

Frequency and reasons for outcome reporting bias in clinical trials: interviews with trialists

R M D Smyth, research associate,^{1,2} J J Kirkham, research associate,¹ A Jacoby, professor of medical sociology,² D G Altman, professor of statistics in medicine,³ C Gamble, senior lecturer,¹ P R Williamson, professor of medical statistics¹

¹Centre for Medical Statistics and Health Evaluation, University of Liverpool, Liverpool, UK

²Division of Public Health, University of Liverpool, Liverpool, UK

³Centre for Statistics in Medicine, University of Oxford, Oxford, UK
Correspondence to: P R Williamson
prw@liv.ac.uk

Cite this as: *BMJ* 2010;341:c7153
doi:10.1136/bmj.c7153

ABSTRACT

Objectives To provide information on the frequency and reasons for outcome reporting bias in clinical trials.

Design Trial protocols were compared with subsequent publication(s) to identify any discrepancies in the outcomes reported, and telephone interviews were conducted with the respective trialists to investigate more extensively the reporting of the research and the issue of unreported outcomes.

Participants Chief investigators, or lead or coauthors of trials, were identified from two sources: trials published since 2002 covered in Cochrane systematic reviews where at least one trial analysed was suspected of being at risk of outcome reporting bias (issue 4, 2006; issue 1, 2007, and issue 2, 2007 of the Cochrane library); and a random sample of trial reports indexed on PubMed between August 2007 and July 2008.

Setting Australia, Canada, Germany, the Netherlands, New Zealand, the United Kingdom, and the United States.

Main outcome measures Frequency of incomplete outcome reporting—signified by outcomes that were specified in a trial's protocol but not fully reported in subsequent publications—and trialists' reasons for incomplete reporting of outcomes.

Results 268 trials were identified for inclusion (183 from the cohort of Cochrane systematic reviews and 85 from PubMed). Initially, 161 respective investigators responded to our requests for interview, 130 (81%) of whom agreed to be interviewed. However, failure to achieve subsequent contact, obtain a copy of the study protocol, or both meant that final interviews were conducted with 59 (37%) of the 161 trialists. Sixteen trial investigators failed to report analysed outcomes at the time of the primary publication, 17 trialists collected outcome data that were subsequently not analysed, and five trialists did not measure a prespecified outcome over the course of the trial. In almost all trials in which prespecified outcomes had been analysed but not reported (15/16, 94%), this under-reporting resulted in bias. In nearly a quarter of trials in which prespecified outcomes had been measured but not analysed (4/17, 24%), the "direction" of the main findings influenced the investigators' decision not to analyse the remaining data collected. In 14 (67%) of the 21 randomly selected

PubMed trials, there was at least one unreported efficacy or harm outcome. More than a quarter (6/21, 29%) of these trials were found to have displayed outcome reporting bias.

Conclusion The prevalence of incomplete outcome reporting is high. Trialists seemed generally unaware of the implications for the evidence base of not reporting all outcomes and protocol changes. A general lack of consensus regarding the choice of outcomes in particular clinical settings was evident and affects trial design, conduct, analysis, and reporting.

INTRODUCTION

Publication of complete trial results is important to allow clinicians, consumers, and policy makers to make better informed decisions about healthcare. Not reporting a study on the basis of the strength and "direction" of the trial results has been termed "publication bias."¹ An additional and potentially more serious threat to the validity of evidence based medicine is the selection for publication of a subset of the original recorded outcomes on the basis of the results,² which is referred to as "outcome reporting bias." The results that the reader sees in the publication may appear to be unselected, but a larger data set from which those results may have been selected could be hidden. This kind of bias affects not just the interpretation of the individual trial but also any subsequent systematic review of the evidence base that includes it.³

The recent scientific literature has given some attention to the problems associated with incomplete outcome reporting, and there is little doubt that non-reporting of prespecified outcomes has the potential to cause bias.⁴⁻⁶ Eight previous studies have described discrepancies with respect to outcomes between the protocol or trial registry entry of randomised trials and subsequent publications.^{4,7-14} In two of these studies,^{4,8} trialists were surveyed to ascertain the frequency and reasons for under-reporting outcomes. Although both studies had relatively low response rates, results showed that lack of clinical importance (7/23; 18/29), lack of statistical significance (7/23; 13/29), and journal space restrictions (7/23) were the reasons most often cited for under-reporting.

A more recent retrospective review of publications and follow-up survey of trial investigators confirmed these reasons for non-reporting.¹⁵ However, information from self reported questionnaires suggested trialists often did not acknowledge the existence of unreported outcomes, even when there was evidence to the contrary in their publication.

We undertook a study comparing the outcomes specified in trial protocols with those in published reports, and obtained detailed information relating to discrepancies from trial investigators by means of a structured telephone interview. Our objectives were to explore all discrepancies between outcomes specified in the protocol and those reported in the trial publication, identify cases of selective reporting of the set of outcomes measured, assess whether outcome reporting bias had occurred, and provide a more detailed understanding of why trialists do not report previously specified outcomes.

METHODS

We compared trial protocols with subsequent reports, and then conducted interviews by telephone with the trialists to obtain information relating to discrepancies. All interviews were conducted in English; because some of the investigators were overseas, we ascertained the feasibility of doing this with the researcher beforehand. The project was not required to be reviewed under the terms of the governance arrangements for research ethics committees in the United Kingdom. However, trialists were informed that confidentiality would be assured and all identifiable data from the interviews would be shared only by immediate members of the research team (AJ, JJK, PRW, and RMDS).

Participants

The study reported here was part of the larger Outcome Reporting Bias in Trials (ORBIT) project to estimate the prevalence and effect of outcome reporting bias in clinical trials on the primary outcomes of systematic reviews from three issues of the Cochrane Library: issue 4, 2006; issue 1, 2007, and issue 2, 2007.³ To assess the risk of outcome reporting bias in this context, a classification system was developed and applied when a trial was excluded from a meta-analysis either because data for the review primary outcome were not reported or because data were reported incompletely (for example, just as “not significant”). The system identified whether there was evidence that: (a) the outcome was measured and analysed but only partially reported; (b) the outcome was measured but not necessarily analysed; (c) it was unclear whether the outcome was measured; or (d) it was clear that the outcome was not measured. One project researcher and the corresponding author of the review independently classified any trial in the review that did not report or partially reported results for the review primary outcome on the basis of all identified publications for that trial.

We had intended originally to interview the investigators of trials where outcome reporting bias was suspected for the review primary outcome and match the

interview to one from an investigator of a trial in the same review where there was no suspicion of outcome reporting bias. However, it became apparent early on that matching would not be possible owing to the low response rate. We therefore interviewed trialists from within these two groups without any attempt at matching. To reduce potential problems with recall, we restricted interviews to authors of trials published since 2002; therefore, trials published since 2002 from Cochrane reviews that included at least one trial where outcome reporting bias was suspected were eligible for the current study.

We also drew on a second source, PubMed, and randomly sampled trial investigators from the 14 758 trial reports indexed on PubMed over the 12 month period from August 2007 to July 2008. We used the advance search option in PubMed and searched for randomised controlled trials in human participants that were published in English.

