

RESEARCH ARTICLE

How Much Participant Outcome Data Is Missing from Sight: Findings from a Cohort of Trials Submitted to a German Research Ethics Committee

Jamie J. Kirkham^{1*}, Kerry M. Dwan², Anette Blümle³, Erik von Elm⁴, Paula R. Williamson¹

1 MRC North West Hub for Trials Methodology Research, Department of Biostatistics, University of Liverpool, Liverpool, United Kingdom, **2** Cochrane Editorial Unit, London, United Kingdom, **3** Cochrane Germany, Faculty of Medicine, University of Freiburg, Freiburg, Germany, **4** Cochrane Switzerland, Institut Universitaire de Médecine Sociale et Préventive, Lausanne, Switzerland

* jjk@liv.ac.uk



OPEN ACCESS

Citation: Kirkham JJ, Dwan KM, Blümle A, von Elm E, Williamson PR (2016) How Much Participant Outcome Data Is Missing from Sight: Findings from a Cohort of Trials Submitted to a German Research Ethics Committee. PLoS ONE 11(6): e0157883. doi:10.1371/journal.pone.0157883

Editor: Martin Voracek, University of Vienna, AUSTRIA

Received: December 18, 2015

Accepted: June 5, 2016

Published: June 17, 2016

Copyright: © 2016 Kirkham et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The original work was supported by the German Research Foundation (DFG), grant number: EL 544/1-1. JJK is funded by the University of Liverpool. PRW is funded by the MRC North West Hub for Trials Methodology Research (Grant Reference Number: MR/K025635/1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

Background

Study publication bias and outcome reporting bias have been recognised as two threats to the validity of systematic reviews. The purpose of this research was to estimate the proportion of missing participant outcome data from randomised controlled trials (RCTs) due to lack of publication of whole studies and due to outcome data missing within study publications.

Methods and Findings

Data were extracted from protocols of clinical research projects submitted to the research ethics committee of the University of Freiburg (Germany) between 2000 and 2002 and associated fully published articles. The total amount of published and unpublished outcome data from all trial participants was calculated for each trial and the overall proportion of missing data from both unpublished and published trials computed. Full and partially reported outcome data was also taken into consideration. The impact of funding source on missingness was also considered at the trial level. From 308 parallel group trials in the study cohort, 167 were published and 141 were unpublished. Overall, 260,563 participants contributed to a total of 2,618,116 participant outcome data across all trials. About half (47%) of the participant outcome data from the 308 trials was reported in full but at least 81% were partially reported. Of the 19% of participant data that were missing, 4% was attributable to missing data from published trials and 15% from unpublished trials. Commercially funded trials had a higher probability of publication (relative risk 1.20, 95% confidence interval 0.86, 1.67; $p = 0.27$) but were less likely to fully report all outcomes than non-commercially funded trials (relative risk 0.64, 95% confidence interval 0.30, 1.38; $p = 0.26$).

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

Conclusions

Missing participant outcome data from both published and unpublished trials is frequent. Clinical trial registration including outcome information not only identifies that clinical trials exist but the systematic examination and monitoring of trial information within a registry can help detect selective reporting of entire studies and of outcome data within studies and possibly prevent it.

Background

Study publication bias and within-study outcome reporting bias have been recognised as two forms of reporting bias that can affect evidence based practice. There is strong evidence of an association between direction of results and publication status; studies that report positive or significant results are more likely to be published and outcomes that are statistically significant have higher odds of being fully reported (range of odds ratios: 2.2 to 4.7) [1,2]. As a result of these two forms of bias, systematic reviews summarising the available evidence are likely to overestimate the effect of treatment [3,4]. Reporting bias may therefore result in inappropriate health care decisions by policy makers, clinicians and patients, which potentially harm patients, waste resources, and misguide future research. While there is growing empirical evidence for the existence of these biases, little is known about the relative size of these two related problems affecting the available research evidence. Using data taken from protocols of clinical research projects submitted to the research ethics committee of the University of Freiburg (Germany) and associated fully published articles, we estimate the proportion of missing participant outcome data due to lack of publication of the study and the proportion due to missing outcome data within a published study.

