinide and glipizide in type 2 diabetes mellitus: a 1-year multicentre study. *Diabet Med* 2001;18:395-401.
8 Marbury T, Huang WC, Strange P, Lebovitz H. Repaglinide versus glyburide: a one-year comparison trial. *Diabetes Res Clin Pract* 1999;43:155-66.
9 University Group Diabetes Program. A study of the effects of hypoglycemic agents on vascular complications in patients with adult-onset diabetes. II Mortality results. *Diabetes* 1970;19:789-830.
10 Owens DR. Repaglinide—Prandial glucose regulator: a new class of oral antidiabetic drugs. *Diabet Med* 1998;15(suppl 4):S28-36.
11 Owens DR, Zinman B, Bolli GB. Insulins today and beyond. *Lancet* 2001;358:739-46.
12 Nissen SE, Wolski K . Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *N Engl J Med* 2007;356:2457-71.
13 Curfman GD, Morrissey S, Drazen J. Expression of concern reaffirmed. *N Engl J Med* 2006;354:1193.
14 Psaty BM, Prentice RL. Minimizing bias in randomized trials: the importance of blinding. *JAMA* 2010;304:793-4.
15 Gøtzsche PC. Why we need easy access to all data from all clinical trials and how to accomplish it. *Trials* 2011;12:249.

Cite this as: *BMJ* 2013;347:f4219

ELSEVIER

Journal of Clinical Epidemiology ■ (2014) ■

Journal of Clinical Epidemiology

BRIEF REPORT

The Food and Drug Administration reports provided more data but were more difficult to use than the European Medicines Agency reports

Jeppe Bennekou Schroll[a,*], Maher Abdel-Sattar[b], Lisa Bero[b,c]

[a]*The Nordic Cochrane Centre, Rigshospitalet, Dept 7810, Blegdamsvej 9, DK-2100 Copenhagen Ø, Denmark*
[b]*Department of Clinical Pharmacy, University of California, San Francisco, Suite 420, Box 0613, 3333 California Street, San Francisco, CA 94118 USA*
[c]*Institute for Health Policy Studies, University of California, San Francisco, Suite 420, Box 0613, 3333 California Street, San Francisco, CA 94118 USA*

Accepted 30 June 2014; Published online xxxx

Abstract

Objectives: To compare the accessibility, comprehensiveness, and usefulness of data available from the European Medicines Agency (EMA) and the Food and Drug Administration (FDA) drug reports.

Study Design and Setting: This is a cross-sectional study. All new molecular drugs approved between January 1, 2011 and December 31, 2012 from the FDA and EMA Web sites were eligible.

Results: We included 27 drug reports. Most were searchable, but the FDA table of contents did not match the file's page numbers. Several FDA documents must be searched compared with a single EMA document, but the FDA reports contain more summary data on harms. Detailed information about harms was reported for 93% of the FDA reports (25 of 27 reports) and 26% of the EMA reports (7 of 27 reports). The reports contained information about trial methodology but did not include trial registry IDs or investigator names. All reports but one contained sufficient information to be used in a meta-analysis.

Conclusion: Detailed data on efficacy and harms are available at the two agencies. The FDA has more summary data on harms, but the documents are harder to navigate.

Keywords: Unpublished data; Systematic reviews; Drug regulation; Harms; FDA; EMA

1. Introduction

Doctors and decision makers cannot depend solely on articles published in medical journals. Articles are often biased [1,2], and some studies are partially published or not published at all [3]. Drug regulators have access to additional data through the companies' approval applications, for instance individual patient data on harms and analysis of efficacy data for multiple outcomes. In the Food and Drug Administration (FDA) drug reviews, some of these data are reported and can provide useful unpublished data for systematic reviews [4–8]. Although unpublished data can be obtained from FDA and, more recently, the European Medicines Agency (EMA) websites, they are rarely used in meta-analysis [9,10]. Difficult access to the FDA Web site could be part of the explanation [11] and other explanations could be lack of guidance on when and how to access data from regulators. Both the FDA and the EMA have made recent changes to the types of information they make available to the public. The purpose of this study was to compare the accessibility, comprehensiveness, and usefulness of information available on the FDA and EMA Web sites.

2. Methods

We identified all new molecular entities approved by the FDA from January 1, 2011 to December 31, 2012 through their Web site (http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm) and paired them with corresponding EMA drug approvals (http://www.ema.europa.eu). As in previous studies [7,12], biologics, orphan drugs, and diagnostics were excluded because they are reviewed using a different approval process. New molecular entities from the EMA Web site (http://www.ema.europa.eu) were also

Conflict of interest: All authors have completed the Unified Competing Interest form at http://www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author). J.B.S. and L.B. declare that they have no conflicts of interest. M.A.S. worked as a paid intern at Genentech (June 2011 to August 2011) and served as a paid liaison for Genentech at University of California, San Francisco (November 2011 to July 2013).

Funding: No specific funding for this project.

* Corresponding author. Tel.: +45-3545-7112; fax: +45-3545-7007.
E-mail address: js@cochrane.dk (J.B. Schroll).

0895-4356/$ - see front matter
http://dx.doi.org/10.1016/j.jclinepi.2014.06.019

What is new?

- Most FDA and EMA reports described trials in sufficient detail to enable them to be included in a meta-analysis.
- Most FDA reports contained detailed information about harms whereas the EMA reports did not.
- The information on the FDA site is harder to navigate, in general, than the information on the EMA site.
- Both agencies should be searched by researches conducting reviews.

identified in the same time period and paired with the corresponding FDA approval reports.

2.1. General description of drugs and documents

The medical review was our primary FDA resource, but we also extracted information from the approval letter, the Risk Evaluation and Mitigation Strategies (REMS), and the risk assessment reviews when available. For EMA-approved drugs, we examined only the European public assessment reports.

To determine how accessible the information was, two researchers (J.B.S. and M.A.S.) assessed whether each regulator provided structured reports, number of pages in the reports, a table of contents, a file that is searchable using text words, reviews in several languages, and lay summaries and whether it was possible to use direct Web links to resources.

To estimate how comprehensive the information was, we assessed whether information was redacted and, if so, whether a reason for the redaction was given and whether each regulator reported on unapproved drugs and relayed internal communications between reviewers and external communications between the applicant and the agency. We also assessed if the original trial protocols or the full trial reports were available and whether the agencies conducted additional statistical analyses.

2.2. Trial characteristics and efficacy data

We assessed the type of trial data that were available from each regulator and whether useful data for meta-analysis were available. Two researchers (M.A.S. and J.B.S.) independently assessed whether the FDA and EMA reports provided (1) an overview of the pivotal trials (the trials that were the basis of the clinical evaluation of the drug), (2) summary reports of each pivotal trial, (3) the number of pivotal trials and other submitted trials included, (4) the ClinicalTrials.gov ID for each trial, (5) names of the investigators, and (6) conflicts of interest among investigators. For the pivotal trials, the two researchers determined whether the inclusion and exclusion criteria for the trials were specified, whether outcomes were specified, whether numerical results were only available in a pooled format, and whether the efficacy results were presented in a manner that would allow for their inclusion in a meta-analysis (i.e., whether standard deviations and number of individuals were reported along with the numerical efficacy data).

2.3. Harms data

Two researchers (M.A.S. and J.B.S.) independently determined whether adverse event tables were present; whether safety data were provided for all completed trials or only for the indications being reviewed in the application; whether all important harms were reported (defined as common adverse events, mortality, serious adverse event, and withdrawals due to adverse events); whether numerical data on harms were reported; whether a risk management plan was included; whether regulators required further studies, follow-up on existing trials, or labeling restrictions; and whether REMS (FDA) or educational materials (EMA) were required by either or both agencies.

Any discrepancies between the two coders were discussed with the third author (L.B.). We planned a descriptive analysis of the differences between the data provided by the EMA and the FDA. We calculated the percentage of our binary outcomes.

3. Results

3.1. Drug characteristics

We found 57 new molecular entities approved by the FDA between 2011 and 2012; 14 orphan drugs and three diagnostic drugs were excluded. Another eight drugs were excluded for not having a corresponding approval on the EMA Web site (presumably the drug approval was never pursued in the European Union), four had only a pediatric plan (which we interpreted as pending), six had a pending status, and two had been withdrawn by the EMA, leaving 20 pairs approved by both agencies as of August 1, 2013. A similar search of the EMA Web site identified 50 new molecular entities approved in the same time period. We excluded 20 orphan drugs, one diagnostic drug, three with no FDA matches, two that were not approved by the FDA, and finally one in which the approval dates between the two agencies were more than 10 years apart and we believed that such a comparison would not be fair. The remaining 23 pairs identified through the EMA database were merged with the 20 pairs found in the FDA database to provide us with a final sample size of 27 unique pairs of drugs after duplicates were removed (Fig. 1). The most commonly approved drugs in our sample were antineoplastic drugs ($n = 6$) and anti-infective agents ($n = 5$).