We aimed to conduct interviews with the trial chief investigator, but also with the lead author where they were different. Several strategies were used to identify current contact details for these two groups of individuals. We began with the email address included on the publication. These email addresses were often no longer valid in cases where a considerable amount of time had elapsed between a publication and this analysis. In this situation, we tried to locate more recent publications by the same author by searching electronic databases. We then attempted to establish contact via the website or directory of the author’s institution, by contacting the journal in which the study was published, or by searching for the author’s name using an internet search engine such as Google. A further strategy if these measures were unsuccessful was to contact other authors the original author had recently published with to request contact information.

Eligible trialists were contacted initially by an email from one of the study team (RMDS) informing them about the study and inviting them to take part in an interview. The trial protocol was requested at this point. One reminder was sent two weeks later to those trial investigators who had not responded to the first email. If no reply had been received within a month of first contact, we assumed the trialist did not want to take part and then consecutively invited coauthors to do so. Trial investigators who actively opted out (declined to be interviewed) were asked to nominate a coauthor who could be approached. To maximise our sample, all those eligible and in agreement were interviewed. Timing of the interview was therefore dependent on when the trialist or coauthor responded to the email.

Sample size

Given the lack of similar studies, it was difficult to estimate the number of trialists who would agree to be interviewed. To maximise information on reasons for discrepancies in reporting of primary outcomes, investigators from all trials included in a Cochrane review that were suspected of outcome reporting bias for the

Table 1 Characteristics of trialists who responded to the invitation to participate in the study and those who did not

| | All (n=268) | Responded (n=59) | Did not respond (n=209) | χ^2 | P value |
|--|----------------|---------------------|----------------------------|----------|---------|
| Cohort | | | | | |
| Randomly selected from PubMed | 85 | 21 (25%) | 64 (75%) | | |
| Identified in Cochrane reviews: | | | | | |
| Suspected of outcome reporting bias* | 85 | 17 (20%) | 68 (80%) | 0.056 | 0.812 |
| Not suspected of outcome reporting bias* | 98 | 21 (21%) | 77 (79%) | | |
| Trial sample size | | | | | |
| <100 | 122 | 28 (23%) | 94 (77%) | 0.627 | 0.731 |
| 100-999 | 128 | 26 (20%) | 102 (80%) | | |
| >1000 | 18 | 5 (28%) | 13 (72%) | | |
| Total | 268 | 59 | 209 | | |
| Funding | | | | | |
| Non-commercial | 105 | 33 (31%) | 72 (69%) | 10.064 | 0.018 |
| Industry | 116 | 21 (18%) | 95 (82%) | | |
| Unfunded | 1 | 0 (0%) | 1 (100%) | | |
| Not stated | 46 | 5 (11%) | 41 (89%) | | |
| Total | 268 | 59 | 209 | | |
| Statistician involved | | | | | |
| Statistician involved | 17 | 7 (41%) | 10 (59%) | 5.012 | 0.082 |
| Statistician involvement unclear | 45 | 12 (27%) | 33 (73%) | | |
| No information provided | 206 | 40 (19%) | 166 (81%) | | |
| Total | 268 | 59 | 209 | | |

*In relation to review primary outcome.

review primary outcome were approached for interview (n=85). In order to interview a similar number of trial investigators from the other two groups—that is, trials from the same Cochrane reviews that were considered to show no outcome reporting bias and the randomly selected cohort of PubMed trials—a further 98 and 85 trialists, respectively, were approached. The sample size was governed by pragmatic considerations relating to the number of eligible trials identified, the resources available, and the willingness of trial investigators to participate.

Comparison of the trial protocol and publication(s)

Before interview, two researchers (RMDS and JJK) independently assessed consistency of the trial protocol with its related publication(s) to identify any discrepancies in outcomes. For each trial we reviewed all associated published articles. All trial reports were written in English; conference abstracts and posters were excluded.

Unreported outcomes were defined as those that were prespecified in the trial protocol but were not reported in any subsequent publications. For each trial where selective outcome reporting was identified, an “aide memoire” for the interview was produced showing the outcomes prespecified in the protocol and the outcomes published in the trial report.

Interview schedule

An interview schedule was devised to identify factors influencing research reporting. The schedule had a

chronological structure that followed the sequence of events for a trial: it started by establishing the research question and discussing the writing of the study protocol; moved to trial conduct; and finally went through the publication of the study findings. Where outcomes were not reported as specified in the trial protocol, we asked trialists to give an explanation and justification for their decision making and an account of how decisions were made. Trial investigators were also asked to provide the statistical significance of these unreported outcomes if the data were collected and analysed.

The schedule comprised three broad areas of questioning: (a) specific questions relating to the presence of selective outcome reporting (that is, selection of a subset of the original recorded outcomes for inclusion in the trial publication (see below)); (b) specific questions relating to writing the manuscript and getting the paper published; and (c) general questions relating to publication bias. This current paper addresses the first of these areas; qualitative data obtained from the interviews that explore trialists’ experiences more generally of carrying out and reporting research will be reported in a separate paper. Trial investigators had a copy of the trial protocol and subsequent publication(s) with them at the time of the interview, as did the interviewer.

Specific questions were asked relating to the possibility of selective outcome reporting; that is, whether a subset of the original recorded outcomes was selected for inclusion in the trial publication:

- Were data for the unreported outcome collected?
 - If not, explore with trialist why not
- What was the trialist’s perception of the importance of the outcome(s)?
- Why were data on the prespecified outcome(s) not collected?
- What did the trialist consider were the benefits, limitations, and difficulties associated with not collecting data on the prespecified outcome(s)?
- If data on the unreported outcome were collected, were they analysed?
 - If not, explore with trialist why outcome data were not analysed
- What did the trialist consider were the benefits, limitations, and difficulties associated with analysing prespecified outcome(s) data?
- Why is reporting all prespecified outcomes important?
- Why were the analysed outcome(s) not reported?
- What did the trialist consider were the benefits, limitations, and difficulties associated with not reporting the analysed outcome(s)?
- What was the statistical significance of the unreported analysed outcome?

Explore with trialists the meaning of statistical significance generally

Explore with trialists the importance or not of the statistical significance of the non-reported outcome

Explore how the statistical significance of an outcome might influence reporting

All interviews were conducted by a researcher with experience of qualitative research and of interviewing clinicians (RMDS). Two trialists requested, and were provided with, their interview transcript in order to add to or clarify issues and to ensure validity of the account produced; neither changed their accounts.

Initially, we had not planned to exclude trialists on the grounds of first language. Three pilot interviews were performed with trialists from non-English speaking institutions whose first language was not English to assess the feasibility of carrying out interviews with such individuals, with respect to information given and interpretation. Email follow-up was then undertaken to verify the authors' responses and our understanding of the information they gave. These interviews are included in the analysis. However, there was some difficulty eliciting data and information because of the language barriers, and follow-up questions were required by email. Given the resource implications of follow-up, we decided that inclusion should be limited to trialists working at institutions where English is usually spoken.

Analysis

Each interview was tape recorded with the trialist's permission, transcribed, and anonymised. All the transcripts were examined in detail by one of the investigators (RMDS). At this stage interpretations were shared with two of the other investigators—one a medical sociologist with extensive qualitative research experience (AJ), the other a statistician with recognised outcome reporting experience (PRW)—both of whom read the transcripts independently. Data obtained from the trial investigators in response to the explicit questions relating to the presence of outcome reporting bias were of particular interest.