Methods

Data source

The data used in this study has previously been identified and evaluated in other publications [5–7]. In brief, the Research Ethics Committee (REC) of the Albert-Ludwigs-University, Freiburg, Germany granted access to the electronic files of all study protocols (including amendments) submitted between 2000 and 2002. Nearly half (408/917; 44%) of the studies approved by the REC Freiburg within this time period were randomised controlled trials (RCTs). Of the 355 RCTs that started, the majority were multicentre trials and in 12%, the primary centre was located in Freiburg. In this research article we consider only the 311 RCTs with parallel group design that started. Trials that never started were excluded since there was no expectation that the trial results would be published. Full study publications were identified through a systematic search of electronic literature databases and verification via contact with the trial authors. Unpublished trials were defined to be trials where it was confirmed by trial authors that the trial was completed or discontinued and there was no full study publication. To ensure that the current dataset was as up-to-date as possible, the publication status of ongoing trials (at the time of the original study [5]) was verified again in July 2015. This enabled us to classify all trials as either published or unpublished.

Data extraction

The total number of trials in the study cohort was separated into those that were published and those that were unpublished. For published trials, the total number of participants was extracted from the trial publication which represented the actual sample size achieved (i.e. the total number randomised), the largest of which was taken if sample sizes differed across multiple publications of the same trial. For unpublished trials, as an approximation to what was actually collected (but not published), planned sample sizes were extracted from the trial protocol. The sample size was assumed to be the same for all trial outcomes. For each trial, published or unpublished, we also extracted the total number of pre-specified outcomes listed in the trial protocol. For published trials, we recorded the total number of published and unpublished outcomes taken from the trial publication.

Published outcomes were categorised into 'full' reporting and 'partial' reporting. An outcome was considered to be fully reported if sufficient information was provided such that the outcome could be included in a potential meta-analysis of a systematic review in accordance to the definitions provided in [8]. Partially reported outcomes were defined as those that were inadequately reported for inclusion into a review meta-analysis (for example, an effect size was presented with no measure of precision or exact p-value). An extended list of the different types of partial reporting can be found in the data supplement (Web table B) of the main ORBIT (Outcome Reporting Bias in Trials) manuscript [3]. The highest order of reporting was taken for each outcome if multiple time-points were considered, for example, if an outcome was partially reported at 3 and 6 months and fully reported at 12 months, then the outcomes was considered to be fully reported. The source of funding was also extracted for each trial. Funding source was categorised into commercially funded (typically funded by pharmaceutical industry), non-commercially funded (typically funded by research councils, research foundations, government ministries and charities) and trials where no source of funding was declared. Trials were classified as commercially funded if they were partially funded by industry but were categorised as non-commercially funded if a company provided the trial drugs (wholly or in part) but had no role in the study design, data collection and analysis.

Data extraction from the original protocols and publications was undertaken in Freiburg in which AB contributed. Using the database of extracted data, the required information for this trial was data extracted independently by two researchers (KD and JJK). Any discrepancies between the information extracted by the two reviewers were resolved through discussion and in consultation with AB. Sample sizes and a summary of the outcome status' (fully reported, unreported and partially reported) for each individual trial can be found in the supporting information file ([S1 File](#)).

Data analyses

Total amount of reported and missing participant outcome data (participant data level). The total amount of participant outcome data from each trial was calculated to be the planned trial sample size multiplied by the total number of pre-specified outcomes. For the study cohort, we computed the total amount of participant data that was fully reported, partially reported and missing from the published trials and the total amount of data that was missing as a result of trials being unpublished. Using the computed metrics summed across all trials and taking the total amount of outcome data across all trials as the denominator, we report the following proportions:

- Proportion of fully published data
- Proportion of partially reported data

- Proportion of fully and partially reported data
- Proportion of missing data from published trials
- Proportion of missing data from unpublished trials
- Proportion of missing data from all trials (published and unpublished).

Missing data from published trials may be due to within study selective reporting. In our computations, we also considered partially reported outcome data as unpublished data. The motivation here is that, although partially reported outcome data are published, they cannot be included into a meta-analysis without obtaining further information (which is usually only available from the authors), and therefore effectively resembles missing data in the systematic review process.

Impact of funding source (trial level). The relative risk of full publication was compared between commercially funded trials and non-commercially funded trials. For the published trials, we also computed the relative risk of fully reporting all outcomes in commercially funded trials compared to non-commercially funded trials.