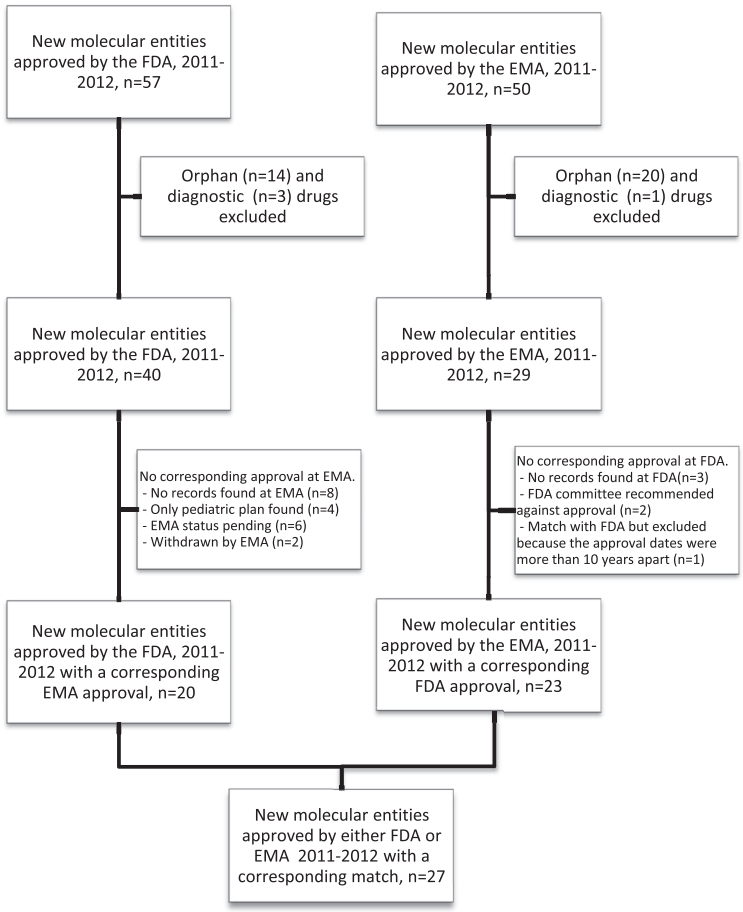


Fig. 1. Flowchart of included drugs. EMA, European Medicines Agency; FDA, Food and Drug Administration.

3.2. Accessibility and comprehensiveness of documents

As summarized in Table 1, the FDA medical review contained substantially more pages than the EMA reports. The FDA reports are structured, but there are often several versions and updates, whereas the EMA releases one final document. Only the EMA reports included a table of contents at the beginning of every document. The FDA reports' typed page numbers did not match the electronic page numbers. Eight FDA reports were not fully searchable, the least accessible ones being the oldest reports.

In the FDA documents, all censored text was obscured, and the number of redacted pages was specified. The reason for each redaction was specified; most frequently because it was considered a trade secret and on a few occasions to prevent investigators and patients from being identified. Any indications that were not approved were redacted in FDA reports. EMA reports included a statement on the first page mentioning that all information of commercial nature was deleted, but it was not possible to see how much information, if any, had been redacted.

Table 1 also summarizes that, although neither of the agencies provides reports in non-English languages, the EMA provides a lay summary which is available in several languages. Letters from the FDA to the applicant are accessible, but communications from the applicants to the FDA are not. Reports from different departments and senior personnel's assessment of reports are only available at the FDA. Neither of the agencies provides full trial reports or protocols on their Web sites.

The FDA conducted and included in their reports additional statistical analyses, whereas the EMA did not. In a few cases, the FDA also acquired case report forms (individual patient data) from the applicants.

Nonapproved or withdrawn drug reports are not available on the FDA Web site; however, the EMA provides such

Table 1. Characteristics of regulatory reports available for 27 new drugs approved between 2011 and 2012

Characteristic	FDA	EMA
Median number of pages (range)[a]	219 (70–602)	88 (37–133)
Table of contents, % (*n*)	70 (19)[b]	100 (27)
Material searchable, % (*n*)	70 (19)	100 (27)
File partially redacted, % (*n*)	100 (27)	0 (0)[c]
Reasons for redaction indicated, % (*n*)	96 (26)	100 (27)
Available in several languages, %	0	0
Lay summaries provided, %	0	100
Communication between regulator and applicant, %	100[d]	0
Full trial reports available, %	0	0
Trial protocols available, %	0	0
Direct links to relevant reports can be saved, %	0	100

Abbreviations: EMA, European Medicines Agency; FDA, Food and Drug Administration.

[a] For FDA, only the pages from the medical review file were calculated. Sometimes, this file contained communications and several versions. The corresponding reports from the EMA also contained information about pharmacology. *P*-value <0.05 for the Student *t*-test.

[b] The FDA documents did not contain a table for contents for the entire document, but they were assessed as acceptable if they had a table of content for the majority of the document.

[c] It was unclear for all EMA reports whether anything was redacted because it was a prepared document for the public. On the first page, it said that commercial information had been removed but it was not possible to see how much data (if any) had been removed.

[d] Redacted only.

Table 2. Characteristics of trials, efficacy, and harms data available for 27 new drugs approved between 2011 and 2012

Characteristic	FDA, % (*n*)	EMA, % (*n*)
ClinicalTrials.gov ID	0 (0)	0 (0)
Summary trial reports of pivotal trials	100 (27)	100 (27)
Names of trial investigators	0 (0)	0 (0)
Patient population specified (inclusion and exclusion criteria)	96 (26)	96 (26)
Intervention and comparison group specified	89 (24)	93 (25)
Outcomes (endpoints) specified	96 (26)	100 (27)
Total number of patients given	100 (27)	100 (27)
Number of patients for each arm	100 (27)	100 (27)
Numerical results provided—efficacy	100 (27)	100 (27)
Individual trial data provided	100 (27)	100 (27)
Results can be used in a meta-analysis	100 (27)	96 (26)
Harms		
Table of common adverse events	96 (26)	67 (18)
Missing trials in safety pool (from other indications)	4 (1)	0 (0)
All important harms reported	93 (25)	26 (7)
Numerical results provided—harms	100 (27)	100 (27)
Risk management plan and/or pharmacovigilance	48 (13)	100 (27)
Further trials and/or studies required	78 (21)	48 (13)[a]
Follow-up existing trials	22 (6)	22 (6)
Labeling restriction	4 (1)	100 (27)
REMS and/or educational material	30 (8)	26 (7)

Abbreviations: EMA, European Medicines Agency; FDA, Food and Drug Administration; REMS, Risk Evaluation and Mitigation Strategies.

[a] For additional 10 reports, it was unclear whether further trials were required or whether the company had voluntarily initiated them.

information along with reasons for withdrawal or nonapproval.

3.3. Trial characteristics and efficacy data

All reports contained an overview of the pivotal trials with internal trial IDs, but they also listed additional trials and sometimes what appeared to the whole trial program with up to 60 trials. This overview was usually in tabular form and often included study design, intervention, and location. However, as summarized in Table 2, none of the trial descriptions contained the corresponding ClinicalTrial.gov IDs. None of the reports listed all trial investigator names, although three FDA reports included names of a few investigators whose sites had been subject to investigation. The FDA had a "financial disclosures" section which, in 17 of the FDA reports, contained relevant information on investigator conflicts of interest and in which the impact on the trials' outcomes was briefly discussed.

Table 2 shows that the patient population, the intervention and comparator groups, and the outcomes were described by both agencies for most drugs.

Table 2 shows that only one EMA report lacked numerical efficacy data and therefore was not suitable for use in a meta-analysis. Standard deviations were missing for individual trials.

3.4. Harms data

Nine (33%) EMA reports had no table of common adverse events, whereas only one (4%) FDA report did not have this table (Table 2). Only 26% (7 of 27) of the EMA reports, compared to 93% of the FDA reports, reported all important information about harms. Compared with the FDA, the EMA reports contained less summary data on harms (Table 2). The risk management plan was always clearly stated in the EMA reports, whereas the FDA relayed the requirements to industry in the approval letter. The FDA required additional studies more frequently than the EMA (Table 2). However, it was primarily the older EMA reports that did not require additional studies. The EMA reports always had clearly stated corrections to the label, whereas such labeling recommendations were either redacted or difficult to find in the FDA reports. REMS and educational materials were equally often required by the two agencies.

4. Discussion

Drug reviews from the FDA and the EMA can be great sources of information for clinicians and researchers conducting meta-analyses. Trial methodology was described, and detailed summary harms data were available in the FDA reports. The FDA provides multiple reports including statistical analysis, assessments of financial

conflicts of interest, and sometimes subgroup analysis of North American patients. The EMA provides a single report with less information, but it is easier to access and navigate compared with the FDA. Methods for navigating drug reports are being developed and could ease handling in the future [13]. Both data sources, but more frequently the FDA, presented a number of challenges including redacted information, internal discrepancies, lack of standardization of reporting, voluminous pages, and documents that are not fully searchable.