Reasons provided by the trialists for not reporting prespecified outcomes were classified by the investigators (AJ, PRW, and RMDS). Categories for outcome reporting were preset and added to or modified as new responses emerged. Through discussion the investigators reached a consensus regarding the categories for classification of the reasons for discrepancies: (a) the outcome was measured and analysed but not reported; (b) the outcome was measured but not analysed; (c) the outcome was not measured; and (d) the outcome was reported but not prespecified in the protocol. A particular reporting practice was deemed to be associated with bias if the reason for non-reporting of an analysis was related to the results obtained; for example, if the result for the outcome was non-significant.

For some trials from Cochrane reviews identified by the ORBIT project, we already suspected outcome reporting bias if trials did not report or partially reported results for the review primary outcome on the basis of all identified publications for that trial. For this reason, estimates of the frequencies of various discrepancies in reporting⁶ are based only on data from the randomly selected cohort identified from PubMed. However, the reasons given by the trial investigators

for selective reporting are provided for all cohorts together.

RESULTS

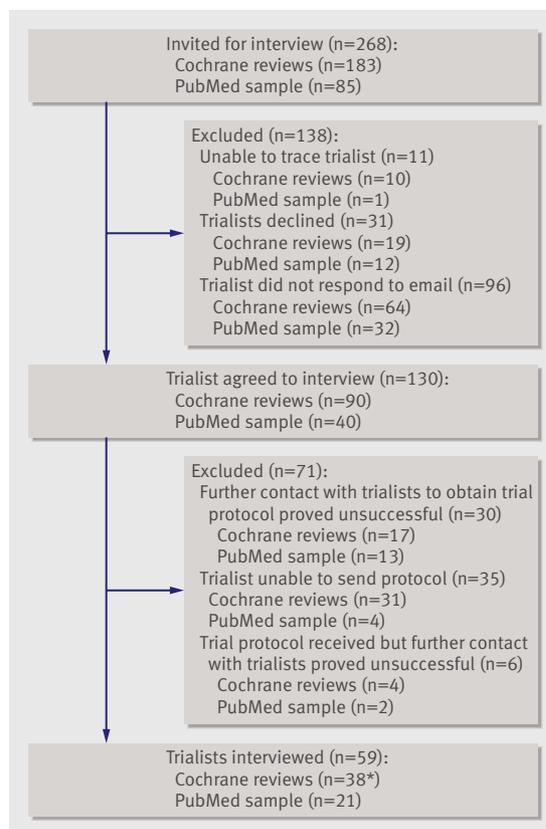
Response rate

A total of 268 trials were identified for inclusion; 183 of these were from the cohort of trials covered in Cochrane systematic reviews and 85 were randomly selected from PubMed. One hundred and sixty one of the 268 respective investigators acknowledged receipt of our email, of whom 31 (19%) declined our invitation. For those authors who did not respond to emails, we were unable to ascertain whether this was because of incorrect or invalid contact information and hence non-receipt of our invitation. We were unable to obtain contact details for chief investigators or lead authors for 19 (7%) trials; however, we established contact with coauthors for eight of these.

Thirty one trialists (including six coauthors) declined to be interviewed. The majority (17 trialists) gave no reason for doing so. Those who did provide reasons for declining cited personal circumstances (five trialists), work commitments (four trialists), and difficulty in recalling the trial (three trialists). One trialist did not wish to be interviewed by telephone, preferring email or written contact only. One trialist requested a copy of the ORBIT protocol, and subsequently declined to be interviewed.

Overall, 130 (81%) trialists initially agreed to be interviewed (113 chief investigators or lead authors; 17 coauthors), but further attempts to establish contact and request their trial protocols proved unsuccessful for 30 of these. A further 35 trialists were unable to provide a copy of their trial protocol and so were not interviewed: 15 were unable to locate a copy, 13 were unable to disclose details of their protocol because of restrictions imposed by funding bodies (funded solely by industry), five had protocols that were not written in English, and two trialists were unwilling to share their protocol with us, both stipulating that it was a confidential document. Six further trialists agreed to be interviewed and provided a copy of their protocol, but did not respond to any further email contact.

All those eligible and in agreement were interviewed. Overall 59 (37%) trialists, 38 identified from the cohort of trials from Cochrane reviews and 21 from trials recently indexed in PubMed, were interviewed (fig 1). Characteristics of the publication(s) were compared between trialists agreeing (n=59) and those not agreeing (n=209) to be interviewed (table 1). A higher proportion of trialists who did not agree to interview were funded by industry. There was some evidence that investigators for trials with reported involvement of a statistician were more likely to agree to be interviewed. There was no evidence of any association between a trialist agreeing to be interviewed and the sample size of the trial. Importantly, there was no evidence of an association between agreement rate and the level of suspicion of outcome reporting bias, as indicated by the two groups of trials identified from Cochrane reviews.



Trialists eligible for interview. *Includes 17 associated with trials suspected of outcome reporting bias, 21 responsible for trials with no outcome bias suspected

Characteristics of those interviewed

Most interviews (48/59, 81%) were conducted with the trial chief investigator, eight of whom were PhD students. A further eight interviews were conducted with the lead author and three with a coauthor. Additional data were obtained at the time of the semistructured interviews to provide a descriptive summary of the sample. Characteristics of the interviewees and trials are shown in table 2. Trialists were based in Australia, Canada, Germany, the Netherlands, New Zealand, the United Kingdom, and the United States.

The median publication year was 2005 (range 2002–2008). In all but one case, interviews were performed with one trial investigator from each trial; for one trial we interviewed the chief investigator, lead author, and statistician simultaneously at their request. Interviews lasted on average 56 minutes (range 19–96 minutes).

Description of protocols

Forty three (73%) trialists sent their full study protocol for review. The quality of the protocols was highly variable, but all lacked some key information such as a clear definition of the primary outcome or a description of the statistical methods planned for the analysis. For the remaining 16 (27%) trialists, no complete protocol was available. Instead, six trialists provided the ethics committee application, two the funding application, one a summary of the full protocol, and three

extracts from relevant chapters of doctoral theses. For one trial the protocol sent was simply a letter with the description of the trial that was sent to the funders. For two trialists the full protocol was available but not in English, so an English summary was provided. For the remaining trial we obtained an abridged version of the protocol from the clinical trials.gov website.

In just under half of the protocols and substituted documents (27/59, 46%), one primary outcome was explicitly specified. In four (7%) protocols more than one outcome was specified, and in 28 (47%) none was specified. In three cases where a primary outcome was not explicitly stated, we assumed that the outcome used in the sample size calculation was the primary outcome; in two of these cases, this matched the stated primary outcome in the publication. In the remaining trial, the stated primary outcome in the report did not match the outcome on which the protocol sample size calculation was based. For three further trials where a primary outcome was not explicitly stated, we assumed that the primary outcome related to the main aim as stated in the research question in the protocol. Across the 37 trials with explicit (31 trials) or implicit (6 trials) primary outcomes, a total of 233 secondary outcomes were specified in the protocols (median of five, range 1–17). For the remaining 22 protocols where we could not deduce clearly which outcome was the primary outcome, a total of 132 outcomes were specified in the protocols (median of five, range 1–21). For the 59 protocols together, a total of 419 outcomes were specified; the median number of all outcomes specified was six (range 1–21). A statistical analysis plan was provided in 40 (68%) protocols.