Results

The study cohort contained 311 RCTs of which three were excluded as they presented no results for the comparator group. This is a high risk of bias form of selective reporting which has previously been documented [3], and our decision to exclude here is that our missing data computations would mask this serious reporting problem by simply reducing the sample sizes of these trials to what was either reported in the treatment arm or not reported in the comparator arm. Of the remaining 308 trials, 167 were published (54%) and 141 (46%) were unpublished. There was a tendency for published trials to specify more outcomes in their protocols than unpublished trials (Table 1). Further, the median number of recruited participants was higher than the number of planned participants in unpublished trials, and a higher proportion of published trials were commercially funded (Table 1). The majority of trials (published and unpublished) were multi-centre trials and of these multi-centre trials, most were international trials (Table 1). The primary centre for the multi-centre trials was Freiberg for 12% of published trials and 14% for unpublished trials.

The 308 published and unpublished trials contained data on 3407 distinct outcomes of 260,563 participants (Table 1). From the published trials 70% (1595/2253) of the outcomes were published in full, 7% (167/2253) of the outcomes went unreported while the remainder were partially reported. Clearly all 1154 outcomes from the unpublished studies went unreported.

At the individual participant level, the total amount of participant outcome data was 2,618,116 (Table 2). Under a half of this data (47%) was fully reported but at least 81% were partially reported. Four percent of the data was missing (within-study selective outcome reporting) from published trials and 15% was missing from entirely unpublished trials.

Sixty percent (119/198) of commercially funded trials were published while 50% (20/40) of non-commercially funded trials were published. A commercially funded trial had a 20% increased probability of being published compared to a non-commercially funded trial although this difference did not reach statistical significance (relative risk 1.20, 95% confidence interval (0.86, 1.67; $p = 0.27$)). This wide confidence interval suggested that there was a high level of uncertainty in this finding.

Of the published trials, nearly a fifth (23/119; 19%) of the commercially funded trials fully reported all pre-specified outcomes compared to 30% (6/20) of the non-commercially funded

Table 1. Outcome, participant and funding characteristics of trials that were published and unpublished.

	All n = 308 (%) ^a	Published n = 167 (%) ^a	Unpublished n = 141 (%) ^a
Outcomes			
Total number of outcomes:	3407	2253	1154
Median (per trial):	10	11	7
Interquartile range:	5 to 15	7 to 17	4 to 11
Trial participants			
Total number of participants:	260,563	210,191	50,372
Median (per trial):	200	210	180
Interquartile range:	80 to 463	87 to 62	78 to 320
Funding			
Commercial:	198 (64)	119 (71)	79 (56)
Non-commercial:	40 (13)	20 (12)	20 (14)
Not stated:	70 (23)	28 (17)	42 (30)
Collaboration			
Single-centre:	65 (21)	28 (17)	37 (26)
Multi-centre:	243 (79)	139 (83)	104 (74)
Only multi-centre studies^b			
International:	162 (67)	107 (77)	55 (53)
National:	80 (33)	32 (23)	48 (46)
Unclear:	1 (<1)	0	1 (1)
Primary centre^b			
Freiburg:	31 (13)	16 (12)	15 (14)
Other:	211 (87)	123 (88)	88 (85)
Unclear:	1 (<1)	0	1 (1)

^acolumn percentages;

^bdenominator is number of multi-centre studies

doi:10.1371/journal.pone.0157883.t001

Table 2. Proportion of reported and missing participant outcome data.

1) Type of data (source)	Total amount	
Participant data (published and unpublished studies)	2,618,116	[A]
Fully reported data (published studies)	1,233,291	[B]
Partially reported data (published studies)	892,892	[C]
Missing data (published studies)	91,381	[D]
Unpublished data (unpublished studies)	400,552	[E]
2) Proportions of total amount of outcome data		
Fully published data	47%	[B]/[A]
Partially reported data	34%	[C]/[A]
Proportion of fully and partially reported data	81%	[B+C]/[A]
Missing data from published studies (<i>within-study selective outcome reporting</i>)	4%	[D]/[A]
Missing data from unpublished studies (<i>selective reporting of entire studies</i>)	15%	[E]/[A]
Missing data from all studies	19%	[D+E]/[A]
Missing data from all studies (when partially reported was considered as unpublished data)	53%	[C+D+E]/[A]

doi:10.1371/journal.pone.0157883.t002

trials. Commercially funded trials were less likely to fully report all outcomes than non-commercially funded trials although this difference again did not reach statistical significance with a wide confidence interval denoting a high level of uncertainty in this finding (relative risk 0.64, 95% confidence interval (0.30, 1.38; $p = 0.26$)).