Crucial information about safety concerns and nonapproved indications were redacted in the FDA reports. EMA does not redact descriptions of nonapproved indications and provides full reports for drugs that were not approved at all. These are two good reasons for also searching EMA. Furthermore, the correspondence between agencies and companies, which is often redacted, is not merely a simple exchange of facts. Some companies see it as strategic negotiation which they want to be held confidential from the public although it might be concerning serious harms [14].

FDA provided information about conflicts of interest, but only researchers receiving over $25,000/yr are required to disclose and financial disclosure statements were in some cases only collected from less than 50% of the investigators.

Neither the FDA nor the EMA documents make it easy to use PubMed or other electronic databases to identify publications that might have resulted from the trials mentioned in the reviews. Investigator names are usually not available, the number of trials done can be difficult to determine, and ClinicalTrials.gov or other trial registry identifiers are not referenced. ClinicalTrials.gov is another source of unpublished data [15].

A possible explanation for the discrepancies in information available is that the two agencies differ substantially in their structure and the laws governing their functions [16–19]. The documents produced by the agencies are not intended for research synthesis, and this is another important reason why information that a reviewer would find essential might not be available. Full clinical study reports which are included in the companies' approval applications would provide an even better understanding of drugs, but they are currently not readily available and require considerable time and resources to extract useful data [20].

4.1. Limitations

Some of our assessments involved some amount of subjectivity, but we attempted to minimize this limitation by having two coders. In documents that were not searchable, we could have overlooked information because of the large number of pages. Although our sample was small, it was a comprehensive sample of new drugs that were recently approved by both the agencies through the standard approval process. We have excluded orphan drugs and biologics which do not go through the standard approval process, so our results cannot be extrapolated to these drugs.

5. Conclusion

The FDA and EMA summary reports contained data that can be used for clinical decision making and meta-analysis. Data on harms are more detailed in the FDA reports, whereas the EMA reports are easier to use and sufficient for the collection of efficacy data.

Acknowledgments

We would like to thank Steffen Thirstrup for providing information about the European approval process and Roberta Scherer for providing comments on the paper.

Author contributions: L.B. and J.B.S. developed the protocol, which was revised by M.A.S. Data extraction was done by J.B.S. and M.A.S. Data analysis was prepared by J.B.S. who also wrote the first draft, which was revised and approved by L.B. and M.A.S. J.B.S. takes responsibility for the integrity of the data and the accuracy of the data analysis.

Ethics approval: In this study, we reviewed redacted publically available FDA and EMA approval documents, and ethical committee approval was not required.

Appendix

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.jclinepi.2014.06.019.

References

[1] Chan A-W, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. JAMA J Am Med Assoc 2004;291:2457–65.

[2] Vedula SS, Bero L, Scherer RW, Dickersin K. Outcome reporting in industry-sponsored trials of gabapentin for off-label use. N Engl J Med 2009;361:1963–71.

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

[4] MacLean CH, Morton SC, Ofman JJ, Roth EA, Shekelle PG, Southern California Evidence-Based Practice Center. How useful are unpublished data from the Food and Drug Administration in meta-analysis? J Clin Epidemiol 2003;56:44–51.

[5] McDonagh MS, Peterson K, Balshem H, Helfand M. US Food and Drug Administration documents can provide unpublished evidence relevant to systematic reviews. J Clin Epidemiol 2013;66:1071–81.

[6] Hart B, Lundh A, Bero L. Effect of reporting bias on meta-analyses of drug trials: reanalysis of meta-analyses. BMJ 2012;344:d7202.

[7] Rising K, Bacchetti P, Bero L. Reporting bias in drug trials submitted to the Food and Drug Administration: review of publication and presentation. Plos Med 2008;5:e217. discussion e217.
[8] Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. N Engl J Med 2008;358:252–60.
[9] Schroll JB, Bero L, Gøtzsche PC. Searching for unpublished data for Cochrane reviews: cross sectional study. BMJ 2013;346:f2231.
[10] Wolfe N, Gøtzsche PC, Bero L. Strategies for obtaining unpublished drug trial data: a qualitative interview study. Syst Rev 2013;2:31.
[11] Turner EH. How to access and process FDA drug approval packages for use in research. BMJ 2013;347:f5992.
[12] Downing NS, Aminawung JA, Shah ND, Braunstein JB, Krumholz HM, Ross JS. Regulatory review of novel therapeutics—comparison of three regulatory agencies. N Engl J Med 2012;366:2284–93.
[13] Jefferson T, Jones MA, Doshi P, Del Mar CB, Hama R, Thompson MJ, et al. Neuraminidase inhibitors for preventing and treating influenza in healthy adults and children. Cochrane Database Syst Rev 2014;4:CD008965.
[14] Hawkes N. Industry and drug regulators disagree on which data should remain confidential. BMJ 2013;347:f5390.
[15] Becker JE, Krumholz HM, Ben-Josef G, Ross JS. Reporting of results in ClinicalTrials.gov and high-impact journals. JAMA J Am Med Assoc 2014;311:1063–5.
[16] Senderowicz AM, Pfaff O. Similarities and differences in the oncology drug approval process between FDA and European Union with emphasis on in vitro companion diagnostics. Clin Cancer Res 2014;20:1445–52.
[17] Tafuri G, Stolk P, Trotta F, Putzeist M, Leufkens HG, Laing RO, et al. How do the EMA and FDA decide which anticancer drugs make it to the market? A comparative qualitative study on decision makers' views. Ann Oncol 2014;25:265–9.
[18] DeMuro C, Clark M, Doward L, Evans E, Mordin M, Gnanasakthy A. Assessment of PRO label claims granted by the FDA as compared to the EMA (2006–2010). Value Health 2013;16:1150–5.
[19] Doua JY, Geertruyden J-PV. Registering medicines for low-income countries: how suitable are the stringent review procedures of the World Health Organisation, the US Food and Drug Administration and the European Medicines Agency? Trop Med Int Health 2014;19:23–36.
[20] Doshi P, Jefferson T. Clinical study reports of randomised controlled trials: an exploratory review of previously confidential industry reports. BMJ Open 2013;3:e002496.

Assessment of harms in protocols, clinical study reports and published papers of trials of orlistat

Jeppe Bennekou Schroll, MD

Elisabeth I. Penninga, MD

Peter C. Gøtzsche, Professor

The Nordic Cochrane Centre, Rigshospitalet, Dept 7811, Blegdamsvej 9, DK-2100 Copenhagen Ø, Denmark.

Corresponding author:

Words: 4037

Keywords: orlistat, adverse drug event, drug approval, unpublished data,

Jeppe Bennekou Schroll

The Nordic Cochrane Centre, Rigshospitalet, Dept 7811,

Blegdamsvej 9, DK-2100 Copenhagen Ø, Denmark.

E-mail: js@cochrane.dk

Tel: +45 35 45 71 12

Fax: +45 35 45 70 07

Abstract

Objective

To study how adverse events for orlistat were handled in trial protocols, clinical study reports (CSR) and published papers.

Design

Information about adverse events was extracted from each source. Individual events from one trial were studied in detail in an exploratory analysis.

Setting

Seven randomised placebo controlled trials of orlistat (4,225 patients) submitted by Roche to the European Medicines Agency (EMA) for marketing approval.

Data sources

CSRs, including protocols, obtained from the EMA in 2011 (8,716 pages). Corresponding journal articles were identified on PubMed.

Results

The protocols contained scant information about how adverse events were to be handled and no guidance was given on how to question the patients. There were nonetheless important discrepancies between the data sources. We found many post hoc changes that favoured orlistat, but were not described in the published reports. For example, it was stated in the CSRs that some gastrointestinal adverse events (which we would expect with orlistat) should only be coded if considered "bothersome" to the patient. Another post hoc change was that serious adverse events should be assessed for relationship to drug by the sponsor. All seven CSRs introduced a new primary outcome for quality of life, which was different from the one specified in the protocol, and the only trial that published any data was the one with the largest effect. We also suspected that the decisions and explanations for withdrawing patients were biased. Asymptomatic patients with hyperglycaemia above a specific threshold were, for instance, withdrawn and inappropriately categorised as "adverse events" in one trial.

In the exploratory analysis, we identified many more adverse events than those listed in the summaries in the CSR and discovered, which the CSR did not explain, that multiple episodes were only counted once. We also discovered that the patients had twice as many days with adverse events on orlistat than on placebo and that the events were more severe on orlistat. None of this was stated in the CSR or in the published trial report.

Conclusion

The information in the published papers on the harms of orlistat is seriously misleading. Clinical study reports, including the protocol and individual patient data, should be the primary data source for systematic reviews of drugs.