Reasons for discrepancies in reporting of an individual outcome

Inconsistencies of reporting were divided into those where an outcome specified in the protocol was omitted in the subsequent publication and those where an outcome presented in the publication was not prespecified in the protocol. All discrepancies identified across all 59 trials are included here. These categories are mutually exclusive for a particular outcome.

Prespecified outcome measured and analysed by the time of the primary publication but not reported

Sixteen trial investigators failed to report a total of 30 analysed outcomes. For the majority (15/16, 94%) of the trials, this was judged to have led to biased under-reporting (table 3). Although the decisions made by trialists not to report outcomes potentially induced bias, the information provided by trial investigators confirmed that, in all but one case (trialist 30), not reporting an outcome was unintentional. The most common reasons given by the trialists for not reporting outcomes were related to a lack of understanding about the importance of reporting “negative” results (trialists 12, 18, 22, 29, 44, 49, and 50), the data being perceived to be uninteresting (trialists 09, 18, and 36), there being too few events worth reporting (trialists 12, 39, and 41),

Table 2 | Characteristics of trial investigators and trials included in the study

| | Trialists from Cochrane review cohort | | Trialists from PubMed cohort (n=21) |
|-------------------------------------|--|--|-------------------------------------|
| | Suspected of outcome reporting bias (n=17) | Not suspected of outcome reporting bias (n=21) | |
| Previous research experience | | | |
| Extensive | 10 (59%) | 7 (33%) | 15 (71%) |
| Coauthor had experience | 7 (41%) | 10 (48%) | 4 (19%) |
| No experience in research team | 0 | 4 (19%) | 2 (9%) |
| Sample size | | | |
| <100 | 6 (35%) | 11 (52%) | 11 (52%) |
| 100-999 | 8 (47%) | 9 (43%) | 9 (43%) |
| >1000 | 3 (18%) | 1 (5%) | 1 (5%) |
| Funding | | | |
| Non-commercial | 10 (59%) | 15 (71%) | 11 (52%) |
| Industry | 7 (41%) | 5 (24%) | 10 (48%) |
| Unfunded | 0 | 1 (5%) | 0 |
| Statistician involved | | | |
| Throughout trial | 9 (53%) | 8 (38%) | 13 (62%) |
| Consulted | 1 (6%) | 5 (24%) | 6 (29%) |
| No statistician involved | 7 (41%) | 8 (38%) | 2 (10%) |

and the need for brevity or perceived space constraints imposed by the journal (trialists 32, 34, and 56).

There was deliberate misrepresentation of the results in one trial, in which a statistically significant increase in harm in the intervention group was suppressed. The investigator for this trial (trialist 30) described the difficulties presented by such a result, including the discussions the team had with each other and the wider clinical and research community, and the potential impact reporting the data would have had on service providers.

Prespecified outcome measured by the time of the primary publication but not analysed or reported

Data on 26 outcomes over 17 trials were collected but subsequently not analysed. Most reasons given by the trial investigators were not considered to indicate bias (table 4). However, in four trials, the direction of the main findings had influenced the investigators' decision not to analyse all the data collected. In three of these trials, results for the primary outcome were non-significant and analyses of some of the secondary outcomes were therefore considered of no value (trialists 05, 39, and 44). The remaining trialist, after finding an increase in harm for the primary outcome, chose not to analyse a secondary outcome (trialist 59).

Other reasons given by the trialists for not analysing collected outcome data included: inadequacies in data collection resulting in missing data (trialists 20 and 23); long term data not being available by the time primary outcome data were published (trialists 08, 23, and 43); delay in obtaining the data (trialists 40, 42, and 53); practical difficulties leading to uncertainty regarding the validity of the results (trialist 55); cuts in government funding meaning full analysis was not financially viable (trialist 15); a large volume of data presented in primary publication (trialist 45); and practical difficulties associated with the organisation of the trial (trialist

27). One trialist (trialist 41) had problems associated with not meeting the prespecified sample size, and an additional trial investigator (trialist 31) failed to analyse one outcome because the trial had a lower than expected recruitment rate (the prespecified sample size was 112, but only 54 participants were recruited), so the researchers decided to analyse fewer outcomes.

Prespecified outcome not measured

Thirteen prespecified outcomes across five trials were not measured over the course of the trial. Reasons given by the trial investigators related to data collection being too expensive or complicated, as well as there being insufficient time and resources to collect less important secondary outcomes.

Outcome reported but not prespecified in the protocol

An additional 11 outcomes were analysed and reported without being prespecified in the five respective trial protocols. The reasons that the trialists gave were associated with poor research practice and were attributed to shortfalls in the writing of the protocol. For example, one trialist acknowledged that only at the point of compiling the case report forms did they consider collecting data on the outcome, a decision that was mostly influenced by the fact that the outcome was routinely measured by clinicians.

Frequency of outcome reporting bias

Estimates of the frequencies of various discrepancies in reporting are based only on data from the randomly selected cohort identified from PubMed, because trials identified from the ORBIT project were not randomly selected. In the 21 PubMed trials, the primary outcome stated in the protocol was the same as in the publication in 18 (86%) cases. Of these 18 studies, three were funded solely by industry, nine were not commercially funded, and six received funding from both industry

Table 3 | Responses from trialists who had analysed data on a prespecified outcome but not reported them by the time of the primary publication (n=16)