Discussion

Our study has shown that (81%) of participant outcome data from trials is either reported in full or partially reported. Direct empirical evidence for the existence of study publication bias and outcome reporting bias has been summarized previously but [1], to our knowledge, this study is novel seeking to quantify how much outcome data is missing from trials and how much of this missingness is attributable to lack of publication of entire studies or of outcomes from studies for which other data has been published. Missing data from published studies which could be associated with within-study selective outcome reporting appeared to be less of a problem than missing data from entirely unpublished studies. Nevertheless we have shown that the amount of missing data from both published and unpublished trials is more than half (53%) if partially reported data is treated as unpublished. Considering partially reported data as unpublished reflects the fact that such data cannot be included in a meta-analysis right away. In a survey, most Cochrane review authors related that they requested additional unpublished data from trialists but only about half of them obtain them [9]. Furthermore, partial reporting has previously been associated with a high risk of outcome reporting bias [3,10]. Our study is also likely to underestimate the impact of partial reporting, since outcomes with multiple time-points were considered to be fully reported if just one time-point was reported in full.

Strengths and limitations of this study

We evaluated a large, unselected cohort of trials, the investigators confirmed whether or not the trial was published if no trial publication was found by us, and two study authors independently extracted and verified all the trial data. A limitation of our study is that we assumed that the sample sizes from the published reports were the same for all trial outcomes. This assumption may not always be realistic and may conflate attrition with the failure to report. However, sample sizes for individual outcomes are not routinely reported for all trials thus our approach could be consistently applied across all trials. All the trials included in the study were identified from protocols submitted to a German REC and hence some of the final trial publications were subsequently published in German language journals. It is possible that the results may therefore not fully represent cohorts of trials from other RECs. Previously, we identified four additional studies that covered the entire process from the publication of the study protocol to the publication of the study outcomes that could have also been used as potential data sources for comparison [1]. However, at the time of this original review, these datasets were unavailable after several requests were made to the study authors and it remains unclear as to whether these studies extracted data on sample sizes needed for inclusion into this current research. These other studies were approved by the Canadian Institute of Health Research (Canada) [11], the Scientific-Ethical Committees for Copenhagen and Frederiksberge (Denmark) [8], Central Sydney Area Health Service Ethics Review Committee (Australia) [12] and Local Research Ethics Committee (UK) [13].

Conclusions

The full publication of outcome data from clinical trials helps avoid unnecessary duplication of research while clinical trial registration allows the relevant stakeholders identify that a clinical trial exists. The systematic recording and monitoring of trial information within a registry goes

some way to prevent and detect the selective publication and selective reporting of clinical research outcomes.

The large proportion of partially reported outcome data suggests that journal editors and reviewers need to more closely monitor and address this problem. Relevant reporting guidelines clearly advocate for complete reporting of outcome data and are endorsed by many journals [14]. However, routine checks of the essential items to be reported are often still not implemented in peer review and editorial procedures. As summarized in recent systematic reviews of non-publication and reporting biases [1] there is solid evidence confirming both areas of problem [2]. Future studies need to explore ways how the known underreporting of outcome data can be tackled effectively e.g. by creating incentives for complete and accurate reporting of outcome data from clinical trials and for making raw trial data more accessible.

Ethical approval

No ethics committee opinion was required for this study.

Supporting Information

S1 File. Samples sizes and outcome summaries for all published and unpublished trials. (XLSX)

Acknowledgments

We are grateful to the research ethics committee of the Albert-Ludwigs-Universität Freiburg/Germany, its administrative staff, and the local investigators who responded to our survey. We also thank Alexander Hellmer and Florian Volz for help with extraction of protocol data, and Patrick Oeller for updating the literature search.