Introduction

Randomised trials generally underreport harms. In 14% of 185 randomised trials published in major medical journals, adverse reactions were not mentioned at all, and in 32% they were not shown for each arm or general statements were used.[1] Only 16% of the trial reports described how adverse events were identified,[1] which is also problematic because the way the investigator obtains information impact greatly on the number[2] and characteristics of the events.[3] Another survey found that only 18% of 107 trials in children reported harms data adequately according to the CONSORT guidelines.[4]

In systematic reviews of harms, the search strategies are generally inadequate and are not reported in detail,[5] and although it is standard to assess the risk of bias in trials included in reviews and state sources of funding, both were done in less than half of a sample of meta-analyses of adverse events.[6]

Industry-sponsored trials are more likely than other trials to conclude that a drug is safe.[7] A similar bias exists in industry-supported reviews of drugs, which are less transparent, have fewer reservations about methodological limitations of the included trials, and have more favourable conclusions than Cochrane reviews of the same drugs.[8]

Selective reporting of harms can be deliberate and it may have disastrous consequences. Merck concealed cases of myocardial infarction and deaths with rofecoxib in pivotal trials.[9,10] Pfizer denied that celecoxib causes heart attacks at an FDA hearing in 2005, despite having unpublished evidence to the contrary[11]. In 2009, they called the evidence "inconclusive" in information to patients invited to participate in a clinical trial.[12] It is estimated that rofecoxib has killed 120,000 people and that celecoxib has killed 75,000 till 2004.[13]

Many steps, decisions and assumptions precede the reporting of an adverse event or the decision to ignore it. Adverse events are coded by the sponsor, which is a highly bias-prone process, as it is rarely blinded. We recently found that no reliable interobserver study of coding has ever been conducted, and that modern coding systems might have made statistical detection of adverse events more difficult because of splitting similar events into several categories.[14]

Except for orlistat, all slimming pills have been withdrawn in Europe because of harms. Aminoxaphen and the fenfluramines were withdrawn because of pulmonary hypertension and cardiovascular death,[15] rimonabant because of psychiatric events,[16] and sibutramine because of myocardial infarction.[17] In 2011, the FDA issued a warning regarding orlistat based on 12 cases of liver failure,[18] and in 2012, the EMA refused to approve a combination of phentermine and topiramate because of adverse effects.[19]

In 2011, we got access to the clinical study reports (CSRs) and their corresponding trial protocols of the placebo controlled trials submitted to the EMA by Roche for obtaining marketing authorisation for its anti-obesity drug, orlistat.[20] The CSRs include individual patient data with narrative descriptions of adverse events. We have used these unique data to study how adverse events and methods for obtaining them were reported in protocols, CSRs and published papers.

Methods

Seven placebo-controlled randomised trials of orlistat were included in the application for marketing authorisation. The CSRs consisted of 8,716 pages in total and included 4,225 patients. They contained the full trial protocols, efficacy data and an overview of adverse events by organ system, data on day of onset and intensity of adverse events for all patients, and detailed narrative descriptions of serious adverse events and events leading to withdrawal from the study.

For the protocols, two investigators (EP and JBS) independently extracted names of authors, withdrawal criteria, coding strategies and information about how adverse events were planned to be handled. Post hoc we added strategies for handling vitamin deficiency (as orlistat decreases absorption of fat from the gut, it might affect the absorption of fat soluble vitamins) and measures of quality of life (which can potentially uncover harms).

For the CSRs, the same investigators noted identifiers such as investigator names, start and end dates, treatment duration and countries, and extracted the following data: From the synopsis, all information about adverse events; from the methods section, information about withdrawal, harms and quality of life; from the results section, the overview of adverse events and, for each treatment arm, number of patients, mean age, mean BMI, gender, patients with withdrawal, adverse events, serious adverse events, gastrointestinal adverse events, deaths, quality of life scores and low vitamin levels; and from the discussion section and the conclusion, all text describing adverse events.

We searched PubMed with “orlistat or Xenical” to find the corresponding publications. The search returned 1433 hits, from which we collected 35 articles as full text. We identified nine papers that described the seven trials individually,[21–29] and seven papers with pooled estimates from the trials that did not contain any data about adverse events.[30–36] Each trial had a detailed primary publication[22-28] and we extracted all information about adverse events from these.

In an exploratory pilot study, we converted all individual patient adverse event listings from one trial (Trial 7) by using text recognition software (ABBYY FineReader 10) and transferred the data to Excel. Trial 7 was chosen because it was the newest and also one of the smallest and had a relatively simple design.

We compared the CSRs with the protocols and the publications to determine if the summaries in the CSRs and published papers reflected the adverse events in an unbiased way.

Results

Trial design

The protocols described 7 phase III randomised trials that all had a placebo arm (Table 1). Orlistat was investigated in regimes of 30 mg, 60 mg and 120 mg t.i.d. Patients and investigators were blinded in all trials. All clinical studies lasted between 52 and 104 weeks. Trial 2 randomised the patients again to either placebo or orlistat after 52 weeks of treatment and trial 5 changed the intensity of treatment after 52 weeks for half the patients. Patients and doctors were blinded to the treatment but whether the coders of adverse events were blinded was not mentioned in any of the documents.

The trials were conducted between 1992 and 1996 in USA and Europe. They all had a “lead-in” period where the patients got placebo along with dietary advice, which mostly lasted for 4-5 weeks. Some patients were excluded based on their performance in this period. The included patients had a

BMI between 28 and 43. Trial 7 included only patients with type II diabetes whereas the other trials excluded such patients.

Protocols
All protocols mentioned that vital signs, adverse events, routine laboratory tests, fat soluble vitamin levels and ECG would be recorded. All protocols had at least eight withdrawal criteria. Apart from "new smokers", which was an additional criterion in five protocols, the withdrawal criteria were the same. They were very vague, as "Administrative reasons" or "other reasons" were sufficient for withdrawal.

Three protocols (trials 1-3) contained an appendix on how to code gastrointestinal adverse events (all events were in American English; for consistency, we have used British English). The investigators were encouraged not to use the term "diarrhoea" as it could cause "misunderstandings". The patients' gastrointestinal symptoms were divided into: "increased defaecation", "liquid stools", "soft stools" (which were "a rather fluid consistency"), "fatty/oily evacuations", "oily spotting", "faecal urgency", "faecal incontinence", "flatus with discharge", "decreased defaecation", "pellets" and "solid stools". If an investigator used the term "diarrhoea", it was recategorised as "liquid stools". This was not mentioned in the protocol but we found out by analysing the tables. Even though protocols 4-7 did not contain this appendix, the tables of coded terms and investigator terms showed that it was still being used.

The patients had between 9 and 17 visits during the first year and adverse events were recorded at each visit on the case report forms. Only a change from the patients' pre-treatment condition was considered an adverse event, and the protocols provided no guidance on how to question the patients. The investigator related the severity to daily function and also judged the relationship to treatment (two appendices offered guidance on this). Protocols 4-7 specified that recorded adverse events from the lead-in phase should be called "complaints".

For quality of life, six protocols specified that the main outcome was "comparative rates of change" for the subscales "health distress and emotional functioning." The scales were not specified in any of the protocols; instead, they referred to a questionnaire, which was a 46-item list divided into 7 groups with no information about how the subscales were to be combined. There were also secondary quality of life outcomes, which were vaguely described as "a variety of scales."

The only information on the statistical handling of adverse events was that the treatment groups would be compared using "descriptive statistics."

Clinical study reports - methods
Like in the protocols, it was not specified how the patients had been questioned about adverse events. What was new compared to the protocols was that gastrointestinal adverse events should be categorised according to a predefined list where some of the terms were accompanied by a star. The unstarred terms should only be considered adverse events when "described as bothersome by the patient" and these included "fatty/oily stool", "liquid stools" (which term the protocol suggested to be used instead of diarrhoea), "increased defaecation", "stools soft", "decreased defaecation" and "pellets". "Bothersome" was not a requirement for adverse events outside the gastrointestinal category and was not mentioned in the protocols. Furthermore, serious adverse events had been assessed for relationship to drug by the sponsor, although this was not prespecified in the protocol.

The adverse events were coded according to a Ciba-Geigy modified WHO glossary, which could be updated by the sponsor. For each adverse event described by the investigator, the sponsor would assign a preferred term from the dictionary.

All the method sections described that adverse events would be presented as listings and summary tables by body system, intensity and relation to drug. For gastrointestinal problems, however, only events more frequent than 1% in four trials, and 3% in three trials would be summarised.

All the method sections noted that the "primary measure" for quality of life was "overweight distress", "depression" and "satisfaction with treatment". We could not find any explanation in the CSRs or in amendments to these as to why the primary outcome for quality of life from the protocol had been changed from "health distress and emotional functioning".

Clinical study reports – results

All CSRs narratively acknowledged that there were many adverse events but also noted that the differences between placebo and active treatment were small, and two CSRs noted that most adverse events were considered unrelated to the drug by the investigator. Only one CSR mentioned the total number of patients with one or more adverse events in the results section. None of the reports mentioned the total number of events where the difference was considerably higher.

The increased number of gastrointestinal adverse events observed in the orlistat group was mentioned but it was emphasised that this was due to the pharmacological effect of the drug, as if that made any difference. It was noted that the numbers of gastrointestinal adverse events per patient were often few (1 to 2), and there was no information on their duration in the main text.