| Trialist | Explanation from trialist | Category |
|----------|--|--|
| 09 | "It was just uninteresting and we thought it confusing so we left it out. It didn't change, so it was a result that we . . . you know, kind of not particularly informative let's say, and was to us distracting and uninteresting." | Bias |
| 12 | "There was no mortality. No mortality at all" "I had the idea that they would change with the intervention, but they didn't change and that is why they were never reported. It was the negative result, which is also a result, but we never report on it. Normally I would have published these, I also publish negative data, but we thought that these data would explain why the intervention should work and that these data did not show, so that's why but normally I would certainly publish negative results." | Bias |
| 18 | "We didn't bother to report it, because it wasn't really relevant to the question we were asking. That's a safety issue thing; there was nothing in it so we didn't bother to report it. It was to keep ethics committee happy. It is not as if we are using a new drug here, it is actually an established one, just an unusual combination, so if we are using new things we report all that sort of stuff, so it's not that experimental. We didn't bother to report it, because it wasn't really relevant to the question we were asking." "They're kind of standard things in the bone world and they never show anything unless you have 5000 people, I'm exaggerating, 500 people in each group. In the end there was nothing really in it, I mean there were no differences, it was a short duration study, very small numbers, so actually the chances of finding anything was very small. Even if we found something it was likely to be confounder, there was a statistical chance. I think the protocol was adapted from a kind of a standard one we used for lots of drug trials and some of those bits could have been left out actually." | Bias |
| 22 | "The whole study showed that there was nothing in the [intervention]. So the whole study was actually a negative result, so I don't think the fact that there was no effect prevented me from putting it into the paper. It was either possibly an oversight or possibly something I thought, 'well this isn't relevant.'" | Bias |
| 29 | "When I take a look at the data I see what best advances the story, and if you include too much data the reader doesn't get the actual important message, so sometimes you get data that is either not significant or doesn't show anything, and so you, we, just didn't include that. The fact that something didn't change doesn't help to explain the results of the paper." | Bias |
| 30 | "When we looked at that data, it actually showed an increase in harm amongst those who got the active treatment, and we ditched it because we weren't expecting it and we were concerned that the presentation of these data would have an impact on people's understanding of the study findings. It wasn't a large increase but it was an increase. I did present the findings on harm at two scientific meetings, with lots of caveats, and we discussed could there be something harmful about this intervention, but the overwhelming feedback that we got from people was that there was very unlikely to be anything harmful about this intervention, and it was on that basis that we didn't present those findings. The feedback from people was, look, we don't, there doesn't appear to be a kind of framework or a mechanism for understanding this association and therefore you know people didn't have faith that this was a valid finding, a valid association, essentially it might be a chance finding. I was kind of keen to present it, but as a group we took the decision not to put it in the paper. The argument was, look, this intervention appears to help people, but if the paper says it may increase harm, that will, it will, be understood differently by, you know, service providers. So we buried it. I think if I was a member of the public I would be saying 'what you are promoting this intervention you thought it might harm people—why aren't you telling people that?'" | Bias |
| 32 | "If we had a found a significant difference in the treatment group we would have reported that, and it certainly would have been something we probably would have been waving the flag about. To be honest, it would have come down to a word limit and we really just cannot afford to report those things, even a sentence used, and often you have a sentence about this, and a sentence about that, and so it doesn't allow you to discuss the more important findings that were positive or were negative as some of our research tends to be, because I guess it's a priority of relevance" | Bias |
| 34 | "No I think probably, it's possible, I am looking on the final one, but probably was each time, reduced and reduced from the start to submitting to the journal. It is very limited on numbers, probably we start to . . . it didn't get accepted so we kind of cut and cut, I believe this is what happened." (the manuscript went to four journals) | Bias |
| 36 | "It's as dull as ditchwater, it doesn't really say anything because the outcome wasn't different, so of course [trial treatment] is going to be more expensive and no more effective so of course it's not going to have a health economic benefit. Because you have got two treatments that don't really differ. I just think, we have got to find a different way of . . . so for example I said well can't we say something about the costs within the groups of those who relapsed and those who didn't, just so that people get a ball park, but it's written and he [coauthor—health economist] wants to put it in as it is, and I don't have a problem with that, it's rather a sense of I am not sure what it tells anybody." | Bias |
| 39 | "We analysed it and there was two patients who had the outcome, you know one in each arm, so we decided the numbers were so small that we didn't think that adding another row to the table to describe two patients added anything." | Bias |
| 41 | "Patients in this particular trial turned out to use very low amounts of drugs. So, there was nothing essentially to compare. The use of other drugs was not an important issue in this population. There was nothing to report. There was no reportable data, no interesting story in the secondary outcome data, and our intention was always to focus on the opiate use not on the other drugs. I did look, we do have data on other drug use, we have collected data as we promised, but essentially there is nothing to report in this data. Patients do not use other drugs heavily. We will present again, I have all the intentions, the data is available for analysis and for presentation if one of my students decide to do some work with this and help me out with this, absolutely it will get published, but I have to pick and chose what I am actually working on." | Bias |
| 44 | "We probably looked at it but again it doesn't happen by magic. So, I can't imagine that there would be a difference. Why we didn't? My guess is that we didn't look at it because that is something that has to be prospectively collected, and so I would assume that we collected it and there was just absolutely no difference, but I don't recall. I am pretty sure there would not be differences, it would be related to temperature, but what the results were I don't remember at this point." | Bias |
| 49 | "Yes because what happened is, I am left with a study where everything is [non-significant], even though we walked in believing that we would see a difference, and even though we had some preliminary information, you know anecdotal, that there should be a difference, there was no difference. So, it really turned out to be a very negative study. So we did collect that information, and again it's a non-result, but there are only so many negative results you can put into a paper." | Bias |
| 50 | "It didn't add anything else to the data, it changed but it wasn't anything that was remarkable, it wasn't a significant change. If it had been something that either added additional strength to the data, or if it was conflicted, if it turned out it went totally against our data, and was counteractive to what we were saying, yes we would have reported." | Bias |
| 54 | "Yes, we have those data on file, and I am sorry to say that we are writing-up so many papers sometimes we do not know what's in the other papers. It has been analysed, I know, because what I know right now is that all the measurements which would be performed in this study as well as in two other studies were done, very simply because the outcome is very difficult to measure, well it's very simple to measure, to get the antibodies is very difficult, and we got it from an organisation that gave us just enough to do the measurements. I know the results, saying from what I have in my head the outcome is going up, we have high levels of it, but I am not sure whether it was a significant increase or whether it was significant compared to the control group or the other group as mentioned in the study." | Delay in writing up of secondary outcomes beyond primary publication |
| 56 | "I actually disagree that this outcome is important, but that was probably a more pragmatic aspect of making sure that our protocol was funded, because I think some reviewers might have said, 'wow you are not measuring this outcome!' That said, there is a vast amount of literature showing that it's of completely no relevance but it was a practical decision to make sure we got money. Once we conducted the study and reflected on our results more we just didn't think it had that much validity in telling us very much about the condition. So for the sake of brevity we didn't report that. I didn't expect there would be much of a difference, and our results show that there wasn't much of a difference." | Bias |

Table 4 Responses from trialists who had collected data on a prespecified outcome but not analysed them by the time of the primary publication (n=17)