Author Contributions

Conceived and designed the experiments: PRW. Performed the experiments: KMD JJK AB. Analyzed the data: KMD JJK AB EvE PRW. Wrote the paper: JJK KMD AB EvE PRW.

References

1. Dwan KM, Gamble C, Williamson PR, Kirkham JJ (2013) Systematic Review of the Empirical Evidence of Study Publication Bias and Outcome Reporting Bias—an updated review. *PLoS ONE*; 8(7): e66844. doi: [10.1371/journal.pone.0066844](https://doi.org/10.1371/journal.pone.0066844) PMID: [23861749](https://pubmed.ncbi.nlm.nih.gov/23861749/)
2. Schmucker C, Schell LK, Portalupi S, Oeller P, Cabrera L, Bassler D, et al (2014). Extent of Non-Publication in Cohorts of Studies Approved by Research Ethics Committees or Included in Trial Registries. *PLoS ONE*; 9(12): e114023. doi: [10.1371/journal.pone.0114023](https://doi.org/10.1371/journal.pone.0114023) PMID: [25536072](https://pubmed.ncbi.nlm.nih.gov/25536072/)
3. Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, et al. (2010). The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ*; 340:c365. doi: [10.1136/bmj.c365](https://doi.org/10.1136/bmj.c365) PMID: [20156912](https://pubmed.ncbi.nlm.nih.gov/20156912/)
4. Egger M, Smith GD. Bias in location and selection of studies. *BMJ* 1998; 316 (7124):61–6. PMID: [9451274](https://pubmed.ncbi.nlm.nih.gov/9451274/)
5. Blümle A, Antes G, Schumacher M, Just H, von Elm E (2008). Clinical research projects at a German medical faculty: follow-up from ethical approval to publication and citation by others. *J. Med. Ethics*; 34: e20. doi: [10.1136/jme.2008.024521](https://doi.org/10.1136/jme.2008.024521) PMID: [18757621](https://pubmed.ncbi.nlm.nih.gov/18757621/)
6. Blümle A, Meerpohl JJ, Schumacher M, von Elm E (2014). Fate of Clinical Research Studies after Ethical Approval—Follow-Up of Study Protocols until Publication. *PLoS ONE*; 9(2):e87184. doi: [10.1371/journal.pone.0087184](https://doi.org/10.1371/journal.pone.0087184) PMID: [24586265](https://pubmed.ncbi.nlm.nih.gov/24586265/)
7. Blümle A, Meerpohl JJ, Rucker G, Antes G, Schumacher M, von Elm E (2011). Reporting of eligibility criteria of randomised trials: cohort study comparing trial protocols with subsequent articles. *BMJ*; 342: d1828. doi: [10.1136/bmj.d1828](https://doi.org/10.1136/bmj.d1828) PMID: [21467104](https://pubmed.ncbi.nlm.nih.gov/21467104/)

8. Chan A-W, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG (2004). Empirical evidence for selective reporting of outcomes in randomized trials—: comparison of protocols to published articles. *JAMA*; 291(20): 2457–2465. PMID: [15161896](#)
9. Schroll JB, Bero L, Gøtzsche PC (2013). Searching for unpublished data for Cochrane reviews: cross sectional study. *BMJ*; 346: f2231 doi: [10.1136/bmj.f2231](#) PMID: [23613540](#)
10. Saini P, Loke YK, Gamble C, Altman DG, Williamson PR, Kirkham JJ(2014). Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews. *BMJ*; 349:g6501. doi: [10.1136/bmj.g6501](#) PMID: [25416499](#)
11. Chan AW, Krolez-Jerić K, Schmid I, Altman DG (2004). Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *CMAJ*; 171(7):735–40. PMID: [15451835](#)
12. Gherzi D (2006). Issues in the design, conduct and reporting of clinical trials that impact on the quality of decision making: University of Sydney.
13. Hahn S, Williamson PR, Hutton JL (2002). Investigation of within-study selective reporting in clinical research: follow-up of applications submitted to a local research ethics committee. *J Eval Clin Pract* 8: 353–359. PMID: [12164983](#)
14. Schulz KF, Altman DG, Moher D, for the CONSORT Group (2010). CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *PLoS Medicine*; 7(3):e1000251. doi: [10.1371/journal.pmed.1000251](#) PMID: [20352064](#)