More patients in the treatment group were withdrawn due to adverse events whereas more patients in the placebo group were withdrawn for "any reason". We suspect that the decisions and explanations were biased. For example, in trial 2, more patients "lost to follow-up" were withdrawn from the placebo group (22 vs. 12) and also more patients who "did not cooperate" (26 vs. 13). In trial 4, more placebo patients were excluded due to "administrative reasons" (29 vs. 10 during the first year). See also Trial 7 below.

Many patients receiving orlistat had low vitamin measurements, even in a trial where everyone received a multivitamin tablet.

Publications

A brief summary of the papers describing the seven clinical trials are listed in Table 2.[21–27] There were between 71 and 270 times as many pages in the CSRs as in the corresponding publications. Six papers described that "all adverse events were recorded," and one noted that the Ciba-Geigy dictionary was used.

Five papers mentioned that a special dictionary was developed for the expected gastrointestinal adverse events, but none described that only "bothersome" adverse events should be recorded and none described that "diarrhoea" was discouraged as a term.

All papers had severe restrictions on which adverse events were reported and only four papers presented a table summarising adverse events. Two papers censored all events that had been considered "unrelated" by the sponsor and only reported events occurring in 3% or 5% of patients.

One paper censored both “unrelated” and “remotely related” events. Three papers reported only adverse events that were twice as frequent in the orlistat group as in the placebo group, and two of them had the additional criterion that only events occurring in at least 5% of the patients would be reported. These two papers only reported the adverse event rate for the orlistat group.

For four trials, we could extract data on the number of adverse events, and between 3% and 33% of those reported in the CRSs were also reported in the publications. However, the true percentage is lower, as the grand total in the CSRs was also too low (see trial 7 below).

Only trial 3, which had the biggest difference between placebo and orlistat, reported on quality of life, but there were no data in the paper, only p-values.

Trial 1 lumped the gastrointestinal adverse events into two new main categories: “Uncontrolled oily discharge,” which included faecal incontinence, flatus with discharge and oily spotting, and “loose stools,” which included oily evacuation, fatty/oily stool, liquid stools and soft stools.

Trial 7, patients with type II diabetes
Almost all patients experienced one or more adverse events (157 patients (96%) in the orlistat groups and 150 (94%) in the placebo group). A total of 3,446 adverse events were listed (2,008 in the orlistat group and 1,438 in the placebo group). These numbers could not be found in any of the summaries in the CSR or in the publication, and more events were missing for orlistat than for placebo: In an appendix in the CSR, the total was 1,198 for orlistat (60%) and 930 for placebo (65%). We discovered that multiple episodes were only counted once; this was not explained in the CSR. We calculated that each patient had 12.8 adverse events, on average, in the orlistat group and 9.6 in the placebo group, or 3.2 (95% CI: 1.2-5.2) more adverse events in the orlistat group. This was not mentioned in the report or publication.

The duration of each adverse event was carefully recorded but was not summarised neither in the CSR nor in the publication. We calculated that the average duration was 22.5 days (95% CI: 20.0-25.1) in the orlistat group and 14.7 days (12.9-16.6) in the placebo group and that the number of days each person was affected by an adverse event was 288 days in the orlistat group and 141 days in the placebo group. Thus, on average, orlistat led to double as many days with adverse events as placebo did.

The CSR noted that most adverse events were mild to moderate in intensity. However, we found that the events were more severe in the orlistat group ($p<0.001$, chi^2 test, not adjusted for dependent observations), which was not mentioned in the CSR or the publication. The RR for having a mild adverse event in the orlistat group compared to the placebo group was 0.93 (95% CI: 0.89-0.96), a moderate event 1.29 (95% CI: 1.13-1.48) and a severe event 1.39 (95% CI: 0.75-2.59).

More placebo patients were withdrawn due to adverse events but 14 of the 23 withdrawn patients in the placebo group were discontinued due to abnormal fasting glucose. The protocol stated that fasting glucose above 220 mg/dl would lead to discontinuation, but it seems inappropriate to code this as an adverse event. An additional sign that this was inappropriate was that only 2 of the 14 withdrawals were listed as an adverse event in the detailed list of adverse events for each patient. Furthermore, a baseline imbalance could perhaps partly explain the difference (HbA1c was 8.05 in the active group and 8.20 in the placebo group).

Since the company seemed to assume that an inert placebo can cause hyperglycaemia, we decided to look for the opposite effect, namely if orlistat could cause hypoglycaemia, which is more plausible due to the weight loss. There was scant information about this in the CSR. In the first quarter of the trial, 14% of patients on orlistat had a hypoglycaemic episode versus 10% on placebo. In the second quarter, the rates were 12% and 6%, respectively. The CSR referred to an appendix, but this was missing. We found 426 hypoglycaemia events in the orlistat group and 300 in the placebo group and an average of 2.7 events per patient in the orlistat group and 2.0 in the placebo group (p=0.10, unpaired t-test).

Discussion

The reporting of harms in the orlistat trials was deceptive and we identified many manoeuvres, both pre and post hoc, that contributed to concealing the true nature, severity and duration of adverse effects caused by orlistat. The trial protocols contained very little information on how adverse events were planned to be collected, handled and presented. Specific coding guidelines encouraging investigators only to code certain complaints if they were "bothersome" were likely post hoc decisions, as they were not mentioned in the protocols but only in the CSRs. Two papers censored all events that had been considered "unrelated" by the investigator, and the sponsor decided for all trials whether serious adverse events were related to the drug. None of this was mentioned in the protocols and it was also not mentioned whether the sponsor was blinded. It is difficult to imagine a larger conflict of interest than to let the sponsor decide whether an adverse event is caused by the sponsor's own drug.

The protocols explicitly discouraged the use of the term "diarrhoea" claiming it was not well defined, and gastrointestinal events that more or less expressed the same thing were split into several categories, which can decrease the power of a study to identify adverse events.[14]

The quality of life subscales changed from the protocol to the CSRs without any explanation or even an acknowledgement that this had been done. The composition of the subscales was obscure and many questions were closely related to the assumed effect of treatment rather than the quality of life. Only one paper reported on quality of life even though all trials collected and analysed this important outcome, and that trial was the one with the largest effect.

All publications reported on adverse events but only a fraction of them were reported due to various censoring filters, none of which were predefined. Other studies have also found that only a fraction of adverse events were reported in published papers compared to the summaries in CSRs.[37,38]

Our in-depth analyses of trial 7 revealed that considering the duration of an adverse event can change the perception of harms dramatically. We found that a patient on orlistat will experience double as many days with adverse events as a patient on placebo, which tells a very different story than the CSR does: "The total percentage of patients with adverse events was large for both treatment groups, but adverse event profile differences between orlistat- and placebo-treated patients were small or non-existent, except for adverse events of the gastrointestinal system". In the ICH guideline about CSRs, an analysis of duration is optional,[39] but perhaps it should not be. It was bizarre that patients with a blood glucose level above an arbitrary level were not only withdrawn from the trial but were also called patients with an adverse event. This was likely a post hoc decision, as the protocol did not specify that hyperglycaemia should be categorised as an adverse event.

All trials had a long lead-in period on placebo, which is also of concern. Adverse events were only registered if there was a worsening compared to this baseline. In trial 6, there was a 24-week lead-in during which more than 90% of the patients reported at least one adverse event. Since gastrointestinal complaints are normal in healthy people, this type of censoring might have made it more difficult to detect gastrointestinal adverse events caused by orlistat.

The harms reported in CSRs and papers were, despite all the censoring manoeuvres, reported in a way that downplayed them, e.g. with sweeping statements that most of the adverse events were considered unrelated to the drug and that they were generally mild to moderate, although the data showed that they were more severe with orlistat.

Since the published trial reports of orlistat are seriously misleading, it is worthwhile to analyse observational studies as well. Slimming pills are often discontinued by the patients.[40] A Canadian study of 16,968 patients on orlistat showed that after one year, only 6% of the patients were still taking the drug, and after two years, it was only 2%.[40] This reflects not only the high price of orlistat but also its harms and its poor effect. In published reports, that suffer from publication bias and analytic biases, such as using the last observation carried forward,[41] the effect is only a 3% decrease in body weight.[40]

Limitations

Our study was explorative and restricted to one drug tested in the mid-1990s; our results might therefore not be applicable for newer drugs. Performing text recognition of individual adverse events is labour intensive.

Conclusion

The protocols, CSRs and publications all reported poorly on how adverse events were planned to be collected, handled and analysed. Censoring filters and decisions that were not prespecified, were introduced post hoc, and the guidance on how to code adverse events differed between protocols and CSRs and was absent in publications. The duration of the adverse events was not included in any of the analyses conducted by the company even though the difference between orlistat and placebo was large, and their severity was downplayed in several different ways. Clinical study reports, including the protocol and individual patient data, should be the primary data source for systematic reviews of drugs.

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

Authors' contributions
Authors' contributions: PCG came up with the idea and JBS developed the protocol in cooperation with PCG. Data extraction was done by JBS and EP. Data analysis was prepared by JBS who also wrote the first draft which was revised and approved by PCG and EP. JBS takes responsibility for the integrity of the data and the accuracy of the data analysis.