| Trialist | Explanation from trialist | Category |
|----------|--|--|
| 05 | "I think, you know, the issue with cost is it would have been relatively easy to get hospital charges, but actual costs are more complicated analysis, and given that there was no difference between the groups we didn't go onto that." | Secondary data not analysed because no difference in primary outcome |
| 08 | "Part of the problem is, as our economists keep on telling us, we have been collecting economic data, but we haven't had enough of them. So for instance, for someone going into treatment, you know, we still don't have a lot of participants who are receiving treatment. So, one thing that we now have is older participant rates, so the economists are just starting to work on that, saying: 'OK, now that we have older rates, we can project out what potential benefits in terms of like earnings over a life course, we might expect.'" | Long term outcome data not obtained by the time primary outcome data published |
| 15 | "Unfortunately, the year we were funded the funders levied large across the board cuts to all approved projects. Our recruitment also took longer than expected. As a result, we had neither the time nor the resources to analyse for effects on the several less important secondary or exploratory outcomes we mentioned in our initial proposal." | Insufficient time, resources, or both to analyse less important outcomes |
| 20 | "The problem was that they weren't completed as well as all the other measures, it was only just over two thirds of people who consistently completed their diary, so it would have reduced our numbers considerably." | Not analysed because of the amount of missing data |
| 23 | "It was missing, a lot of people were missing data, and because we were looking at pre- and post- if they were missing at either point then we had to throw out that person, so we just didn't have enough, and then across three different conditions, so if you have only six or seven people with data in each condition it really wasn't worth looking at." "Oh I know what it was; I do know what it was. Well I know why it wouldn't have been in the dissertation. I ended up getting pregnant, and so I actually defended the dissertation early, so I defended the dissertation before all of the six months data were collected, even though I said I would do all of it, and I did, but I wanted to get the defence done before I went on maternity leave, so they allowed me to defend just based on the pre-post data, and we didn't collect the health services data until six months, so I didn't have those data at the time that I defended. Now I did include the follow-up data in this published article but I probably yes, I didn't go back and probably didn't even code and clean all the data we collected on health services." | Not analysed because of the amount of missing data Long term outcome data not obtained by the time primary outcome data published |
| 27 | "Samples were taken and one could argue that this is unethical actually because those samples were taken, and stored, and frozen, and the results of analysis have never been published because the analysis was never done. Speaking as someone who is interested in trial ethics, that is probably unethical to take samples and not analyse them. Here is the answer to that question then, it's really simple issue, which is the reality of doing a trial. This was an utter nightmare. For the analysis we did all the preparation abroad, it was an utter nightmare. I wouldn't say I still wake up screaming. For a small aspect of a trial, you know, some of the more interesting outcomes and laboratory based things to support whatever the hypothesis is, it grew to assume gargantuan proportions because there was no electricity, it was 40 degrees, there were bugs crawling around everywhere, and this turned into a nightmare. Getting it back to . . . there were no facilities to do these in the country and getting it back to us was a nightmare. And, the reason for the reporting differences for this biochemical stuff is actually I suspect not to do with bias but to do with pragmatics." | Not analysed because of practical difficulties |
| 31 | "The main reason was limited power, so we had fewer variables up front, because of small sample size. Most of it was driven by the small sample size, you know when you have 54 participants and you are trying to look at outcomes, so we made decisions about what seemed to be the most important variable. I think I tend to be conservative and part of this is just the sample was so small. I have tried to be very focussed in the papers and there is lots of data that is lying dormant essentially because of that." | Limited analysis undertaken owing to poor recruitment |
| 39 | "No, we haven't analysed it yet, you know we saw no effect on [primary outcome], which means that the drug is not ever going to get, well not in the foreseeable future, going to get on the market. So I think it's less important, people aren't going to have access to it, it's less important to share, to make that data public given the low likelihood that it's going to show anything. It's not going to have any immediate clinical implications for anybody." | Secondary data not analysed because no difference in primary outcome |
| 40 | "Well, we will be doing it, so what happens is you send someone down to a laboratory, they do the various procedures and then you have to have someone who scores the protocol, which requires various software and code writing, and then someone to spend the time to actually analyse the data. And, it just, it takes a while and so our primary interest was getting out the information, these other measures were more secondary outcome measures. It gets into issues around priorities, staff leaving, new staff having to be trained, it has just taken us a while to get this stuff analysed and as I said it's, it has been analysed and we are hoping to write the manuscript for this, this fall." | Delay in obtaining the data |
| 41 | "The cost effectiveness was really not conducted in this study. I am not an expert on cost effectiveness, there is a different team that works with us on cost effectiveness analysis, and this is their own survey now. And also the size, the effect sizes to be expected for the cost effectiveness analysis are much smaller and therefore the sample size did not really afford this type of analysis." "There is also a limit of what number of statistical analyses you can actually reliably do with a small sample size, so considering the sample size of this particular study, we decided to report only on the really main primary selected outcomes and not run many analysis, because we understand that they lose the power, and they lose the power of the conclusion." | Not analysed because of poor recruitment |
| 42 | "Well that, it turns out that our funders [industry] kind of went through some internal reorganisation, and as a result there are delays in the analyses of this outcome, and so we are still waiting to receive the final data from that, to publish it. We have got partial data completed on that part of the project but not complete. The data, it's really held by us. The funders have the samples, they just have that one set of samples, and they send us the data and we own the data and do the analysis, they just did the assay. We are interpreting it and analysing it and everything." | Delay in obtaining the data |
| 43 | "The economic evaluation, it's in processes. Well, we were supposed to have it analysed by the end of the month, I can't give you any preliminary. We know about the utilisation already but I don't know about the related costs. Our goal is to have it actually submitted this year." | Long term outcome data not obtained by the time primary outcome data published |
| 44 | "So it was a negative trial, it was not a failed trial, we have statistical power, but it was negative basically, we showed that in the context of our environment the intervention did not increase the primary outcome so basically the outcomes were identical in the two groups. Given this there was absolutely no reason to expect the difference in anything we would measure in the blood. So, we had collected all that, it was sitting in the refrigerator, in the freezer, so we didn't do the analysis because it was completely obvious that it would be negative." | Samples not analysed because no difference in primary outcome |

| Trialist | Explanation from trialist | Category |
|----------|---|---|
| 45 | "That outcome is a paper still in the making. That data was kept completely aside, you know, as you can see the paper that we are talking about has got so much in it and we did hum and ah whether we were going to break it down or present it separately, or any of that sort of stuff, but when we did come down to it, this outcome was just going to be too much again for that paper and so we chose to leave it and it's still a paper in the making, it's still data which is sitting there and hopefully will be a standalone paper in itself. I haven't got round to analysing it, I have gone out of a PhD and jumped straight into something else, I have still got data and all sorts of stuff that I haven't even have time to enter. I get the impression from talking to others as well that it's always common when you collect a lot of data to have one or two variables that never make it to the cutting board. So this outcome is not in my thesis. I would like to think that it will reach publication one day. I think at the moment I am very, I am very rushed with all the new things that I have got on my plate, but there is a number of variables that I breeze back on and think to myself 'oh I will do something about that one day, I will do something about that one day.' So yes I would like to think that it wasn't, it wasn't a wasted time in collecting it." | Outcome not analysed because volume of data presented in primary publication |
| 53 | "Right, we did that and just because it takes absolutely for ever to score it, and although I am a trained scorer as a primary investigator I can't do the scoring, so I had to send one of my team members to get trained and it just takes, it is taking forever to get through them all, so it's still in the pipeline. I would have had to postpone the writing of the primary outcome paper if I had to wait for this data. Yes, we have almost just about done now and so in a separate publication we will report that." | Delay in obtaining the data |
| 55 | "Doing the measurements is really hard and I wasn't sort of specifically trained to do them. I mean the clinicians told me how to do them, but it can be quite difficult to do. Which is a shame because that is a nice sort of, a nice measurement to have. We did it the whole way through but kind of knowing that it was pointless. I think, because we ended up doing it for the trial participants rather than for us, because they expected it to be done. So we just did it and noted it down but never analysed it, because we didn't believe in it, even if it had shown something brilliant, we wouldn't have thought it was true." | Not analysed because of practical difficulties and uncertainty about validity of data |
| 59 | "I think it had just fallen off our radar screen and we were focused on the primary outcome, because the story would have been 'you reduce the outcome, and a higher proportion of people remain independent.' In fact, we sort of then began telling another story. It didn't seem important once we had already frightened ourselves by showing that we caused harm and I suppose the story could have been 'you cause harm and you cause more people to have to go into care.' I think it just fell off the radar screen because we then began to worry about why, why this didn't work." | Not analysed because harmful effect of intervention on primary outcome |

and non-commercial sources. Discrepancies between primary outcomes specified in the protocol and those listed in the trial report were found in three trials. In one trial, the primary outcome as stated in the protocol was downgraded to a secondary outcome in the publication, and a non-primary outcome with statistically significant results ($P < 0.05$) was changed to the primary outcome. One trial omitted reporting three prespecified primary outcomes and included in the publication a new primary outcome ($P < 0.05$) not stated in the protocol. Both trials were not commercially funded. The remaining trial measured but did not analyse one of the eight prespecified primary outcomes; funding here came from both industry and non-commercial sources.

A total of 159 outcomes were prespecified across the 21 trials. Across the 21 trials, 26 (16%) of the 159 outcomes were not reported, of which four were prespecified primary outcomes in two trials. There was at least one unreported efficacy or harm outcome in 14 trials (67%), of which seven (64%) were not commercial trials and seven (70%) had joint industry and non-commercial funding. Unreported secondary outcomes were either not measured (two trials), measured but not analysed (12 trials), or measured and analysed but not reported (eight trials). Four trials reported five outcomes that were not prespecified.

Given the stated reasons for discrepancies, six (29%) of the cohort of 21 trials were found to have displayed outcome reporting bias—that is, non-reporting of an outcome was related to the results obtained in the analysis (trialists 39, 41, 44, 49, 50, and 56).

DISCUSSION

Researchers have a moral responsibility to report trial findings completely and transparently.¹⁶ Failure to communicate trial findings to the clinical and research

community is an example of serious misconduct.¹⁷ Previous research has shown that non-publication of whole studies or selected results is common, but there has been little previous research into the awareness among researchers of their responsibilities to publish fully their research findings or of the possible harm associated with selective publication.

Principal findings

When comparing the trial protocol with the subsequent publication, we carefully distinguished the cases where an outcome was measured and analysed but not reported, measured but not analysed or reported, and not measured. There are two key messages from our work. Firstly, most unreported outcomes were not reported because of a lack of a significant difference between the trial treatments, thus leading to an overoptimistic bias in the published literature as a whole. However, we believe that non-reporting mainly reflected a poor understanding of how any individual trial forms part of the overall evidence base, rather than intentional deception. Secondly, from the reasons given for outcomes being either measured but not analysed or not measured, there was a general lack of clarity about the importance and feasibility of data collection for the outcomes chosen at the time of protocol development.

We identified one study, however, in which the trialists intentionally chose to withhold data on a statistically significant increase in harm associated with the trial intervention. This type of "distorted reporting" in relation to adverse events has been previously described.¹⁸ The reported findings from the trial were published in a prestigious journal, and since publication the paper has been cited by others on a number of

occasions. We have strongly encouraged the trialist to make these unreported results publicly available.

We identified an apparent over reliance on the arbitrary but widely used cut off of 0.05 to distinguish between significant and non-significant results. In many cases trialists described subdividing P values into “significant” and “non-significant,” and deciding to not report “non-significant” results. Yet statistically significant results may not always be clinically meaningful. Likewise an observed difference between treatment groups might be clinically important, yet fail to reach statistical significance. Whether or not a finding is reported should not depend on its P value.

Comparison with other studies

Our work provides richer information on trialists’ decision making regarding selective outcome reporting than previous work has been able to capture. Previous questionnaire surveys asked trialists about any unpublished outcomes, whether those showed significant differences, and why those outcomes had not been published.^{4,8,15} Our work supports their findings and provides a more detailed understanding of why trialists do not report outcomes. Previous studies did not establish whether or not trialists intentionally chose to mislead by not reporting non-significant outcomes. Our study suggests deliberate misrepresentation of the results may be rare, although we cannot rule out the possibility that trialists may have framed their responses to our interview questions to hide intentional under-reporting of outcomes.

An ongoing Cochrane review has identified studies that compare trial protocols with subsequent publications (K Dwan, personal communication, 2010). The effect of journal type was examined in one study but was not found to be associated with selective outcome reporting.¹² The effect of funding source on outcome reporting bias seems to be heterogeneous, with studies showing that commercially funded trials are more (D Gherzi, personal communication, 2010), less,¹⁹ or equally²⁰ likely to display inconsistencies in the reporting of primary outcomes compared with studies that were not commercially funded. One study reported no effect of funding source on whether outcomes were completely reported, after adjusting for statistical significance (E von Elm, personal communication, 2010), whereas another found that commercially funded trials were more likely to completely report outcomes (D Gherzi, personal communication, 2010). For the 85 studies suspected of outcome reporting bias and the 98 studies assumed to be free of outcome reporting bias in this project, there was little difference between the groups in terms of sample size or the involvement of a statistician. However, there was some evidence that trials suspected of outcome reporting bias were more likely to not state their source of funding. The implications of our low response rate among trialists involved in commercial trials is not clear. It is possible that our results for the prevalence of outcome reporting bias are conservative, if those declining to

be interviewed did so to avoid exposure of such bias with their studies.

Recent work has shown the harmful effect of selective outcome reporting on the findings of systematic reviews³ and also the lack of awareness among people who conduct systematic reviews of this potential problem. Two recent initiatives are relevant here: firstly, the inclusion of selective outcome reporting as a key component of the new Cochrane risk of bias tool,²¹ which all review authors should be using in future to assess the trials included in a Cochrane review; and secondly, the recommendation to consider this issue when reporting a systematic review.²¹ We hope these developments will lead to greater recognition by reviewers and users of systematic reviews of the potential impact of reporting bias on review conclusions.

Strengths and limitations of the study

The strength of our study rests in its use of interviews to provide valuable insight into the decision making process from the planning to the publication stage of clinical trials. Our study adds to the growing body of empirical evidence on research behaviour.²² Outcome reporting bias and publication bias more broadly could be viewed as forms of “fabrication,” according to the definition given by the US Office of Research Integrity. Our study leads us to agree with Martinson et al that “mundane ‘regular’ misbehaviours present greater threats to the scientific enterprise than those caused by high profile misconduct cases such as fraud.”²³ Outcome reporting bias and publication bias more broadly were not explicitly mentioned in Martinson et al’s survey, however, suggesting that the scientists from whom the authors drew the specific misbehaviours to be surveyed were unaware of the harmful effects of selective reporting.

It is important to acknowledge that the results presented come only from those trialists who agreed to be interviewed, and thus we urge caution in the interpretation of our study findings given the low response rate (trialists who declined our invitation or could not send their trial protocol were not interviewed). There is limited evidence relating to the factors associated with selective outcome reporting and even less with respect to outcome reporting bias directly, making an assessment of response bias in this current study difficult.