Acknowledgements
None

Funding
This project was funded The Danish Council for Independent research. Grant ID 12-126864. The funder had no influence on the study design, interpretation of data, or the decision to publish the results.

Data sharing:
Anonymised data sets are available on request from the corresponding author: js@cochrane.dk

What this paper adds

What is already known on this subject

- Selective reporting of benefits and harms data is common.
- Only a fraction of adverse events from clinical study reports are reported in publications.

What this study adds

- Individual patient data, for instance from clinical study reports, are necessary for a comprehensible overview of harms.
- Important coding and reporting guidelines were not mentioned in the published papers and many post hoc changes were introduced that favoured the drug over placebo.
- Quality of life was measured in all trials but only published for the one that showed the largest effect, and the primary outcome was changed in all trials without any explanation.

- Incorporating the duration of the adverse events changed the perception of the harms of orlistat dramatically.
- The published papers are seriously unreliable.

References

1. Loke YK, Derry S. Reporting of adverse drug reactions in randomised controlled trials - a systematic survey. *BMC Clin Pharmacol*. 2001;1:3.

2. Olsen H, Klemetsrud T, Stokke HP, Tretli S, Westheim A. Adverse drug reactions in current antihypertensive therapy: a general practice survey of 2586 patients in Norway. *Blood Press*. 1999;8(2):94-101.

3. Ioannidis JPA, Evans SJW, Gøtzsche PC, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med*. 2004;141(10):781-788.

4. De Vries TW, van Roon EN. Low quality of reporting adverse drug reactions in paediatric randomised controlled trials. *Arch Dis Child*. 2010;95(12):1023-1026. doi:10.1136/adc.2009.175562.

5. Golder S, Loke Y, McIntosh HM. Poor reporting and inadequate searches were apparent in systematic reviews of adverse effects. *J Clin Epidemiol*. 2008;61(5):440-448. doi:10.1016/j.jclinepi.2007.06.005.

6. Golder S, Loke Y, McIntosh HM. Room for improvement? A survey of the methods used in systematic reviews of adverse effects. *BMC Med Res Methodol*. 2006;6:3. doi:10.1186/1471-2288-6-3.

7. Golder S, Loke YK. Is there evidence for biased reporting of published adverse effects data in pharmaceutical industry-funded studies? *Br J Clin Pharmacol*. 2008;66(6):767-773. doi:10.1111/j.1365-2125.2008.03272.x.

8. 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(7572):782. doi:10.1136/bmj.38973.444699.0B.

9. Mukherjee D, Nissen SE, Topol EJ. Risk of cardiovascular events associated with selective COX-2 inhibitors. *JAMA J Am Med Assoc*. 2001;286(8):954-959.

10. Curfman GD, Morrissey S, Drazen JM. Expression of concern reaffirmed. *N Engl J Med*. 2006;354(11):1193. doi:10.1056/NEJMe068054.

11. Harris G. Pfizer says internal studies show no Celebrex risks. *New York Times*. http://www.nytimes.com/2005/02/05/business/05drug.html?_r=0. Published May 2, 2005. Accessed June 13, 2014.

12. De Andrade M. Research Ethics . In clear sight. *BMJ*. 2009;339:b3443.

13. Gøtzsche PC, Smith R, Rennie D. *Deadly Medicines and Organised Crime: How Big Pharma Has Corrupted Healthcare*. 1 edition. London: RADCLIFFE MEDICAL PRESS LTD; 2013.

14. Schroll JB, Maund E, Gøtzsche PC. Challenges in coding adverse events in clinical trials: a systematic review. *PloS One*. 2012;7(7):e41174. doi:10.1371/journal.pone.0041174.

15. Curfman GD. Diet pills redux. *N Engl J Med*. 1997;337(9):629-630. doi:10.1056/NEJM199708283370909.

16. The European Medicines Agency recommends suspension of the marketing authorisation of Acomplia. 2008. Available at: http://www.ema.europa.eu/docs/en_GB/document_library/Press_release/2009/11/WC500014774.pdf. Accessed July 3, 2014.

17. Curfman GD, Morrissey S, Drazen JM. Sibutramine--another flawed diet pill. *N Engl J Med*. 2010;363(10):972-974. doi:10.1056/NEJMe1007993.

18. FDA Drug Safety Communication: Completed safety review of Xenical/Alli (orlistat) and severe liver injury. 2010. Available at: http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm213038.htm. Accessed December 6, 2014.

19. Assessment report. Qsiva. 2013. Available at: http://www.ema.europa.eu/docs/en_GB/document_library/EPAR_-_Public_assessment_report/human/002350/WC500144300.pdf. Accessed June 21, 2014.

20. Gøtzsche PC, Jørgensen AW. Opening up data at the European Medicines Agency. *BMJ*. 2011;342:d2686.

21. Finer N, James WPT, Kopelman PG, Lean MEJ, Williams G. One-year treatment of obesity: a randomized, double-blind, placebo-controlled, multicentre study of orlistat, a gastrointestinal lipase inhibitor. *Int J Obes Relat Metab Disord*. 2000;24(3).

22. Sjöström L, Rissanen A, Andersen T, et al. Randomised placebo-controlled trial of orlistat for weight loss and prevention of weight regain in obese patients. European Multicentre Orlistat Study Group. *Lancet*. 1998;352(9123):167-172.

23. Rössner S, Sjöström L, Noack R, Meinders A, Noseda G. Weight loss, weight maintenance, and improved cardiovascular risk factors after 2 years treatment with orlistat for obesity. *Obes Res*. 2000;8(1):49–61.

24. Hauptman J, Lucas C, Boldrin MN, Collins H, Segal KR. Orlistat in the long-term treatment of obesity in primary care settings. *Arch Fam Med*. 2000;9(2):160.

25. Davidson MH, Hauptman J, DiGirolamo M, et al. Weight control and risk factor reduction in obese subjects treated for 2 years with orlistat: a randomized controlled trial. *JAMA J Am Med Assoc*. 1999;281(3):235-242.

26. Hill JO, Hauptman J, Anderson JW, et al. Orlistat, a lipase inhibitor, for weight maintenance after conventional dieting: a 1-y study. *Am J Clin Nutr*. 1999;69(6):1108–1116.

27. Hollander PA, Elbein SC, Hirsch IB, et al. Role of orlistat in the treatment of obese patients with type 2 diabetes: a 1-year randomized double-blind study. *Diabetes Care*. 1998;21(8):1288–1294.

28. James WP, Avenell A, Broom J, Whitehead J. A one-year trial to assess the value of orlistat in the management of obesity. *Int J Obes Relat Metab Disord J Int Assoc Study Obes*. 1997;21 Suppl 3:S24-30.

29. Karhunen L, Franssila-Kallunki A, Rissanen P, et al. Effect of orlistat treatment on body composition and resting energy expenditure during a two-year weight-reduction programme in obese Finns. *Int J Obes Relat Metab Disord J Int Assoc Study Obes*. 2000;24(12):1567-1572.

30. Zavoral JH. Treatment with orlistat reduces cardiovascular risk in obese patients. *J Hypertens*. 1998;16(12 Pt 2):2013-2017.

31. Zhi J, Mulligan TE, Hauptman JB. Long-term systemic exposure of orlistat, a lipase inhibitor, and its metabolites in obese patients. *J Clin Pharmacol*. 1999;39(1):41-46.

32. Heymsfield SB, Segal KR, Hauptman J, et al. Effects of weight loss with orlistat on glucose tolerance and progression to type 2 diabetes in obese adults. *Arch Intern Med*. 2000;160(9):1321-1326.

33. Rissanen A, Lean M, Rössner S, Segal KR, Sjöström L. Predictive value of early weight loss in obesity management with orlistat: an evidence-based assessment of prescribing guidelines. *Int J Obes*. 2003;27(1):103-109. doi:10.1038/sj.ijo.0802165.

34. Jacob S, Rabbia M, Meier MK, Hauptman J. Orlistat 120 mg improves glycaemic control in type 2 diabetic patients with or without concurrent weight loss. *Diabetes Obes Metab*. 2009;11(4):361-371. doi:10.1111/j.1463-1326.2008.00970.x.

35. Reaven G, Segal K, Hauptman J, Boldrin M, Lucas C. Effect of orlistat-assisted weight loss in decreasing coronary heart disease risk in patients with syndrome X. *Am J Cardiol*. 2001;87(7):827-831.

36. Morris M, Lane P, Lee K, Parks D. An Integrated Analysis of Liver Safety Data from Orlistat Clinical Trials. *Obes Facts*. 2012;5(4):485-494. doi:10.1159/000341589.

37. Maund E, Tendal B, Hrobjartsson A, et al. Benefits and harms in clinical trials of duloxetine for treatment of major depressive disorder: comparison of clinical study reports, trial registries, and publications. *BMJ*. 2014;348(jun04 2):g3510-g3510. doi:10.1136/bmj.g3510.