It is possible that different methods—such as postal questionnaires—might have resulted in a higher proportion of trialists agreeing to be interviewed. However, structured telephone interviews were considered the most appropriate form of data collection when taking into account the aims of our study. The style of the interview allowed similar topics to be covered with each of the trialists, while still allowing the trialists to have flexibility in their answers. It allowed the prespecified topics to be discussed and explored in detail, as well as new areas or ideas to be uncovered. In addition, such interviews allowed the interviewer to check that they had understood the respondent’s meaning, instead of relying on their own assumptions.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Outcome reporting bias is the selection for publication of a subset of the original recorded outcomes on the basis of the results

Outcome reporting bias has been identified as a threat to evidence based medicine because clinical trial outcomes with statistically significant results are more likely to be published

The extent of outcome reporting bias in published trials has been previously investigated using follow-up surveys of authors, but these studies found that trialists often did not acknowledge the issue of unreported outcomes even when there was evidence of their existence in their publications

WHAT THIS STUDY ADDS

The prevalence of incomplete outcome reporting is high

This study has, for the first time, provided a detailed understanding of why trialists do not report previously specified outcomes

Trialists seem to be generally unaware of the implications for the evidence base of not reporting all outcomes and protocol changes

There is a general lack of consensus regarding the choice of outcomes in particular clinical settings, which can affect trial design, conduct, analysis, and reporting

A further limitation of our work is the small sample size. All trialists eligible for interview were invited (n=268), of whom 59 (22%) finally agreed. Completion of interviews was governed by pragmatic reasons (the number of trialists who ultimately agreed), rather than when there was no extra data gained from the latest interview (data saturation).

Conclusions and policy implications

We are the first group to conduct interviews targeting instances of outcome reporting bias and protocol changes. Our findings add important new insights, pointing to a lack of awareness and understanding of the problem generally. Trialists seemed unaware of the implications for the evidence base of not reporting all outcomes and protocol changes. A general lack of consensus regarding the choice of outcomes in particular clinical settings was evident and affected trial design, conduct, analysis, and reporting.

It has been suggested that comparing publications with the original protocol constitutes the most reliable means of evaluating the conduct and reporting of research,²⁴ and the Cochrane Collaboration has recommended such comparison to detect if complete or incomplete reporting is indicated.²¹ It is widely believed that the introduction of trial registration would remove the practice of retrospective changes to outcomes, but a recent study has shown that there are frequent discrepancies between outcomes included on trial registers and those reported in subsequent publications.¹² We have found that some differences between the protocol and the publication are a result of outcome reporting bias, but also that some changes are made during the course of the trial as a result of poor outcome choice at the protocol development stage. Such differences do not lead to bias. Our work suggests that for the assessment of reporting bias, reliance on the protocol alone without contacting trial authors will often be inadequate.

Our findings emphasise the need to improve trial design and reporting. The recent update of the CONSORT reporting guideline²⁵ includes an explicit request that authors report any changes to trial outcomes after the trial commenced and provide reasons for such changes. Readers can then judge the potential for bias. Unfortunately, in our study none of the reports in which trial outcomes were not measured, or were measured but not analysed, acknowledged that protocol modifications had been made.

Our finding of a lack of understanding of the importance of reporting outcomes in various clinical areas lends further support to the development of core outcome sets.²⁶⁻²⁸ International efforts to develop consensus on a minimum set of outcomes that should be measured and reported in later phase trials in a particular clinical area would lead to an improved evidence base for healthcare.

Contributors: The study was conceived by PRW, DGA, CG, and AJ. PRW, AJ, DGA, CG, and RMDS designed the study protocol. RMDS contacted the trialists to conduct the interviews. A comparison of trial protocols with trial reports was carried out by RMDS and JJK. Interviews were performed by RMDS. Analysis was performed by RMDS and JJK, with input and supervision from PRW and AJ. RMDS prepared the initial manuscript; PRW, AJ, DGA, and CG were involved in revisions of this manuscript. All authors commented on the final manuscript before submission. PRW is the guarantor for the project.

Funding: The ORBIT (Outcome Reporting Bias In Trials) project was funded by the Medical Research Council (grant number G0500952). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of this manuscript. DGA is supported by Cancer Research UK.

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; and no other relationships or activities that could appear to have influenced the submitted work.

Data sharing: No additional data available.

- Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H Jr. Publication bias and clinical trials. *Control Clin Trials* 1987;8:343-53.
- Hutton JL, Williamson PR. Bias in meta-analysis due to outcome variable selection within studies. *Appl Stat* 2000;49:359-70.
- Kirkham JJ, Dwan KM, Altman DA, Gamble C, Dodd S, Smyth RMD, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010;340:c365.
- 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.
- Williamson PR, Gamble C, Altman DG, Hutton JL. Outcome selection bias in meta-analysis. *Stat Methods Med Res* 2005;14:515-24.
- Dwan K, Altman DG, Amaiz JA, Bloom J, Chan AW, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS One* 2008;3:e3081.
- Hahn S, Williamson PR, Hutton JL. Investigation of within-study selective reporting in clinical research: follow-up of applications submitted to a local research ethics committee. *J Eval Clin Pract* 2002;8:353-9.
- Chan AW, Kroleza-Jeric K, Schmid J, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *CMAJ* 2004;171:735-40.
- Ghera D. Issues in the design, conduct and reporting of clinical trials that impact on the quality of decision making. PhD thesis. School of Public Health, Faculty of Medicine, University of Sydney, 2006.
- Von Elm E, Rollin A, Blumle A, Huwiler K, Witschi M, Egger M. Publication and non-publication of clinical trials: longitudinal study of applications submitted to a research ethics committee. *Swiss Med Wkly* 2008;138:197-203.
- 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.

- 12 Mathieu S, Boutron I, Moher D, Altman DG, Ravaud P. Comparison of registered and published primary outcomes in randomized controlled trials. *JAMA* 2009;302:977-84.
- 13 Al-Marzouki S, Roberts I, Evans S, Marshall T. Selective reporting in clinical trials: analysis of trial protocols accepted by *The Lancet*. *Lancet* 2008;372:201.
- 14 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.
- 15 Chan AW, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ* 2005;330:753.
- 16 Moher D. Reporting research results: a moral obligation for all researchers. *Can J Anaesth* 2007;54:331-5.
- 17 Chalmers I. Underreporting research is scientific misconduct. *JAMA* 1990;263:1405-8.
- 18 Ioannidis JP. Adverse events in randomized trials: neglected, restricted, distorted, and silenced. *Arch Intern Med* 2009;169:1737-9.
- 19 Bourgeois FT, Murthy S, Mandl KD. Outcome reporting among drug trials registered in ClinicalTrials.gov. *Ann Intern Med* 2010;153:158-66.
- 20 Ewart R, Lausen H, Millian N. Undisclosed changes in outcomes in randomized controlled trials: an observational study. *Ann Fam Med* 2009;7:542-6.
- 21 Higgins JPT. Cochrane handbook for systematic reviews of interventions. Version 5.0.2. The Cochrane Collaboration, 2009.
- 22 Fanelli D. How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS One* 2009;4:e5738.
- 23 Martinson BC, Anderson MS, de Vries R. Scientists behaving badly. *Nature* 2005;435:737-8.
- 24 Chan AW, Upshur R, Singh JA, Ghersi D, Chapuis F, Altman DG. Research protocols: waiving confidentiality for the greater good. *BMJ* 2006;332:1086-9.
- 25 Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c869.
- 26 Cooney RM, Warren BF, Altman DG, Abreu MT, Travis SP. Outcome measurement in clinical trials for ulcerative colitis: towards standardisation. *Trials* 2007;8:17.
- 27 Clarke M. Standardising outcomes for clinical trials and systematic reviews. *Trials* 2007;8:39.
- 28 Sinha I, Jones L, Smyth RL, Williamson PR. A systematic review of studies that aim to determine which outcomes to measure in clinical trials in children. *PLoS Med* 2008;5:e96.

Accepted: 5 October 2010