38. Wieseler B, Wolfram N, McGauran N, et al. Completeness of Reporting of Patient-Relevant Clinical Trial Outcomes: Comparison of Unpublished Clinical Study Reports with Publicly Available Data. Ghersi D, ed. *PLoS Med*. 2013;10(10):e1001526. doi:10.1371/journal.pmed.1001526.

39. STRUCTURE AND CONTENT OF CLINICAL STUDY REPORTS. E3. 1995. Available at: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E3/E3_Guideline.pdf. Accessed June 26, 2014.

40. Padwal R, Kezouh A, Levine M, Etminan M. Long-term persistence with orlistat and sibutramine in a population-based cohort. *Int J Obes 2005*. 2007;31(10):1567-1570. doi:10.1038/sj.ijo.0803631.

41. Jørgensen A. Robustness of results and conclusions in systematic reviews, trials and abstracts [PhD thesis]. 2011.

Tables

Table 1. Overview of trials included in the present study

Trial ID	Start/end date	Lead-in* /duration	Country	Patients	Treatment arms (1. year/2. year)	N, patients	Threshold for GI adverse events
1 - BM14119B	1 June 1992 - 25 August 1994	4 weeks/ 52weeks	UK	BMI 30-43	120mg TID Placebo	114 114	3%
2 - BM14119C	14 May 1992 - 9 October 1995	4 weeks/ 104weeks	Austria, Denmark, Finland, France, Germany, Netherlands, Sweden and Switzerland	BMI 30-43	120 mg TID /120 mg TID 120 mg TID /placebo placebo/ 120mg TID placebo/placebo	135 138 127 126	1%
3 - BM14149	May 3, 1993- February 15, 1996.	4 weeks/ 104 weeks	Austria, Finland, France, Germany, Netherlands, Sweden and Switzerland	BMI 28-43	60mg TID 120mg TID Placebo	242 244 243	1%
4 - NM14161	26 February 1993-1 December 1995	4 weeks/ 104 weeks	USA, primary care	BMI 30-43	60mg TID 120mg TID Placebo	214 214 214	1%
5 - NM14185	October 15, 1992- October 26, 1995	4 weeks/ 104 weeks	USA	BMI 31-43	120 mg TID /120 mg TID 120 mg TID /60 mg TID 120 mg TID/placebo placebo / placebo	153 152 138 133	1%
6 - NM14302	May 25, 1993- March 7, 1996	24 weeks/ 52 weeks	USA	BMI 28-38	30mg TID 60mg TID 120mg TID placebo	187 173 181 188	3%
7 - NM14336	December 21, 1993 - January 4, 1996	5 weeks/ 52 weeks	USA	BMI 28-40 per oral treated Type 2 Diabetics	120mg TID placebo	163 159	3%

* Lead-in was a period where both groups got placebo
TID: Three times a day
GI: Gastrointestinal

Table 2. Overview of publications included in the present study

	Citation	Time between completion and publication	Restriction of published adverse events	No Adverse event in CSR	No adverse events in published paper	Percentage of adverse events published	Compression factor*
Trial 1	Finer2000[21]	6 years	Adverse events considered "unrelated" by investigator were	661 534	220 112	33% orlistat 21% placebo	71

			censored. Only adverse events more common than 3% published				
Trial 2	Sjoström1998[22]	3 years	Adverse events considered "unrelated" by investigator censored. Only adverse events more common than 5% published	1511 1086	483 162	32% orlistat 15% placebo	254
Trial 3	Rossner2000[23]	4 years	"Common" adverse events reported	1097 1280 1087	164 208 38	15% orlistat 60mg 16% orlistat 120mg 3% placebo	88
Trial 4	Hauptman2000[24]	5 years	Only gastrointestinal adverse events that were considered possibly or probably related to treatment where incidence in active arm is twice that of placebo were reported.	1728 1737 1327	253 240 35	15% orlistat 60mg 14% orlistat 120mg 3% placebo	173
Trial 5	Davidson1999[25]	4 years	No table. Adverse events more frequent than 5% and more than twice as common in orlistat were reported (only orlistat arm events shown).	1359 1483 1387 1147	Adverse events not reported for placebo arm	Calculation not possible.	270
Trial 6	Hill1999[26]	3 years	No table. "Some gastrointestinal events occurred in a greater percentage of patients in the orlistat-treated groups" were reported in text but the total number of patients was unclear	1138 1083 1243 894	Adverse events not clearly reported	Calculation not possible.	128
Trial 7	Hollander1998[27]	2 years	No table. Adverse events more frequent than 5% and more than twice as common in orlistat were reported (only orlistat arm).	1198** 930	Adverse events not reported for placebo arm in publication	Calculation not possible.	113

* Pages in CSRs divided by pages in publication.
** During our study we discovered that the counts for trial 7 were too small and we expect that this is also the case for the remaining 6 trials.

GRADUATE SCHOOL OF HEALTH AND MEDICAL SCIENCES
UNIVERSITY OF COPENHAGEN

DECLARATION OF CO-AUTHORSHIP

About this form

- Declarations of co-authorship must be included in your PhD thesis if your thesis contains or is based on co-authored manuscripts or published articles.
- You must submit a declaration of co-authorship for all articles/manuscripts included.
- The declarations are important to the assessment committee's work as they must be able to identify and assess your contribution to the article(s).
- If you have more co-authors for an article than there is space for in the table, simply just complete another declaration for the same article with the remaining co-authors.
- Remember that you and your principal supervisor must sign the declarations.
- Please submit the co-author declarations together with your thesis. Either insert the declarations in the back of the thesis or create a joint PDF which you submit in the email together with the thesis.

Information on PhD student:	
Name of PhD student	Jeppe Bennekou Schroll
E-mail	jschroll@gmail.com
Date of birth	15. May 1980
Work place	The Nordic Cochrane Centre
Principal supervisor	Peter C. Gøtzsche

Title of PhD thesis:
Unpublished data, particularly in relation to harms, in clinical trials

This declaration concerns the following article:
Challenges in coding adverse events in clinical trials: a systematic review

The PhD student's contribution to the article: *(please use the scale (A,B,C) below as benchmark*)*	**(A,B,C)**
1. Formulation/identification of the scientific problem that from theoretical questions need to be clarified. This includes a condensation of the problem to specific scientific questions that is judged to be answerable by experiments	B
2. Planning of the experiments and methodology design, including selection of methods and method development	C

3. Involvement in the experimental work	B
4. Presentation, interpretation and discussion in a journal article format of obtained data	C

*Benchmark scale of the PhD student's contribution to the article		
A. refers to:	*Has contributed to the co-operation*	*0-33 %*
B. refers to:	*Has contributed considerably to the co-operation*	*34-66 %*
C. refers to:	*Has predominantly executed the work independently*	*67-100 %*

Signature of the co-authors:			
Date:	Name:	Title:	Signature:
28/7 2014	Emma Maund	MsC	
30/9 2014	Peter C Gøtzsche	Professor	

Signature of the PhD student and the principal supervisor:	
Date: 28/7 2014	Date: 30/9 2014
PhD student:	Principal supervisor:

GRADUATE SCHOOL OF HEALTH AND MEDICAL SCIENCES
UNIVERSITY OF COPENHAGEN

DECLARATION OF CO-AUTHORSHIP

About this form

- Declarations of co-authorship must be included in your PhD thesis if your thesis contains or is based on co-authored manuscripts or published articles.
- You must submit a declaration of co-authorship for all articles/manuscripts included.
- The declarations are important to the assessment committee's work as they must be able to identify and assess your contribution to the article(s).
- If you have more co-authors for an article than there is space for in the table, simply just complete another declaration for the same article with the remaining co-authors.
- Remember that you and your principal supervisor must sign the declarations.
- Please submit the co-author declarations together with your thesis. Either insert the declarations in the back of the thesis or create a joint PDF which you submit in the email together with the thesis.

Information on PhD student:	
Name of PhD student	Jeppe Bennekou Schroll
E-mail	jschroll@gmail.com
Date of birth	15. May 1980
Work place	The Nordic Cochrane Centre
Principal supervisor	Peter C. Gøtzsche

Title of PhD thesis:
Unpublished data, particularly in relation to harms, in clinical trials

This declaration concerns the following article:
Searching for unpublished data for Cochrane reviews: cross sectional study

The PhD student's contribution to the article: *(please use the scale (A,B,C) below as benchmark*)*	**(A,B,C)**
1. Formulation/identification of the scientific problem that from theoretical questions need to be clarified. This includes a condensation of the problem to specific scientific questions that is judged to be answerable by experiments	B
2. Planning of the experiments and methodology design, including selection of methods and method development	B

3. Involvement in the experimental work	C
4. Presentation, interpretation and discussion in a journal article format of obtained data	C

**Benchmark scale of the PhD student's contribution to the article*		
A. refers to:	*Has contributed to the co-operation*	*0-33 %*
B. refers to:	*Has contributed considerably to the co-operation*	*34-66 %*
C. refers to:	*Has predominantly executed the work independently*	*67-100 %*

Signature of the co-authors:			
Date:	**Name:**	**Title:**	**Signature:**
8/6/14	Lisa Bero	Professor	Lisa Bero
30/9 2014	Peter C Gøtzsche	Professor	PCGøtzsche

Signature of the PhD student and the principal supervisor:	
Date: 30/9 2014 **PhD student:**	**Date:** 30/9 2014 **Principal supervisor:** PCGøtzsche

DECLARATION OF CO-AUTHORSHIP

About this form

- Declarations of co-authorship must be included in your PhD thesis if your thesis contains or is based on co-authored manuscripts or published articles.
- You must submit a declaration of co-authorship for all articles/manuscripts included.
- The declarations are important to the assessment committee's work as they must be able to identify and assess your contribution to the article(s).
- If you have more co-authors for an article than there is space for in the table, simply just complete another declaration for the same article with the remaining co-authors.
- Remember that you and your principal supervisor must sign the declarations.
- Please submit the co-author declarations together with your thesis. Either insert the declarations in the back of the thesis or create a joint PDF which you submit in the email together with the thesis.

Information on PhD student:	
Name of PhD student	Jeppe Bennekou Schroll
E-mail	jschroll@gmail.com
Date of birth	15. May 1980
Work place	The Nordic Cochrane Centre
Principal supervisor	Peter C. Gøtzsche

Title of PhD thesis:
Unpublished data, particularly in relation to harms, in clinical trials

This declaration concerns the following article:
Deaths in trials should always be reported

The PhD student's contribution to the article: *(please use the scale (A,B,C) below as benchmark*)*	**(A,B,C)**
1. Formulation/identification of the scientific problem that from theoretical questions need to be clarified. This includes a condensation of the problem to specific scientific questions that is judged to be answerable by experiments	C
2. Planning of the experiments and methodology design, including selection of methods and method development	C

3. Involvement in the experimental work	C
4. Presentation, interpretation and discussion in a journal article format of obtained data	C

*Benchmark scale of the PhD student's contribution to the article		
A. refers to:	*Has contributed to the co-operation*	*0-33 %*
B. refers to:	*Has contributed considerably to the co-operation*	*34-66 %*
C. refers to:	*Has predominantly executed the work independently*	*67-100 %*

Signature of the co-authors:			
Date:	Name:	Title:	Signature:

Signature of the PhD student and the principal supervisor:	
Date: 30/9 2014	Date: 30/9 2014
PhD student:	Principal supervisor:

GRADUATE SCHOOL OF HEALTH AND MEDICAL SCIENCES
UNIVERSITY OF COPENHAGEN

DECLARATION OF CO-AUTHORSHIP

About this form

- Declarations of co-authorship must be included in your PhD thesis if your thesis contains or is based on co-authored manuscripts or published articles.
- You must submit a declaration of co-authorship for all articles/manuscripts included.
- The declarations are important to the assessment committee's work as they must be able to identify and assess your contribution to the article(s).
- If you have more co-authors for an article than there is space for in the table, simply just complete another declaration for the same article with the remaining co-authors.
- Remember that you and your principal supervisor must sign the declarations.
- Please submit the co-author declarations together with your thesis. Either insert the declarations in the back of the thesis or create a joint PDF which you submit in the email together with the thesis.

Information on PhD student:	
Name of PhD student	Jeppe Bennekou Schroll
E-mail	jschroll@gmail.com
Date of birth	15. May 1980
Work place	The Nordic Cochrane Centre
Principal supervisor	Peter C. Gøtzsche

Title of PhD thesis:
Unpublished data, particularly in relation to harms, in clinical trials

This declaration concerns the following article:
FDA reports provided more data but were more difficult to use than EMA reports

The PhD student's contribution to the article: *(please use the scale (A,B,C) below as benchmark*)*	**(A,B,C)**
1. Formulation/identification of the scientific problem that from theoretical questions need to be clarified. This includes a condensation of the problem to specific scientific questions that is judged to be answerable by experiments	B
2. Planning of the experiments and methodology design, including selection of methods and method development	B

3. Involvement in the experimental work	B
4. Presentation, interpretation and discussion in a journal article format of obtained data	B

**Benchmark scale of the PhD student's contribution to the article*		
A. refers to:	*Has contributed to the co-operation*	*0-33 %*
B. refers to:	*Has contributed considerably to the co-operation*	*34-66 %*
C. refers to:	*Has predominantly executed the work independently*	*67-100 %*

Signature of the co-authors:			
Date:	**Name:**	**Title:**	**Signature:**
8/5/14	Lisa Bero	Professor	Lisa Bero
	Maher Abdel-Sattar	Pharmacist	

Signature of the PhD student and the principal supervisor:	
Date: 30/9 2014 **PhD student:**	**Date:** 30/9 2014 **Principal supervisor:**

GRADUATE SCHOOL OF HEALTH AND MEDICAL SCIENCES
UNIVERSITY OF COPENHAGEN

DECLARATION OF CO-AUTHORSHIP

About this form

- Declarations of co-authorship must be included in your PhD thesis if your thesis contains or is based on co-authored manuscripts or published articles.
- You must submit a declaration of co-authorship for all articles/manuscripts included.
- The declarations are important to the assessment committee's work as they must be able to identify and assess your contribution to the article(s).
- If you have more co-authors for an article than there is space for in the table, simply just complete another declaration for the same article with the remaining co-authors.
- Remember that you and your principal supervisor must sign the declarations.
- Please submit the co-author declarations together with your thesis. Either insert the declarations in the back of the thesis or create a joint PDF which you submit in the email together with the thesis.

Information on PhD student:	
Name of PhD student	Jeppe Bennekou Schroll
E-mail	jschroll@gmail.com
Date of birth	15. May 1980
Work place	The Nordic Cochrane Centre
Principal supervisor	Peter C. Gøtzsche

Title of PhD thesis:
Unpublished data, particularly in relation to harms, in clinical trials

This declaration concerns the following article:
FDA reports provided more data but were more difficult to use than EMA reports

The PhD student's contribution to the article (please use the scale (A,B,C) below as benchmark):	(A,B,C)
1. Formulation/identification of the scientif ic problem that from theoretical questions need to be clarified. This includes a condensation of the problem to specific scientific questions that is judged to be answerable by experiments	B
2. Planning of the experiments and methodology design, including selection of methods and method development	B

3. Involvement in the experimental work	B
4. Presentation, interpretation and discussion in a journal article format of obtained data	B

*Benchmark scale of the PhD student's contribution to the article		
A. refers to:	*Has contributed to the co-operation*	*0-33 %*
B. refers to:	*Has contributed considerably to the co-operation*	*34-66 %*
C. refers to:	*Has predominantly executed the work independently*	*67-100 %*

Signature of the co-authors:

Date:	Name:	Title:	Signature:
	Lisa Bero	Professor	
7/28/2014	Maher Abdel-Sattar	Pharmacist	

Signature of the PhD student and the principal supervisor:

Date: 30/9 2014	Date: 30/9 2014
PhD student:	Principal supervisor:

DECLARATION OF CO-AUTHORSHIP

About this form

- Declarations of co-authorship must be included in your PhD thesis if your thesis contains or is based on co-authored manuscripts or published articles.
- You must submit a declaration of co-authorship for all articles/manuscripts included.
- The declarations are important to the assessment committee's work as they must be able to identify and assess your contribution to the article(s).
- If you have more co-authors for an article than there is space for in the table, simply just complete another declaration for the same article with the remaining co-authors.
- Remember that you and your principal supervisor must sign the declarations.
- Please submit the co-author declarations together with your thesis. Either insert the declarations in the back of the thesis or create a joint PDF which you submit in the email together with the thesis.

Information on PhD student:	
Name of PhD student	Jeppe Bennekou Schroll
E-mail	jschroll@gmail.com
Date of birth	15. May 1980
Work place	The Nordic Cochrane Centre
Principal supervisor	Peter C. Gøtzsche

Title of PhD thesis:
Unpublished data, particularly in relation to harms, in clinical trials

This declaration concerns the following article:
Assessment of harms in clinical trials of the anti-obesity drug orlistat

The PhD student's contribution to the article: *(please use the scale (A,B,C) below as benchmark*)*	(A,B,C)
1. Formulation/identification of the scientific problem that from theoretical questions need to be clarified. This includes a condensation of the problem to specific scientific questions that is judged to be answerable by experiments	B
2. Planning of the experiments and methodology design, including selection of methods and method development	B

3. Involvement in the experimental work	B
4. Presentation, interpretation and discussion in a journal article format of obtained data	C

*Benchmark scale of the PhD student's contribution to the article		
A. refers to:	*Has contributed to the co-operation*	*0-33 %*
B. refers to:	*Has contributed considerably to the co-operation*	*34-66 %*
C. refers to:	*Has predominantly executed the work independently*	*67-100 %*

Signature of the co-authors:

Date:	Name:	Title:	Signature:
30/9 2014	Peter C Gøtzsche	Professor	
30/9'14	Elisabeth Penninga	MD	

Signature of the PhD student and the principal supervisor:

Date: 30/9 2014	Date: 30/9 2014
PhD student:	Principal supervisor: