

# Systematic Review of the Empirical Evidence of Study Publication Bias and Outcome Reporting Bias

Kerry Dwan<sup>1\*</sup>, Douglas G. Altman<sup>2</sup>, Juan A. Arnaiz<sup>3</sup>, Jill Bloom<sup>4</sup>, An-Wen Chan<sup>5</sup>, Eugenia Cronin<sup>6</sup>, Evelyne Decullier<sup>7</sup>, Philippa J. Easterbrook<sup>8</sup>, Erik Von Elm<sup>9,10</sup>, Carrol Gamble<sup>1</sup>, Davina Gherzi<sup>11</sup>, John P. A. Ioannidis<sup>12,13</sup>, John Simes<sup>14</sup>, Paula R. Williamson<sup>1</sup>

**1** Centre for Medical Statistics and Health Evaluation, University of Liverpool, Liverpool, United Kingdom, **2** Centre for Statistics in Medicine, University of Oxford, Oxford, United Kingdom, **3** Clinical Pharmacology Unit, UASP Hospital Clinic, Barcelona, Spain, **4** Moorfields Eye Hospital, London, United Kingdom, **5** Randomized Controlled Trials Unit, Canadian Institutes of Health Research, Ottawa, Canada, **6** Healthier Communities/Public Health, Greenwich Council, London, England, **7** Clinical Epidemiology Unit, DIM-Hospices Civils de Lyon, Lyon, France, **8** Department of HIV/GUM, King's College London, London, United Kingdom, **9** Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland, **10** German Cochrane Centre, Department of Medical Biometry and Medical Informatics, University Medical Centre Freiburg, Freiburg, Germany, **11** NHMRC Clinical Trials Centre, Camperdown, Australia, **12** Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece, **13** Institute for Clinical Research and Health Policy Studies, Department of Medicine, Tufts Medical Center, Tufts University School of Medicine, Boston, Massachusetts, United States of America, **14** National Health and Medical Research Council (NHMRC) Clinical Trials Centre, University of Sydney, Sydney, Australia

## Abstract

**Background:** The increased use of meta-analysis in systematic reviews of healthcare interventions has highlighted several types of bias that can arise during the completion of a randomised controlled trial. Study publication bias has been recognised as a potential threat to the validity of meta-analysis and can make the readily available evidence unreliable for decision making. Until recently, outcome reporting bias has received less attention.

**Methodology/Principal Findings:** We review and summarise the evidence from a series of cohort studies that have assessed study publication bias and outcome reporting bias in randomised controlled trials. Sixteen studies were eligible of which only two followed the cohort all the way through from protocol approval to information regarding publication of outcomes. Eleven of the studies investigated study publication bias and five investigated outcome reporting bias. Three studies have found that statistically significant outcomes had a higher odds of being fully reported compared to non-significant outcomes (range of odds ratios: 2.2 to 4.7). In comparing trial publications to protocols, we found that 40–62% of studies had at least one primary outcome that was changed, introduced, or omitted. We decided not to undertake meta-analysis due to the differences between studies.

**Conclusions:** Recent work provides direct empirical evidence for the existence of study publication bias and outcome reporting bias. There is strong evidence of an association between significant results and publication; 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. Publications have been found to be inconsistent with their protocols. Researchers need to be aware of the problems of both types of bias and efforts should be concentrated on improving the reporting of trials.

**Citation:** Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan A-W, et al. (2008) Systematic Review of the Empirical Evidence of Study Publication Bias and Outcome Reporting Bias. PLoS ONE 3(8): e3081. doi:10.1371/journal.pone.0003081

**Editor:** Nandi Siegfried, Medical Research Council South Africa, South Africa

**Received:** December 7, 2007; **Accepted:** June 20, 2008; **Published:** August 28, 2008

**Copyright:** © 2008 Dwan et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work forms part of the first author's PhD, funded by the NICE Liverpool Reviews and Implementation Group. Douglas Altman is supported by Cancer Research UK. Funders were not involved in the work.

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

\* E-mail: kerry.dwan@liverpool.ac.uk

## Introduction

Study publication bias arises when studies are published or not depending on their results; it has received much attention [1,2]. Empirical research consistently suggests that published work is more likely to be positive or statistically significant ( $P < 0.05$ ) than unpublished research [3]. Study publication bias will lead to overestimation of treatment effects; it has been recognised as a threat to the validity of meta-analysis and can make the readily available evidence unreliable for decision making. There is additional evidence that research without statistically significant results takes longer to achieve publication than research with

significant results, further biasing evidence over time [4–6,29]. This “time lag bias” (or “pipeline bias”) will tend to add to the bias since results from early available evidence tend to be inflated and exaggerated [7,8].

Within-study selective reporting bias relates to studies that have been published. It has been defined as the selection on the basis of the results of a subset of the original variables recorded for inclusion in a publication [9]. Several different types of selective reporting within a study may occur. For example, selective reporting of analyses may include intention-to-treat analyses versus per-protocol analyses, endpoint score versus change from baseline, different time points or subgroups [10]. Here we focus on

the selective reporting of outcomes from those that were originally measured within a study; outcome reporting bias (ORB).

Randomised controlled trials (RCTs) are planned experiments, involving the random assignment of participants to interventions, and are seen as the gold standard of study designs to evaluate the effectiveness of a treatment in medical research in humans [11]. The likely bias from selective outcome reporting is to overestimate the effect of the experimental treatment.

Researchers have considered selective outcome reporting to be a major problem, and deserving of substantially more attention than it currently receives [12]. Recent work [13–19] has provided direct empirical evidence for the existence of outcome reporting bias. Studies have found that statistically significant results had a higher odds of being fully reported compared to non-significant results for both efficacy and harm outcomes. Studies comparing trial publications to protocols are accumulating evidence on the proportion of studies in which at least one primary outcome was changed, introduced, or omitted.

Thus, the bias from missing outcome data that may affect a meta-analysis is on two levels: non-publication due to lack of submission or rejection of study reports (a study level problem) and the selective non-reporting of outcomes within published studies on the basis of the results (an outcome level problem). While much effort has been invested in trying to identify the former [2], it is equally important to understand the nature and frequency of missing data from the latter level.

The aim of this study was to review and summarise the evidence from empirical cohort studies that have assessed study publication bias and/or outcome reporting bias in RCTs approved by a specific ethics committee or other inception cohorts of RCTs.

## Methods

### Study inclusion criteria

We included research that assessed an inception cohort of RCTs for study publication bias and/or outcome reporting bias. We focussed on inception cohorts with study protocols being registered before start of the study as this type of prospective design were deemed more reliable. We excluded cohorts based on prevalence archives, in which a protocol is registered after a study is launched or completed, since such cohorts can already be affected by publication and selection bias.

Both cohorts containing exclusively RCTs or containing a mix of RCTs and non-RCTs were eligible. For those studies where it was not possible to identify the study type (i.e. whether any included studies were RCTs), we attempted to contact the authors to try to resolve this. In cases where it could not be resolved, studies were excluded. Those studies containing exclusively non-RCTs were excluded.

The assessment of RCTs in the included studies had to involve comparison of the protocol against all publications (for outcome reporting bias) or information from trialists (for study publication bias).

### Search strategy

The first author (KD) alone conducted the search. No masking was used during the screening of abstracts. MEDLINE (1950 to 2007), SCOPUS (1960 to 2007) and the Cochrane Methodology Register (1898 to 2007) were searched without language restrictions (final search December 2007 - see Appendix S1 for all search strategies). SCOPUS is a much larger database than EMBASE, it offers more coverage of scientific, technical, medical and social science literature than any other database. Over 90% of

the sources indexed by EMBASE are also indexed by SCOPUS plus many other indexed sources as well.

Additional steps were taken to complement electronic database searches: First, the references given in the empirical evidence section of the HTA report of Song et al [1] were checked for relevance. Second, the lead reviewer of the protocol on the Cochrane library entitled 'Publication bias in clinical trials' [20] (Sally Hopewell) was contacted in November 2007 for references to studies included and excluded in their review. Their search strategy was compared to our own and differences in included studies were discussed between PRW, KD and Sally Hopewell. Finally, the lead or contact authors of all identified studies were asked to identify further studies.

### Quality assessment

To assess the methodological quality of the included studies, we applied the same criteria as a recent Cochrane review [20]. In addition, we examined whether protocols were compared to publications in those studies that purported to investigate outcome reporting bias.

#### 1. Was there an inception cohort?

Yes = a sample of clinical trials registered at onset or on a roster (e.g. approved by an ethics committee) during a specified period of time.

No = anything else

Unclear

#### 2. Was there complete follow up (after data-analysis) of all the trials in the cohort?

Yes  $\geq 90\%$

No  $< 90\%$

Unclear

#### 3. Was publication ascertained through personal contact with the investigators?

Yes = personal contact with investigators, or searching the literature and personal contact with the investigator.

No = searching the literature only

Unclear

#### 4. Were positive and negative findings clearly defined?

Yes = clearly defined

No = not clearly defined

Unclear

#### 5. Were protocols compared to publications?

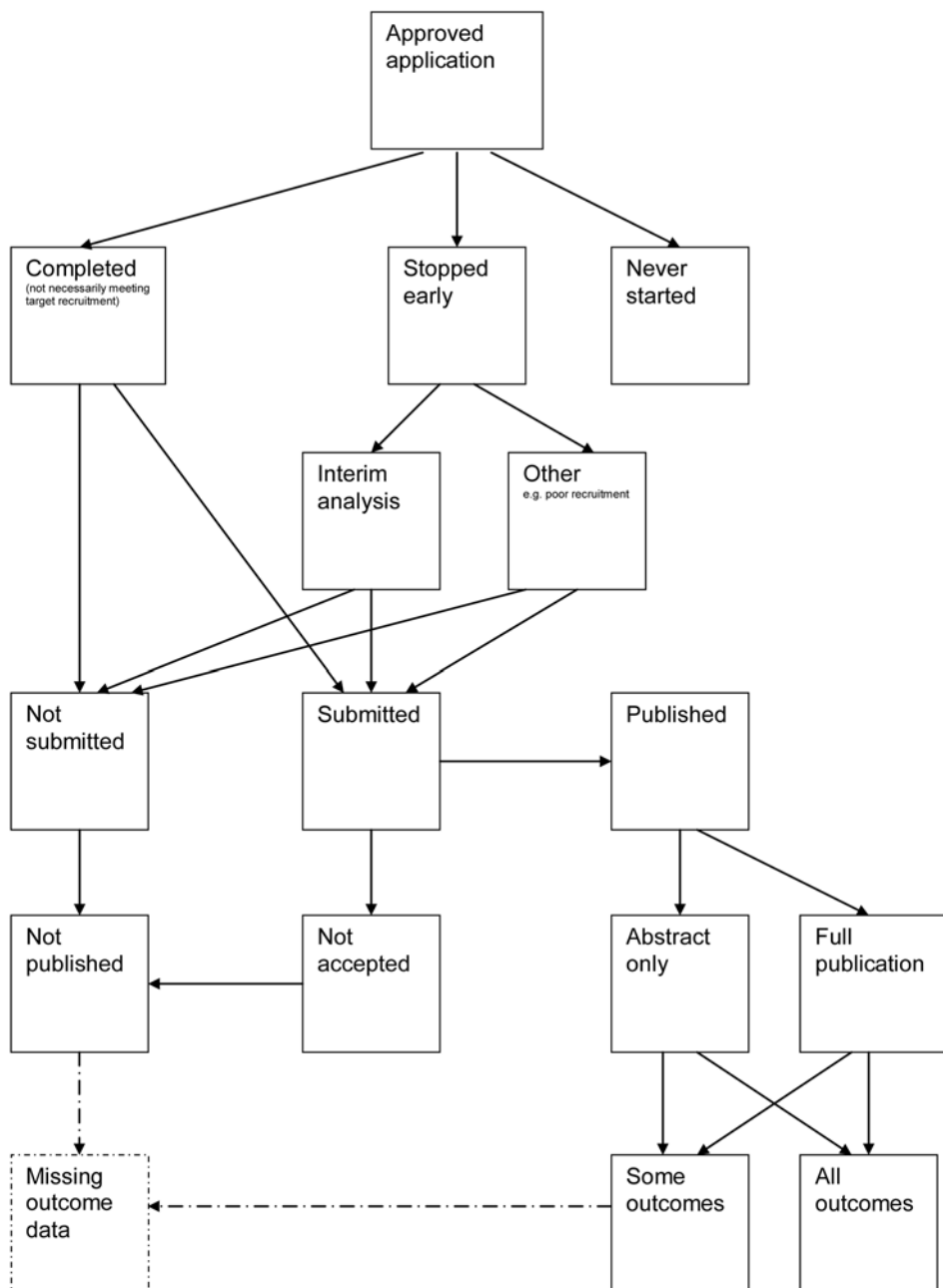
Yes = protocols were compared to publications

No = protocols were not considered in the study

Unclear

### Data extraction

A flow diagram (Figure 1, text S1) to show the status of approved protocols was completed for each empirical study by the first author (KD) using information available in the publication or further publications. Lead or contact authors of the empirical studies were then contacted by email and sent the flow diagram for



**Figure 1. Study flow diagram.**  
doi:10.1371/journal.pone.0003081.g001

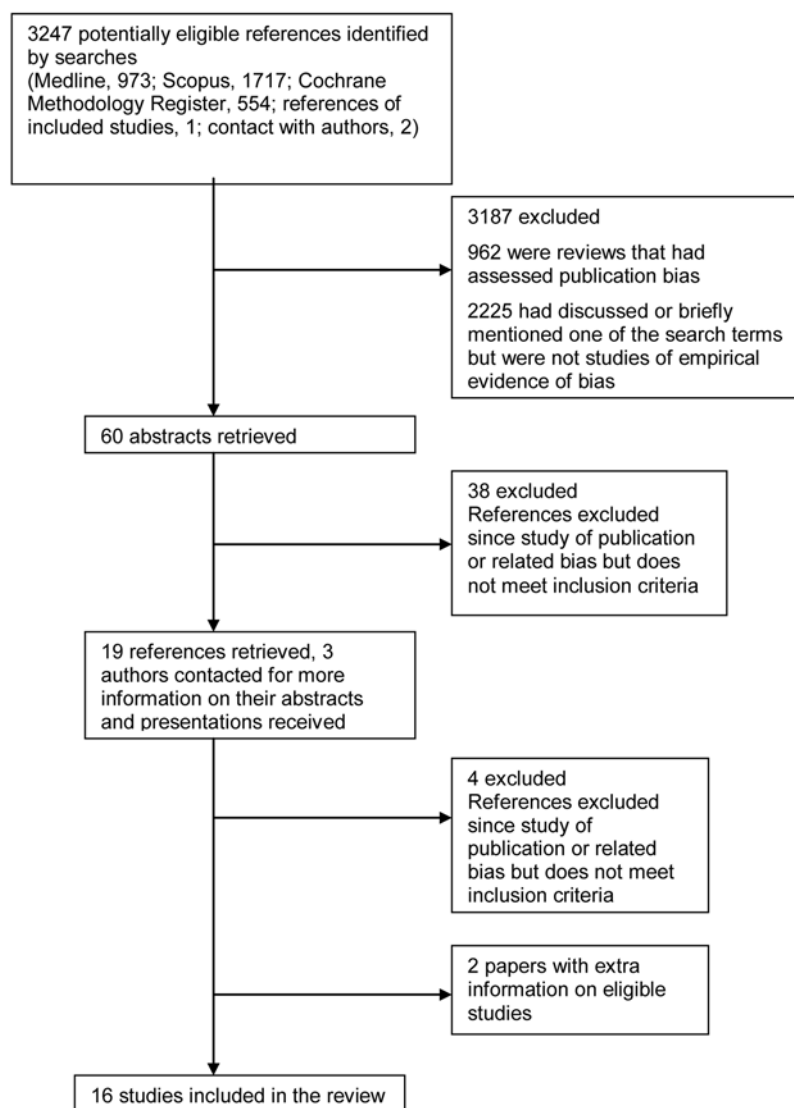
their study to check the extracted data along with requests for further information or clarification of definitions if required. No masking was used and disagreements were resolved through discussion between KD and the lead or contact author of the empirical studies. Where comments from the original author were not available, PRW reviewed the report and discussed queries with KD.

Characteristics of the cohorts were extracted by the first author for each empirical study and issues relating to the methodological quality of the study were noted. We recorded the definitions of 'published' employed in each empirical study. Further, we looked at the way the significance of the results of the studies in each cohort were investigated (i.e. direction of results and whether the

study considered a  $p$ -value  $\leq 0.05$  as definition of significance and where there were no statistical tests whether the results were categorised as negative, positive, important or unimportant). We extracted data on the number of positive and negative trials that were published in each cohort and we extracted all information on the main objectives of each empirical study and separated these according to whether they related to study level or outcome level bias.

### Data analysis

This review provides a descriptive summary of the included empirical studies. We refrained from statistically combining results from the different cohorts due to the differences in their design.



**Figure 2. QUOROM flow diagram.**  
doi:10.1371/journal.pone.0003081.g002

## Results

### Search results

The search of MEDLINE, SCOPUS and the Cochrane Methodology Register led to 973, 1717 and 554 references, respectively. Titles were checked by the first author (KD) and abstracts obtained for 57 potentially relevant studies. Abstracts were assessed for eligibility by the first author; 38 were excluded and full papers were obtained for 16. Only meeting abstracts were available for three studies [17,18,21] and their authors were contacted. Copies of their presentations were received and relevant data extracted.

Four studies were excluded; two were not inception cohorts as they considered completed studies submitted to drug regulatory authorities [22,23], in one study authors were not contacted for information on publication [24] and in another we could not confirm if any of the included studies were RCTs [25]. Fifteen empirical studies were deemed eligible [3–5,13–15,17,18,21,26–29,31,32].

The MEDLINE search identified eight of the included empirical studies [4,5,13–15,26,27,29]. SCOPUS identified eight of the included empirical studies [3–5,13–15,26,29]. The search of the Cochrane Methodology Register identified 15 included empirical studies [3–5,13–15,17,18,21,26–29,31,32]. Seven studies were identified by all three databases [4,5,13,14,15,26,29]. Two studies were identified by two of the three databases [3,27] and six studies were only identified by the Cochrane Methodology Register [17,18,21,28,31,32], three of these studies were abstracts presented at the Cochrane Colloquium.

The HTA report of Song et al [1] led to four potentially eligible empirical studies [3,4,26,27], all of which had been identified previously. References from the included empirical studies led to another paper [33] which gave extra information on the type of publication (full, abstract, none or unknown) for four eligible empirical studies [3,4,26,27]. The reference list provided by Sally Hopewell did not lead to any further studies.

Through contact with the authors, one reference [30] was located and found to be eligible and another [34] was identified

**Table 1.** Study characteristics for inception cohorts.

Study	Objective	Committee approving protocols (country)	Period protocols approved	Date of follow up	Included study designs; Number of studies/Total number of studies (percentage of studies included)	Funding source for all studies	Conclusions
Easterbrook, 1991 [26]	<b>Study publication bias:</b> Evidence of publication bias	Central Oxford research Ethics committee (UK)	1984–1987	1990	Analysed <sup>1</sup> : RCTs 148/285 (52%), observational 86/285 (30%), non-RCT 51/285 (18%)	17% unfunded, 20% NHS or department, 13% Government, 38% Pharmaceutical industry, 12% private/charity	Studies with statistically significant results were more likely to be published, also more likely to lead to a greater number of publications and presentations and to be published in journals with a high citation impact factor.
Dickersin, 1992 [27]	<b>Study publication bias:</b> To investigate factors associated with the publication of research findings, in particular, the association between 'significant' results and publication.	Institutional Review Boards that serve The John Hopkins Health Institutions (USA)	1980	1988	Completed <sup>2</sup> : RCTs 168/514 (33%), observational 273/514 (53%), other experimental 73/514 (14%)	45% NIH, 12% other government, 8% Drug industry, 63% Other, 4% Internal, 18% None.	There is a statistically significant association between significant results and publication.
Dickersin, 1993 [3]	<b>Study publication bias:</b> To investigate the association between trial characteristics, findings and publication.	National Institutes of Health (USA)	1979	1988	RCTs 310/310 (100%)	50% Grant, 30% Contract, 20% Intramural.	Publication bias is a significant problem
Stem, 1997 [4]	<b>Study publication bias and time lag bias:</b> To determine the extent of publication bias and whether publication was delayed for studies with negative results in comparison with those with positive results.	Approved Royal Prince Alfred hospital ethics committee application (Australia)	1979–1988	1992	Total: RCTs 418/748 (56%), observational 165/748 (22%), non trial experiment 165/748 (22%)	117/321 Internal, 206/321 External	Confirms the evidence of publication bias found in other studies. Identifies delay in publication as an additional important factor
Completed questionnaires: RCTs 277/520 (53%), observational 129/520 (25%), non trial experiment 114/520 (22%)							
Cooper, 1997 [32]	<b>Study publication bias:</b> To determine the fate of studies approved by their departmental human subjects review committee	Department of Psychology Human Subjects Committee or Institutional Review Board, Midwestern, research oriented, state university, USA	1986–1988	NI	Analysed <sup>3</sup> : RCTs 167/321 (52%), observational 90/321 (28%), non trial experiment 64/321 (20%)	NI	Significant findings were more likely than non-significant findings to be submitted for meeting presentation or publication.
Wormald, 1997 [21]	<b>Study publication bias:</b> To determine the outcome of all randomised controlled trials processed through the pharmacy of Moorfields eye hospital and to determine whether the publication status of these trials is associated with observed effect of treatment.	Trials processed through the pharmacy of Moorfields Eye Hospital (UK)	1963–1995	1997	RCTs 61/61 (100%)	NI	There was limited evidence of publication bias

Table 1. cont.

Study	Objective	Committee approving protocols (country)	Period protocols approved up	Date of follow up	Included study designs; Number of studies/Total number of studies (percentage of studies included)	Funding source for all studies	Conclusions
Ioannidis, 1998 [5]	<b>Study publication bias and time lag bias:</b> To evaluate whether the time to completion and time to publication of randomized phase 2 and phase 3 trials are affected by the statistical significance of results.	Efficacy clinical trials conducted by AIDS Clinical Trials Group and Terry Bein Community Programs for Clinical Research on AIDS (USA)	1986–1996	1996	RCTs 109/109 (100%)	Data managed by: 10% Pharmaceutical industry, 90% Other federally sponsored.	There is a time lag in the publication of negative findings that occurs mostly after the completion of the trial follow up.
Pich, 2003 [28]	<b>Publication rate:</b> To assess the outcome of protocols submitted to the HCEC.	Hospital Clinic Ethics Committee (Spain)	1997	2001	RCTs 158/158 (100%)	89% Pharmaceutical industry, 11% Other.	Only 64% of trials that started were finally implemented and finished in accordance with the original protocol. Only 31% of closed clinical trials were published or in-press in peer reviewed journals.
Cronin, 2004 [31]	<b>Study publication bias:</b> Assess the degree to which research project findings were published and explore factors that influenced publication	R&D projects funded by the NHS and commissioned by the North Thames Regional Office (UK)	1993–1998	1995–1998	NI	100% government	Funders should consider the significant number of studies that did not result in publication and the higher rate of publication in peer reviewed journals from some programs
Decullier, 2005 [29]	<b>Study publication bias and time lag bias:</b> To describe the fate of approved protocols and assess publication bias at a national level.	French Research Ethics Committees (France)	1994	2000–2002	Total: RCTs 345/649 (53%), descriptive/observational 91/649 (14%), non-randomised 213/649 (33%)  Completed: RCTs 269/501 (54%), descriptive/observational 66/501 (13%), non-randomised 166/501 (33%)	8% No funding, 73% Private funding, 13% Public, 6% Mixed.	Too many studies are not completed and too many are not published.
Decullier, 2006 [30]	<b>Study publication bias:</b> To investigate the fate of protocols submitted for funding, whether they were funded or not.	Greater Lyon regional scientific committee (France)	1997	2003	RCTs 20/142 (14%), experimental 15/142 (10%), descriptive 45/142 (32%), analytical 27/142 (19%), not clinical 28/142 (20%), not available 7/142 (5%)	38% committee funded	Some protocols submitted for funding were initiated and completed without any funding declared. To our understanding this means that not all protocols submitted really needed funding and also that health care facilities are unaware that they implicitly financially support and pay for biomedical research.
Hahn, 2002 [13]	<b>Outcome reporting bias:</b> To examine the extent of within-study selective reporting in clinical research	Local Research Ethics Committee (UK)	1994	1999	Of 15 published: RCTs 2/15 (13%), non RCT 2 (13%), uncontrolled trial 2 (13%), case control 1 (7%), survey 2 (13%), cohort and case control 1 (7%), method evaluation study 5 (34%)	Not recorded	Within-study selective reporting may be examined qualitatively by comparing the study report with the protocol. The results suggest that it might well be substantial; the bias could only be broadly identified as protocols were not sufficiently precise.

Table 1. cont.

Study	Objective	Committee approving protocols (country)	Period protocols approved	Date of follow up	Included study designs; Number of studies/Total number of studies (percentage of studies included)	Funding source for all studies	Conclusions
Chan, 2004a [14]	<b>Outcome reporting bias:</b> To determine whether outcome reporting bias would be present in a cohort of government funded trials subjected to rigorous peer review.	Canadian Institutes of Health Research (Canada)	1990–1998	2002/2003	RCTs 108/108 (100%)	42% jointly funded by industry and CIHR/MRC, 58% no industry funding.	Selective reporting of outcomes frequently occurs in publications of high-quality government-funded trials.
Chan, 2004b [15]	<b>Outcome reporting bias:</b> To study empirically the extent and nature of outcome reporting bias in a cohort of RCTs	Scientific-Ethical Committees for Copenhagen and Frederiksberg, Denmark	1994–1995	2003	RCTs 304/304 (100%)	55% Full industry, 17% Partial industry, 22% Non-industry, 7% non declared.	Reporting of trials is frequently incomplete, biased and inconsistent with protocols.
Gherzi, 2006 [17]	<b>Outcome reporting bias:</b> To identify discrepancies in the identity and definition of the primary outcome and to investigate factors associated with the completeness of reporting of the primary outcome.	CSAHS Ethics review committee (Australia)	1992–1996	NI	RCTs 318/318 (100%)	37% commercial funding, 63% no commercial funding.	NI
Von Elm, 2008 [18]	<b>Outcome reporting bias:</b> To study trial outcomes specified in protocols of RCTs and reported in subsequent full publications and to estimate publication rate. Investigate whether outcomes are discrepant and to investigate factors that are associated with complete reporting (e.g. statistical significance, funding)	University of Berne/CH ethics committee (Switzerland)	1988–1998	2006	Total: RCTs 451/1698 (27%) In depth analyses: 451/451 (100%)	81% industry, 10% other <sup>4</sup> ,	About half of drug trials are not published. A high prevalence of pre-specified outcomes are not reported and discrepancies includes primary outcomes. Completeness of reporting of an outcome is associated with statistical significance.

<sup>1</sup>Easterbrook et al assumed that only studies that had been analysed had the potential for being written up and published, so tests for study publication bias were restricted to these.

<sup>2</sup>Studies for which there was a full interview by the researchers of the cohort study and for which information on the nature of results and publication was provided.

<sup>3</sup>Of the 520 studies with completed questionnaires, 321 had analysis undertaken with results available and were included in further analysis of the association between study outcome and time to publication.

<sup>4</sup>Both groups are not mutually exclusive. 4% had a statement of both sources of funding. In the remainder, the protocols did not include information on how study was funded.

NI No information available.

doi:10.1371/journal.pone.0003081.t001

**Table 2.** Methodological Quality Assessment.

Quality criteria	Inception cohort	Complete follow up of all trials	Publication ascertained through personal contact with investigators	Definition of positive and negative findings clearly defined	Comparison of protocol to publication
Easterbrook, 1991 [26]	Y	N (25% lost to follow up)	Y	Y (positive: $p < 0.05$ /striking, negative: $p \geq 0.05$ /definite but not striking, null: no difference observed between the groups/null findings.)	NA
Dickersin, 1992 [27]	Y	Y	Y	Y (positive: $p < 0.05$ /statistically significant, negative: suggestive trend but not statistically significant, null: no trend or difference. In terms of importance when statistical tests were not performed: great, moderate or little.)	NA
Dickersin, 1993 [3]	Y	N (14% refused to participate)	Y	Y (positive: $p < 0.05$ significant/of great importance, negative: showing a trend in either direction but not statistically significant/moderate importance/no difference/little importance.)	NA
Stern, 1997 [4]	Y	N (only 70% of questionnaires were completed)	Y	Y (positive: $p < 0.05$ significant/striking/important/definite, negative: non-significant trend $0.05 \leq p < 0.10$ or non-significant or null $p \geq 0.10$ /unimportant and negative)	NA
Cooper, 1997 [32]	Y	Y	Y	N (significant and non-significant)	NA
Wormald, 1997 [21]	Y	Y	Y	Y (positive: $p < 0.05$ , negative: $p \geq 0.05$ )	NA
Ioannidis, 1998 [5]	Y	Y	Y	Y (positive: $p < 0.05$ significant and in favour of experimental therapy arm or any arm when there is no distinct control, negative: nonstatistically significant findings or favouring the control arm )	NA
Pich, 2003 [28]	Y	Y	Y	NA	NA
Cronin, 2004 [31]	Y	Y	Y	U (study showed effect)	NA
Decullier, 2005 [29]	Y	N (only 69% of questionnaires were completed)	Y	Y (confirmatory/inconclusive/invalidating)	NA
Decullier, 2006 [30]	Y	N (only 80% of questionnaires were completed)	Y	Y (scale from 1 to 10 for not important to very important)	NA
Hahn, 2002 [13]	Y	Y	Y	NA	Y
Chan, 2004a [14]	Y	Y	Y	Y (positive: $p < 0.05$ , negative: $p \geq 0.05$ )	Y
Chan, 2004b [15]	Y	Y	Y	Y (positive: $p < 0.05$ , negative: $p \geq 0.05$ )	Y
Gherzi, 2006 [17]	Y	Y	Y	Y (positive: $p \leq 0.05$ , negative: $p > 0.05$ )	Y
Von Elm, 2008 [18]	Y	Y (for drug trials)	Y	Y (positive: $p < 0.05$ )	Y

Y yes.

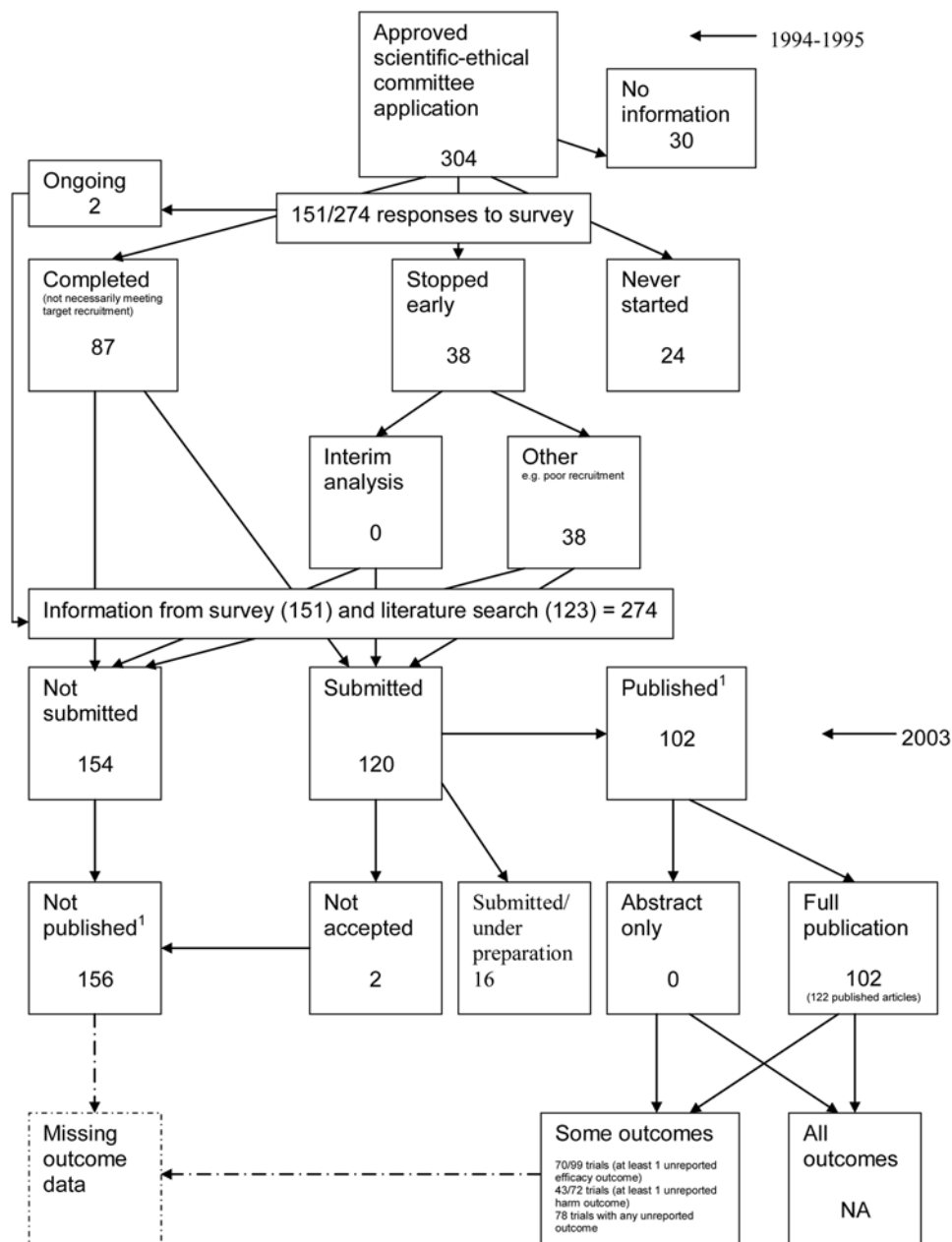
N no.

U unclear.

NA Not applicable.

doi:10.1371/journal.pone.0003081.t002





1. Information from survey and literature search

**Figure 3. Status of approved protocols for Chan 2004b study [15].**  
doi:10.1371/journal.pone.0003081.g003

that gave more information on one of the eligible studies [5]. Thus in total, the search strategy identified 16 eligible empirical studies (Figure 2). We are aware of three further empirical studies currently underway in Italy (D'Amico, personal communication), Germany (Von Elm, personal communication) and the USA (Djulgovic, personal communication), but no further information is available at this stage.

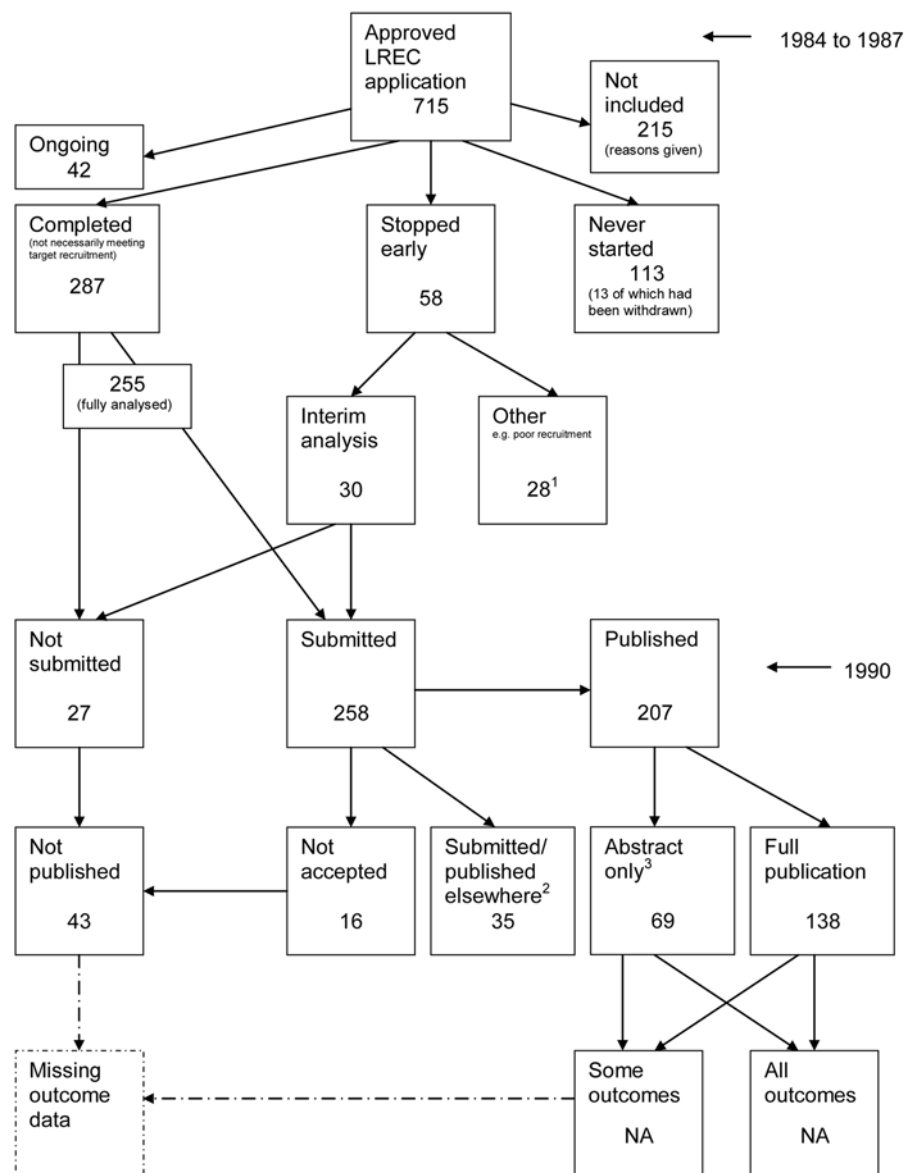
#### Included studies

**Study publication bias.** Eleven empirical studies considered the process up to the point of publication [3–5,21,26–32].

However, two of these empirical studies [28,31] did not consider whether a study was submitted for publication.

Four cohorts included only RCTs [3,5,21,28]; in the remaining seven cohorts [4,26,27,29–32] the proportion of included RCTs ranged from 14% to 56%. The results presented in the flow diagrams relate to all studies within each cohort because it was not possible to separate information for different types of studies (RCTs versus other).

**Outcome reporting bias.** Five empirical studies covered the entire process from the study protocol to the publication of study outcomes [13–15,17,18]. However, three of these empirical studies



1. Information on this group in terms of number submitted/not submitted is not available.
2. 35 submitted or published elsewhere; drafted or under editorial review (book chapters, conference abstracts or proceedings)
3. Presented only

NA = Not Available

**Figure 4. Status of approved protocols for Easterbrook 1991 study [26].**

doi:10.1371/journal.pone.0003081.g004

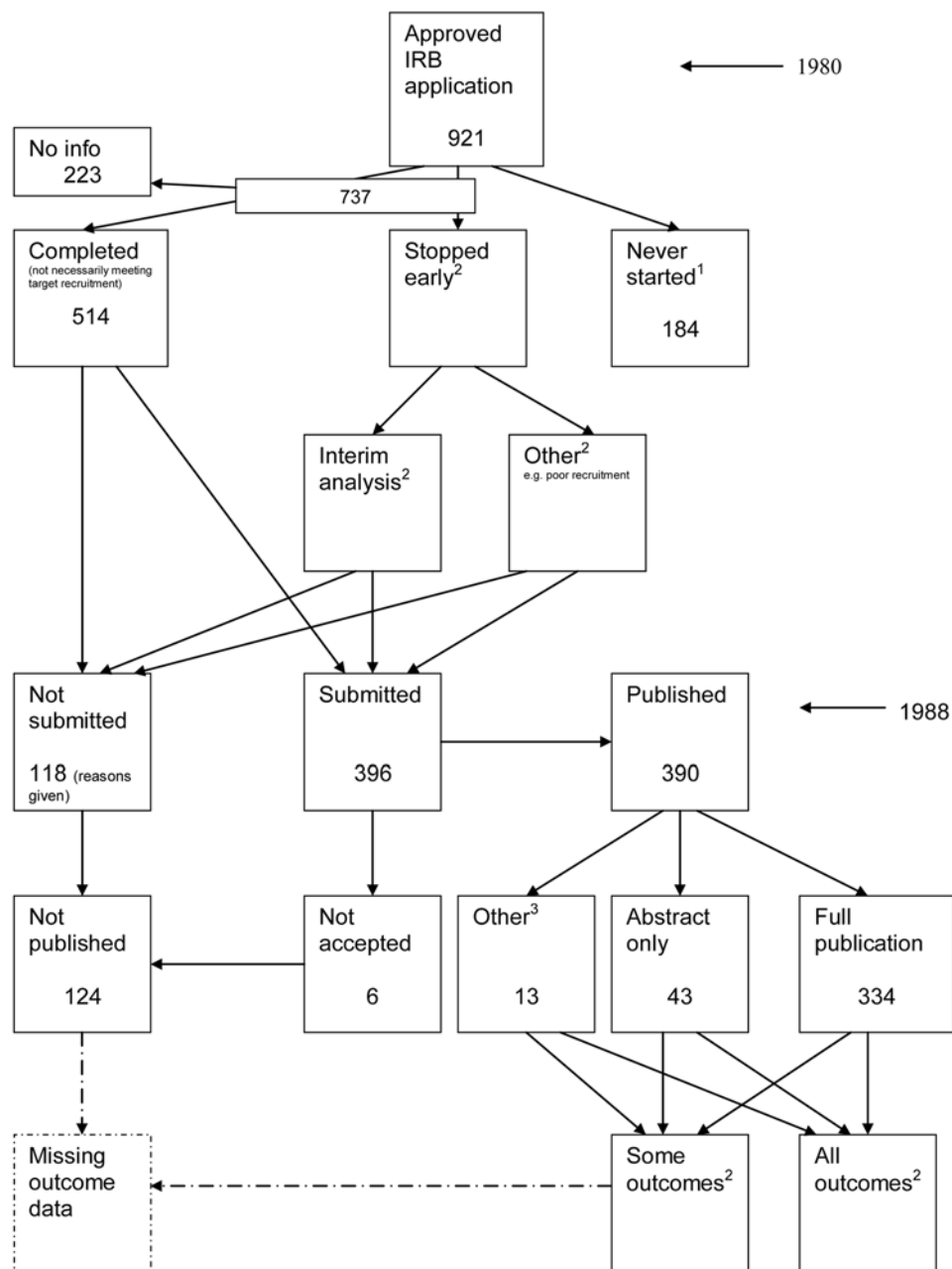
[13,17,18] did not consider whether a study was submitted for publication. Four cohorts included only RCTs [14,15,17,18]; in the remaining cohort [13] the proportion of included RCTs was 13%.

### Study Characteristics

Table 1 contains information on empirical study characteristics. The majority of the empirical study objectives related to study publication bias or outcome reporting bias.

**Study publication bias.** Three of the empirical studies investigating study publication bias also assessed time lag bias [4,5,29], one [28] assessed the outcome of protocols submitted to a

research ethics committee (for example whether trials were started and if they were published) and another considered whether absence of acknowledged funding hampered implementation or publication [30]. Seven of the empirical studies [4,26–30,32] assessed protocols approved by ethics committees, one [3] assessed those approved by health institutes, one assessed trials processed through a hospital pharmacy [21], one assessed studies funded by the NHS and commissioned by the North Thames Regional Office [31] and one empirical study [5] assessed trials conducted by NIH-funded clinical trials groups. The time period between protocol approval and assessment of publication status varied widely (less than one year to 34 years).

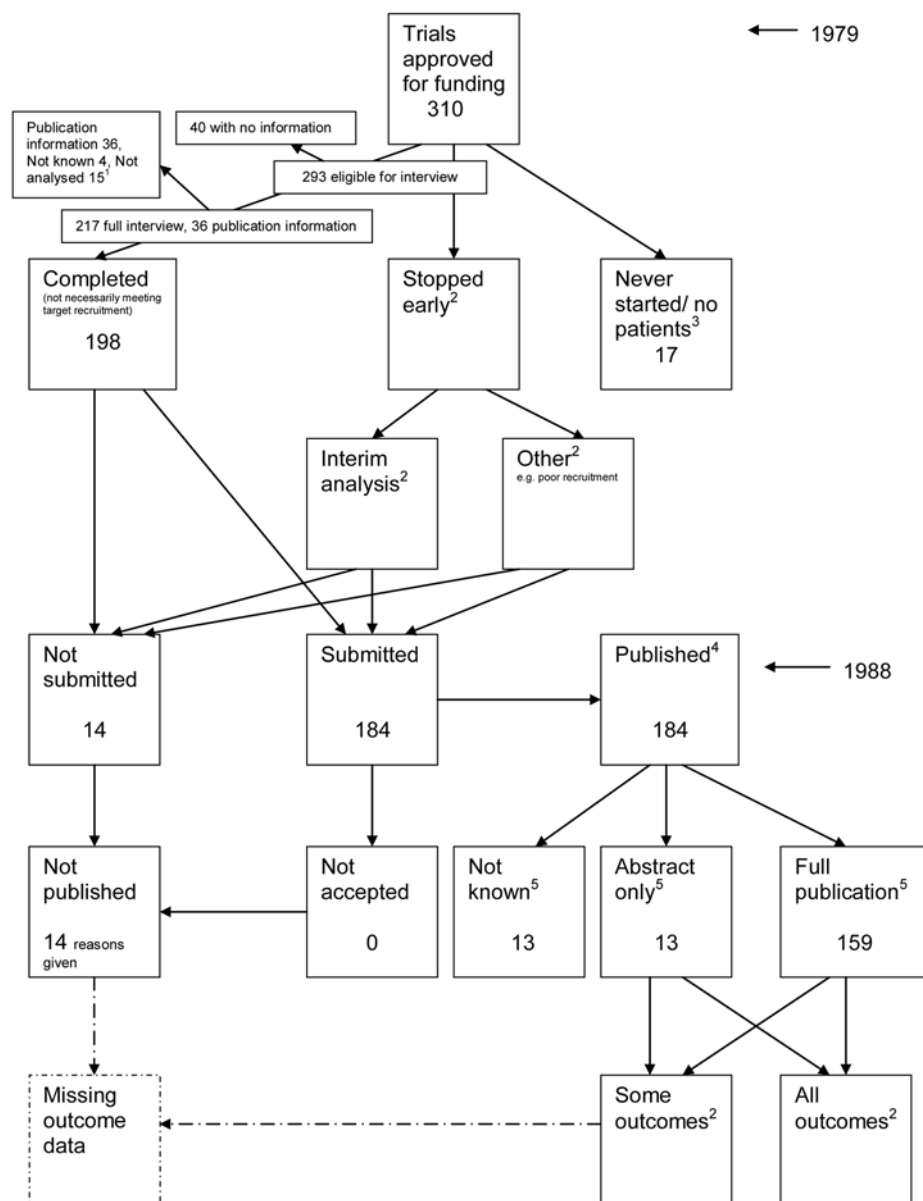


**Figure 5. Status of approved protocols for Dickersin 1992 study [27].**  
doi:10.1371/journal.pone.0003081.g005

**Outcome reporting bias.** Four of the empirical studies [13,15,17,18] assessed protocols approved by ethics committees and one empirical study [14] assessed those approved by a health institute. The time period between protocol approval and assessment of publication status varied from four to eight years.

### Quality Assessment

Details of the methodological quality are presented in Table 2. The overall methodological quality of included empirical studies was good, with more than half of studies meeting all criteria.



1. Unclear as to if these studies were completed but not analysed.
2. No information was available in the paper to fill in these boxes.
3. Unclear as to whether this number is correct.
4. It is unclear what was published from the 9 studies that were not analysed but published.
5. It is not clear why these numbers do not add up;  $159+13+13 \neq 184$ .

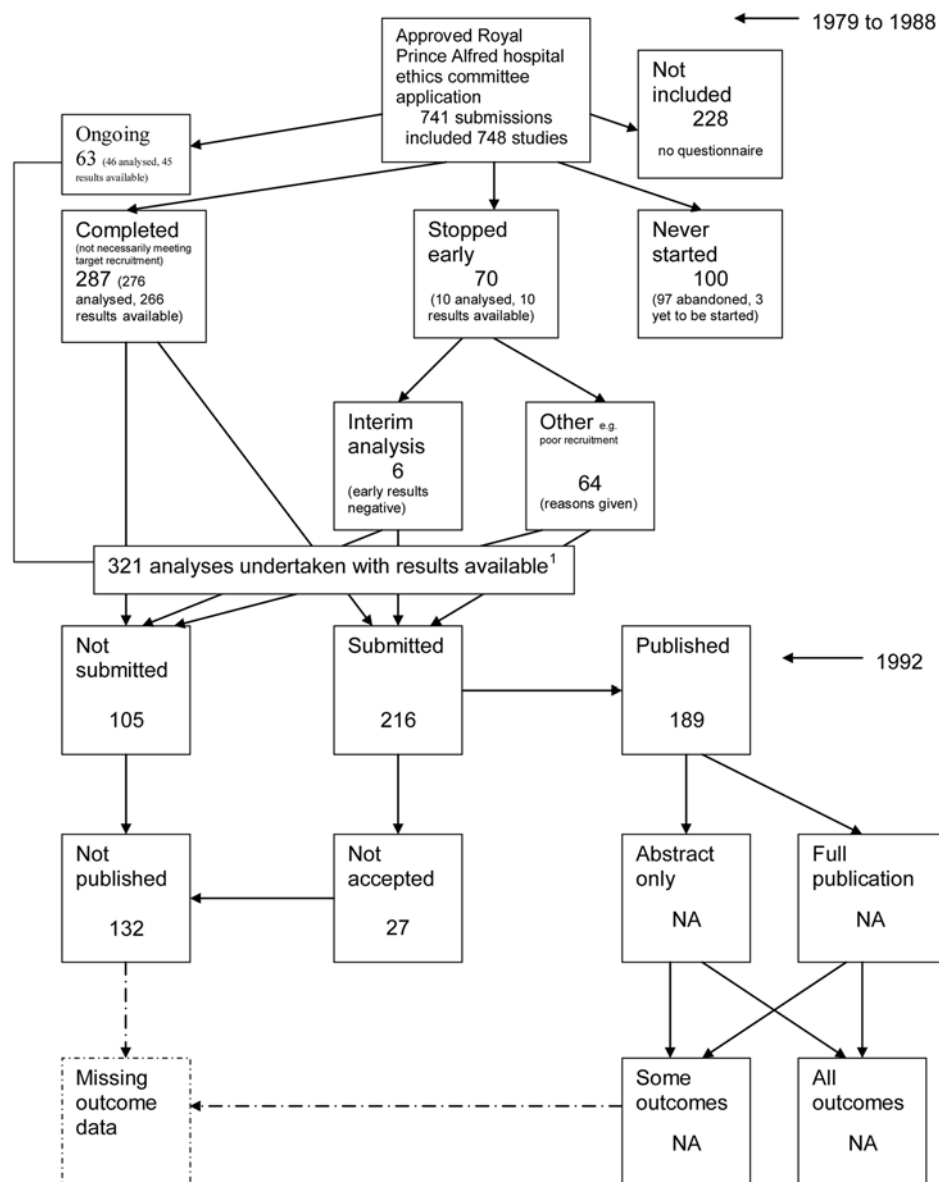
**Figure 6. Status of approved protocols for Dickersin 1993 study [3].**  
doi:10.1371/journal.pone.0003081.g006

**Study publication bias.** Four of the eleven empirical studies [5,21,27,28] met all four of the criteria for studies investigating study publication bias (inception cohort, complete follow up of all trials, publication ascertained through personal contact with the investigator and definition of positive and negative findings clearly defined). In five empirical studies [3,4,26,29,30] there was less than 90% follow up of trials and in 2 empirical studies [31,32] the definition of positive and negative findings was unclear.

**Outcome reporting bias.** All five empirical studies [13–15,17,18] met all five criteria for studies investigating ORB (inception cohort, complete follow up of all trials, publication

ascertained through personal contact with the investigator, definition of positive and negative findings clearly defined and comparison of protocol to publication).

As some studies may have several specified primary outcomes and others none, we looked at how each of the empirical studies dealt with this: Hahn et al [13] looked at the consistency between protocols and published reports in regard to the primary outcome and it was only stated that there were 2 primary outcomes in one study. In both of their empirical studies Chan et al [14,15] distinguished harm and efficacy outcomes but did consider the consistency of primary outcomes between protocols and publica-



1. 266 completed, 45 ongoing and 10 stopped early.

**Figure 7. Status of approved protocols for Stern 1997 study [4].**  
doi:10.1371/journal.pone.0003081.g007

tions and stated how many had more than one primary outcome. Ghera et al [17] included studies with more than one primary outcome and included all primary outcomes in the analysis but excluded studies with primary outcomes that were non identifiable or included more than 2 time points. This is due to complex outcomes being more prone to selective reporting. von Elm et al [18] considered harm and efficacy outcomes and primary outcomes.

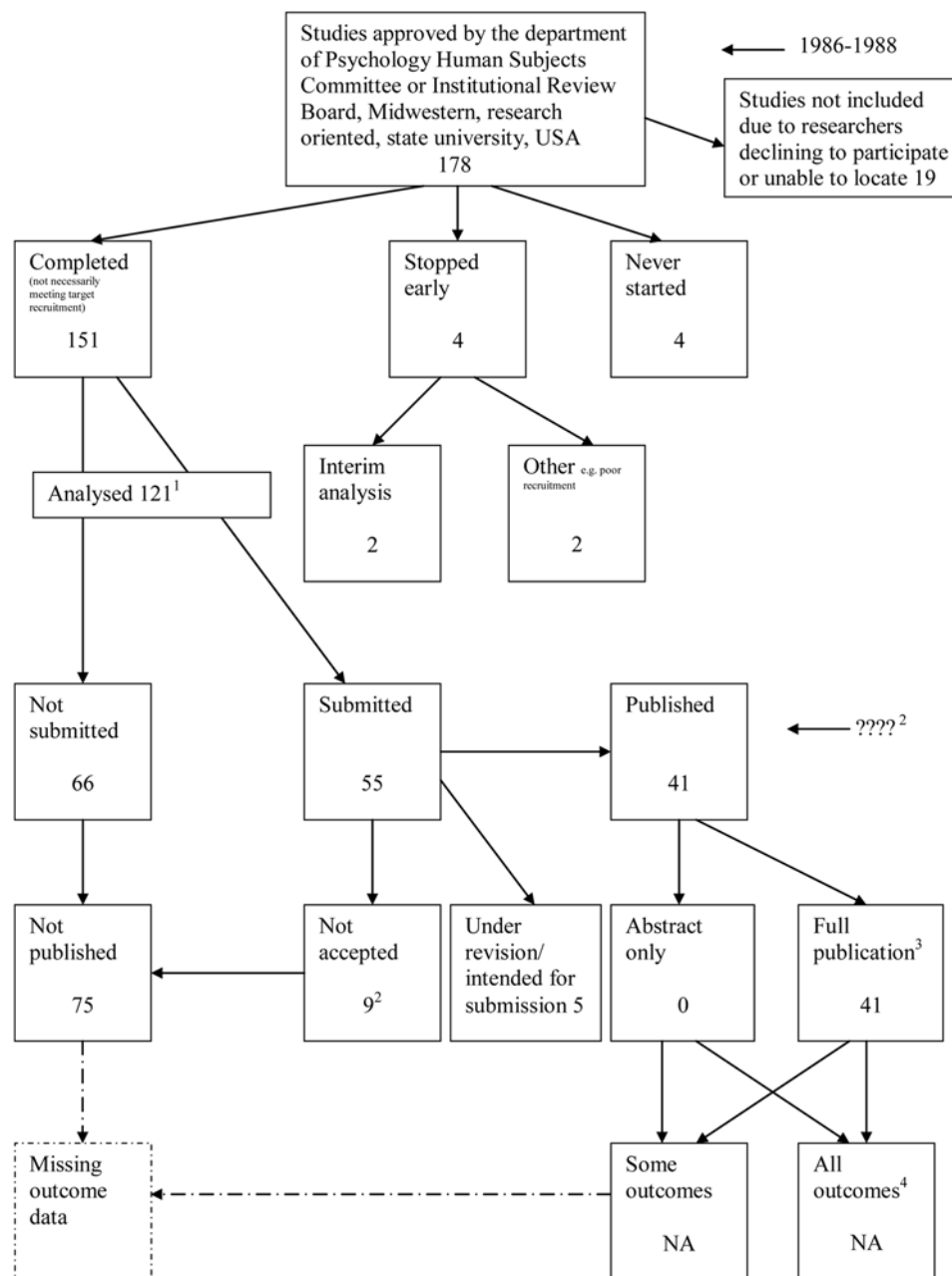
### Flow diagrams

The flow diagrams (Figures 3 to 18) show the status of approved protocols in included empirical studies based on available

publications and additional information obtained such as number of studies stopped early or never started.

**Study publication bias.** No information other than the study report was available for one empirical study [26] due to its age. Information could not be located for three empirical studies [3,27,32]. A conference abstract and poster was only available for one empirical study presented over 10 years ago [21]. Extra information from lead or contact authors was available for six empirical studies [4,5,28–31], including data to complete flow diagrams, information on definitions and clarifications.

**Outcome reporting bias.** A conference presentation only was available for one empirical study which is still to be published



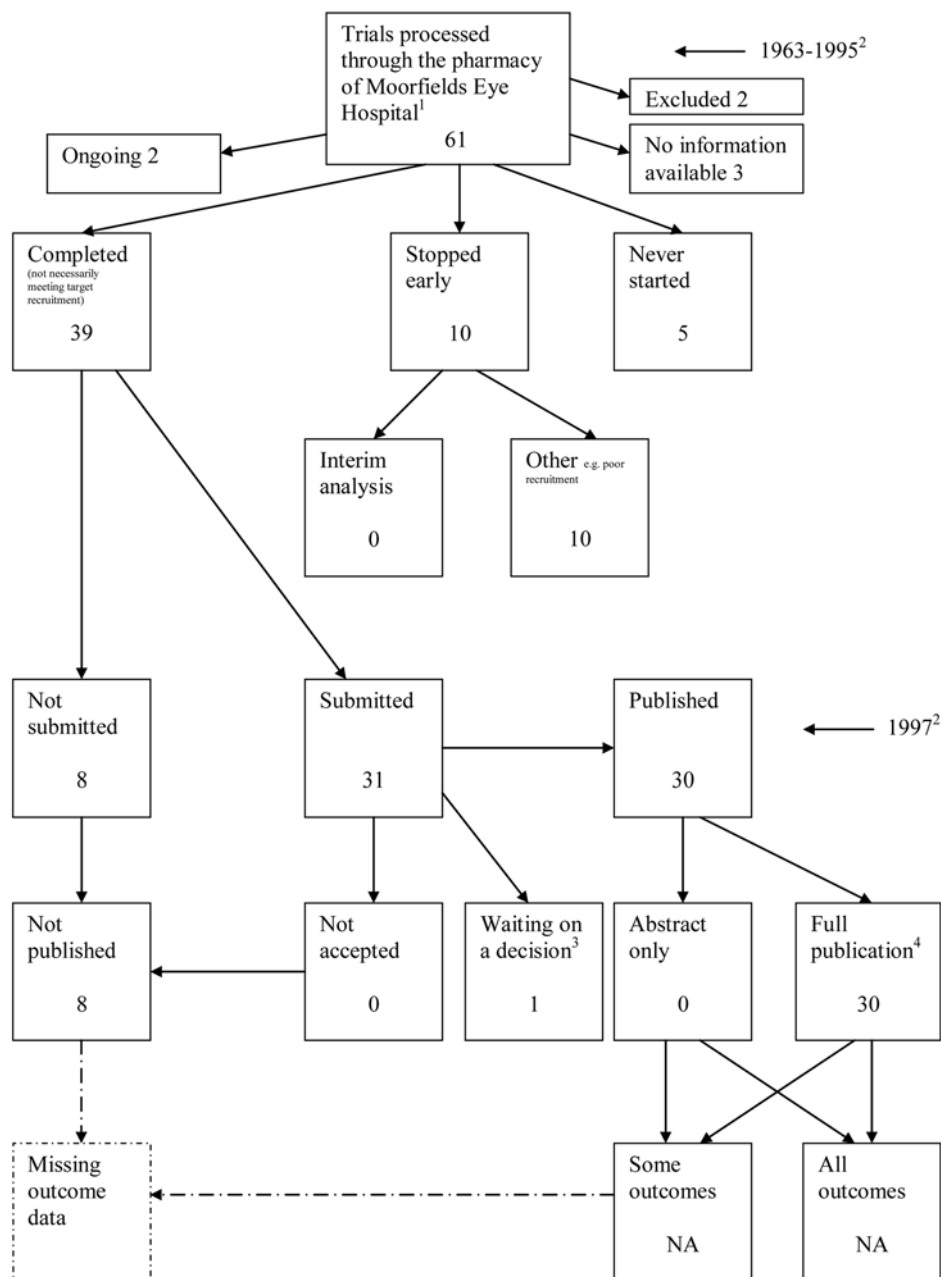
1. Only 121/151 analysed, 27/30 unanalysed were undergraduate class projects.
2. The year the study was conducted is unclear.
3. Of these, 3 are books.
4. Not stated if outcomes were looked at.
5. Not stated how many of the analysed studies or completed studies were RCTs.

**Figure 8. Status of approved protocols for Cooper 1997 study [32].**  
doi:10.1371/journal.pone.0003081.g008

in full [17]. Extra information from lead or contact authors was available for four empirical studies [13–15,18], including data to complete flow diagrams, information on definitions, clarifications and extra information on outcomes. Original flow diagrams and questions asked are available on request.

Figure 3 shows for illustrative purposes the completed flow diagram for the empirical study conducted by Chan et al [15] on the status of 304 protocols approved by the Scientific-Ethical

Committees for Copenhagen and Frederiksberg in 1994–1995. The empirical study was conducted in 2003, which allowed sufficient time for trial completion and publication. Thirty studies were excluded as the files were not found. Surveys were sent to trial investigators with a response rate of 151 out of 274 (55%); of these two were ongoing, 38 had stopped early, 24 studies had never started and 87 studies were completed. Information from the survey responses (151) and the literature search alone (123)



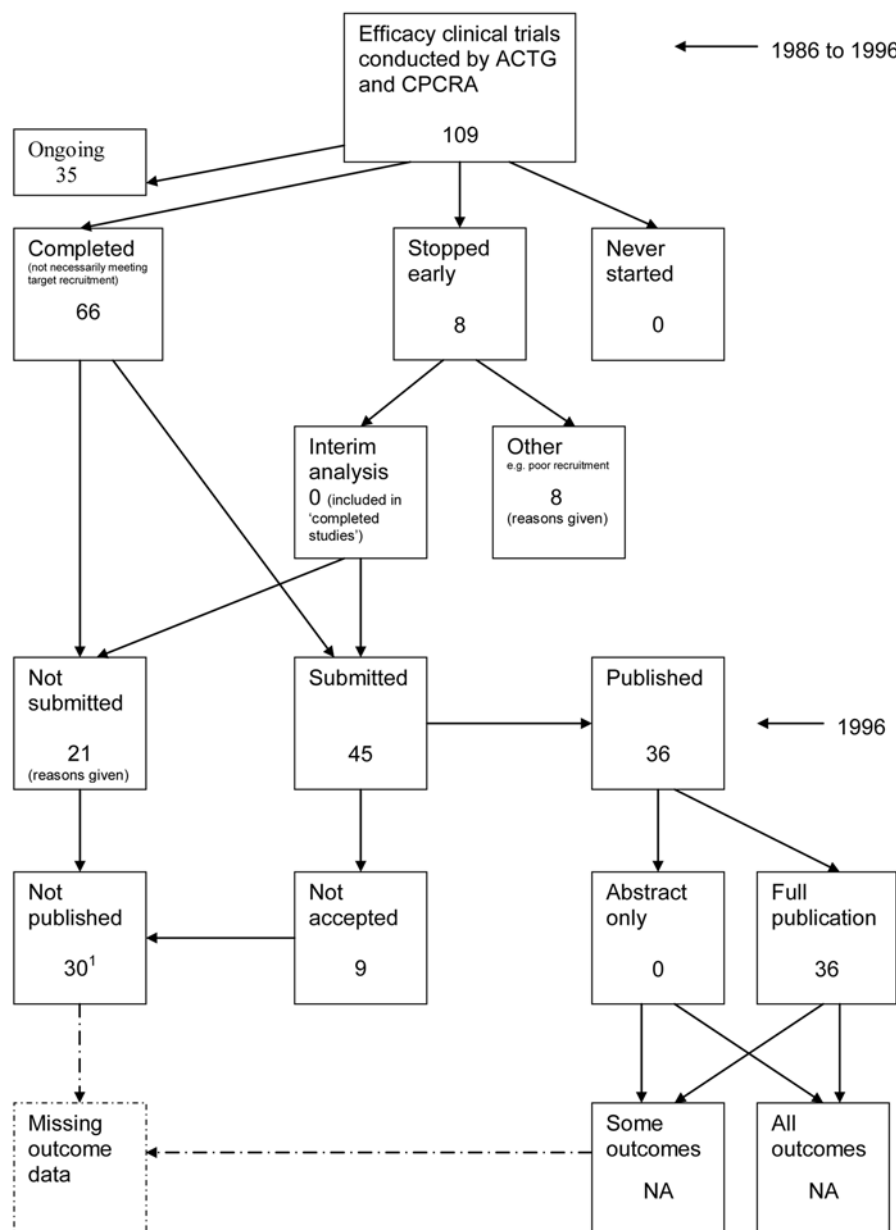
1. In the poster this number is 61, in the abstract it is 68. It is unclear why there is a difference.
2. Unclear if the dates are correct.
3. The 1 study that was submitted – unclear if this waiting for a decision on acceptance or had it been rejected for publication.
4. Unclear if all 30 published studies were full publications. Unclear on definition of 'published'.
5. Unclear if there was complete follow up of all trials.

**Figure 9. Status of trials for Wormald 1997 study [21].**  
doi:10.1371/journal.pone.0003081.g009

indicated that 120 studies had been submitted for publication and 154 studies had not been submitted for publication. Of the 120 submitted studies; 102 had been fully published, 16 had been submitted or were under preparation and two had not been accepted for publication. This resulted in 156 studies not being published.

### Publication and trial findings

**Study publication bias.** Table 3 shows the total number of studies published in each cohort which varies widely from 21% to 93%. Nine of the cohorts [3–5,21,26,27,29,30,32] consider what proportion of trials with positive and negative results are published, ranging from 60% to 98% and from 19% to 85%,



1. Nine studies had been submitted at least once and 8 had been completed less than one year ago. The other 13 had been completed over a year ago and they were still not submitted for full publication at the time data were censored for analysis. Many of the trials that were categorized as: "unpublished" may have been published in abstract form (the same applies also to those that were published), but it is not possible to trace this back at this point.

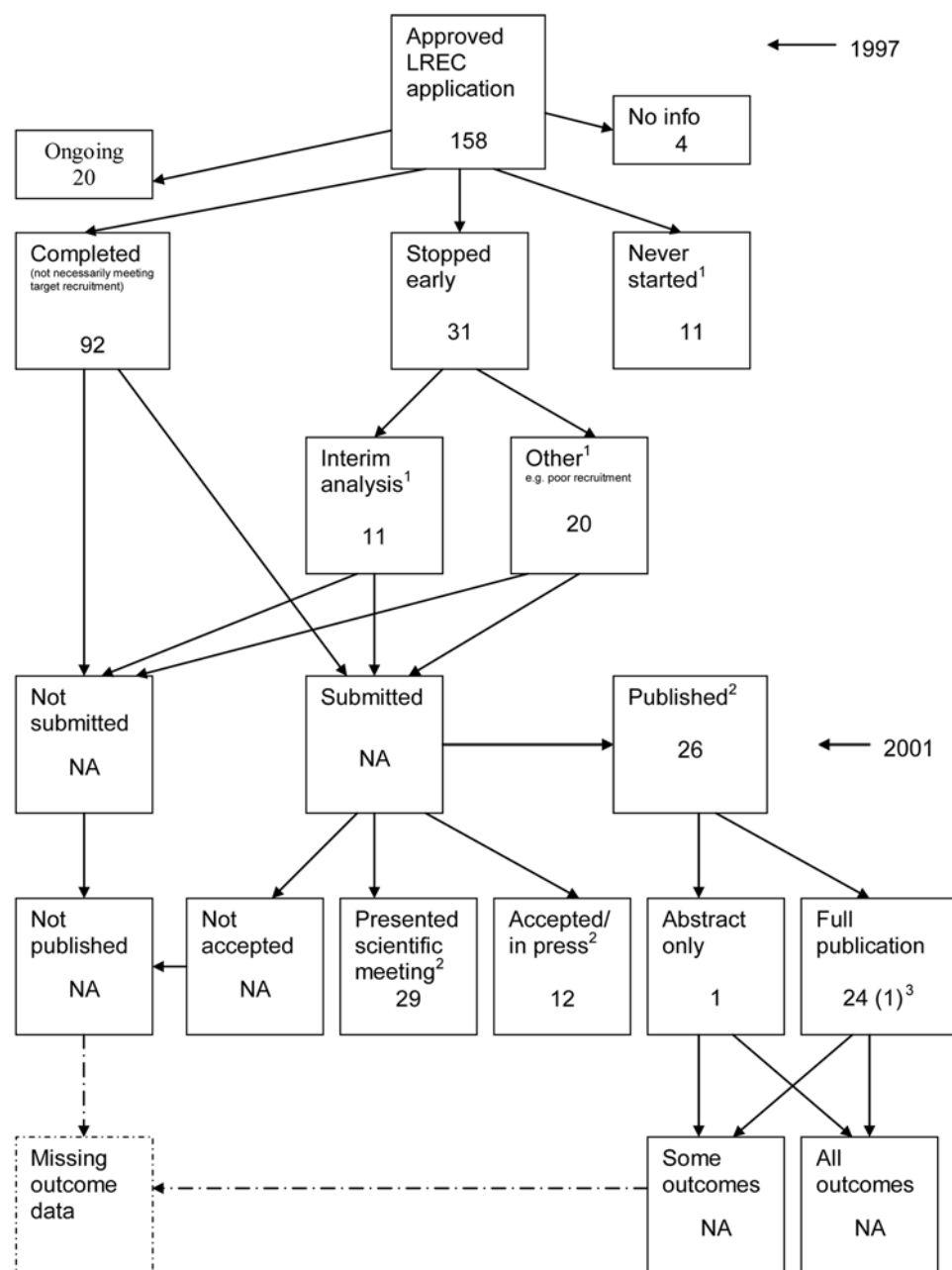
**Figure 10. Status of approved protocols for Ioannidis 1998 study [5].**  
doi:10.1371/journal.pone.0003081.g010

respectively. Only four cohorts [4,26,29,32] consider what percentage of studies with null results (no difference observed between the two study groups,  $p > 0.10$ , inconclusive) are published (32% to 44%). The results consistently show that positive studies are more likely to be published compared to negative studies.

Table 4 shows general consistency in the definition of 'published.' However, two empirical studies [3,27] considered grey literature in their definition of 'published' although information on full publications and grey literature publications are separated (Figures 5, 6). Although not considered in the

definition of 'published', four empirical studies [26,28–30] gave information on the grey literature or reports in preparation. Three empirical studies gave no information on their definition of 'published' [21,31,32]. In addition, results are presented for the percentage of studies not submitted for journal publication (7% to 58%), of studies submitted but not accepted for publication (0 to 20%) by the time of analysis of the cohort and the percentage of studies not published that were not submitted (63% to 100%). This implies that studies remain unpublished due largely to failure to submit rather than rejection by journals.





1. Reasons given in the paper.

2. From 143 clinical trials that started. The 26 published trials include 22 that were completed and 4 that were stopped early.

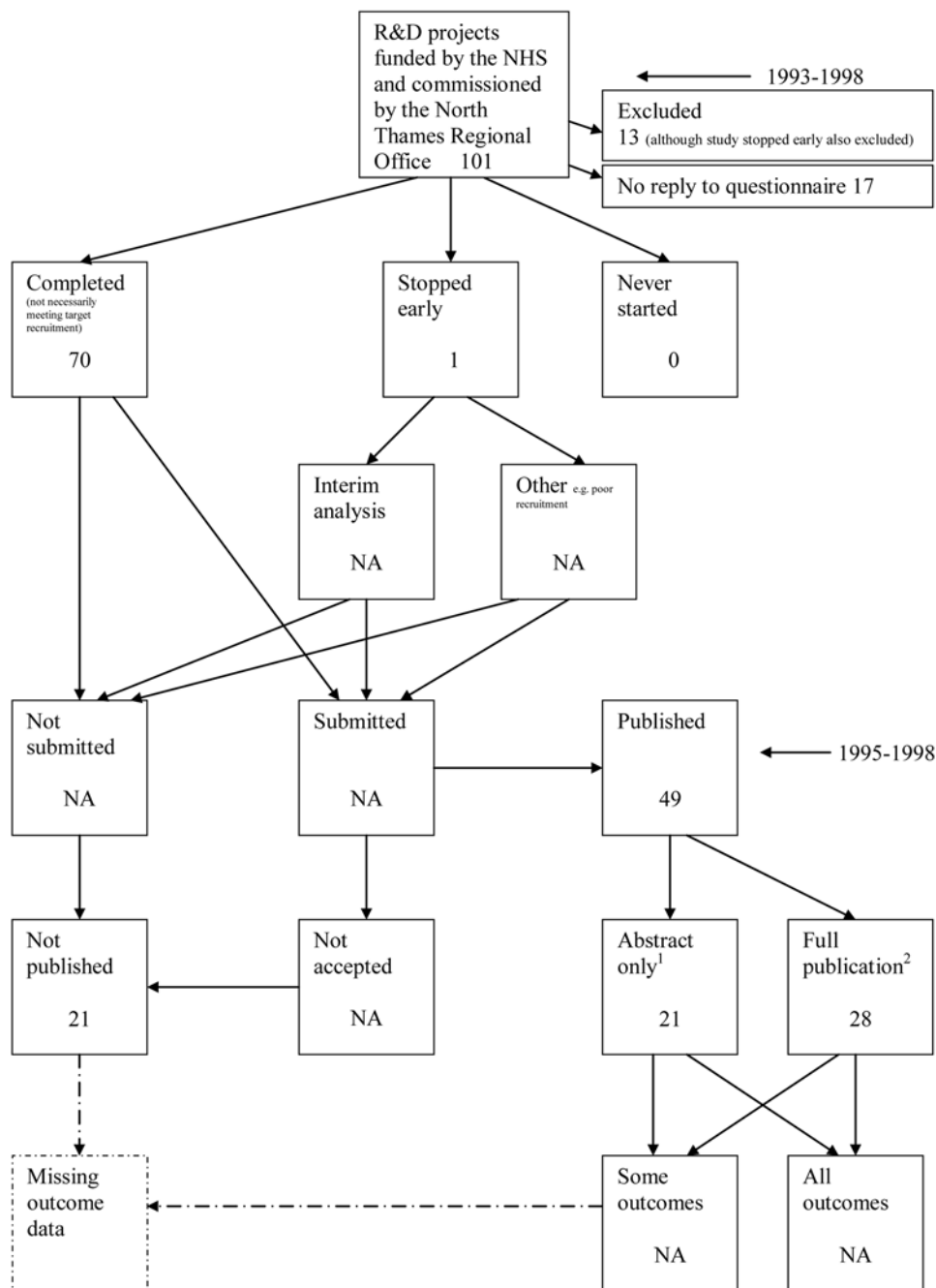
3. Published in a supplement.

**Figure 11. Status of approved protocols for Pich 2003 study [28].**

doi:10.1371/journal.pone.0003081.g011

The main findings of the empirical studies are shown in Table 5 and they are separated into study level and outcome level results. Eight of the included cohort studies [3,4,21,26,27,29,31,32] investigated results in relation to their statistical significance. One empirical study considered the importance of the results as rated by the investigator [30] and another empirical study considered confirmatory versus inconclusive results [29]. Five of the empirical studies [3,4,26,27,29] that examined the association

between publication and statistical significance found that studies with statistically significant results were more likely to be published than those with non-significant results. Stern et al [4] reported that this finding was even stronger for their subgroup of clinical trials (Hazard Ratio (HR) 3.13 (95% confidence interval (CI) 1.76, 5.58),  $p=0.0001$ ) compared to all quantitative studies (HR 2.32 (95% CI 1.47, 3.66),  $p=0.0003$ ). One empirical study [32] found that studies with statistically significant results were more likely to



1. Book chapters, reports, other publications.
2. In peer reviewed journal.

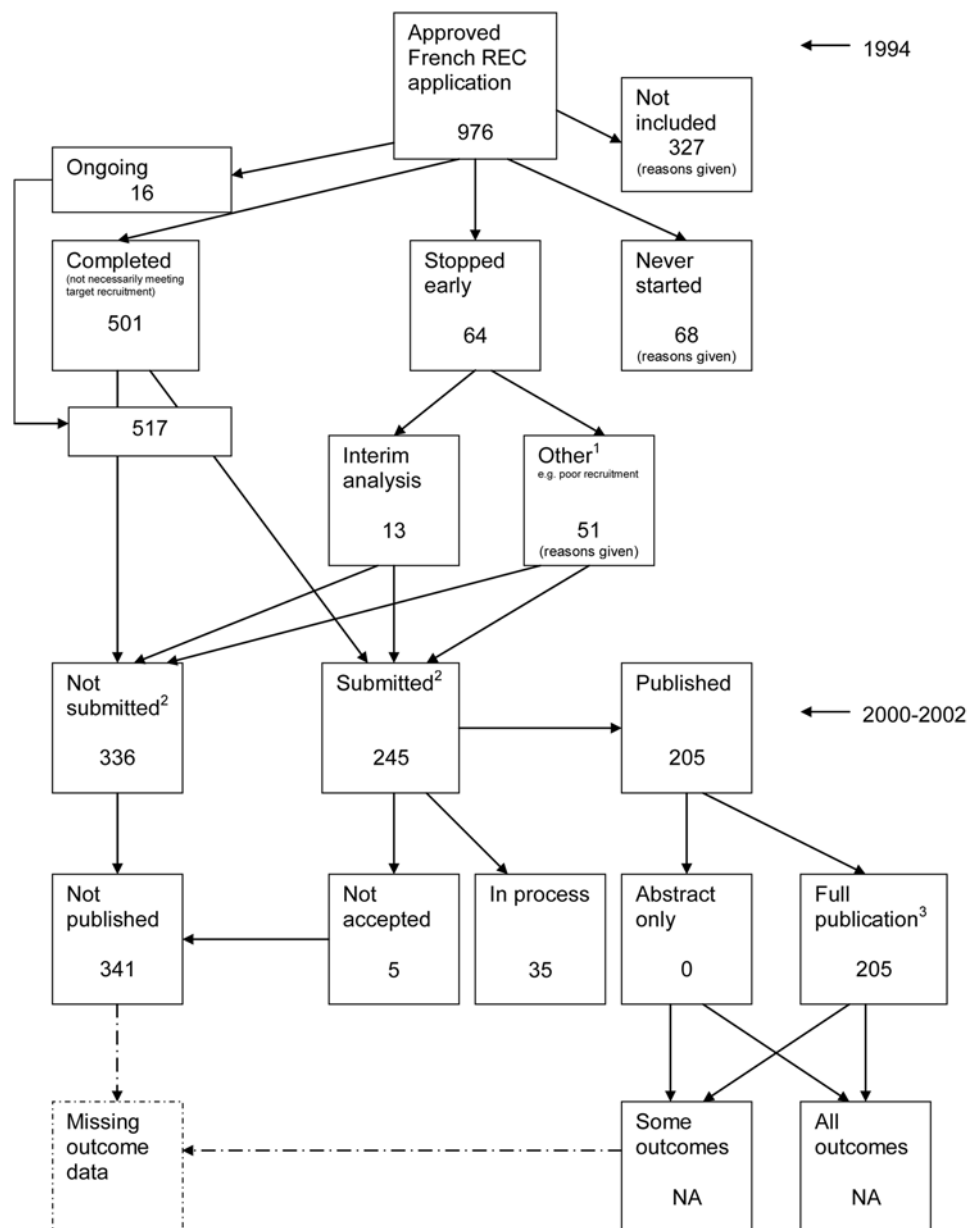
Unsure as to how many clinical trials there were among the 84 studies.

**Figure 12. Status of approved protocols for Cronin 2004 study [31].**  
doi:10.1371/journal.pone.0003081.g012

be submitted for publication than those with non-significant results. Easterbrook et al [26] also found that study publication bias was greater with observational and laboratory-based experimental studies (Odds Ratio (OR) 3.79, 95% CI; 1.47, 9.76) than with RCTs (OR 0.84, 95% CI; 0.34, 2.09). However, two empirical studies [21,31] found no statistically significant evidence

for study publication bias (RR 4 (95% CI 0.6, 32)  $p = 0.1$  and OR 0.53 (95% CI 0.25, 1.1)  $p = 0.1$ ).

Ioannidis et al [5] found that positive trials were submitted for publication more rapidly after completion than negative trials (median 1 vs 1.6 years,  $p < 0.001$ ) and were published more rapidly after submission (median 0.8 vs 1.1 years,  $p < 0.04$ ). Stern et al [4]



1. Reasons given for studies stopped early.
2. Unsure on the contribution to this number from studies stopped early. Only know that 8 were published. Reasons for not submitting are given.
3. Seven from ongoing and 8 from stopped early.

**Figure 13. Status of approved protocols for Decullier 2005 study [29].**  
doi:10.1371/journal.pone.0003081.g013

and Decullier et al [29] also considered time to publication and found that those studies with positive results were published faster than those with negative results (median 4.8 v 8.0 years [4] and HR 2.48 (95% CI 1.36, 4.55) [29], respectively).

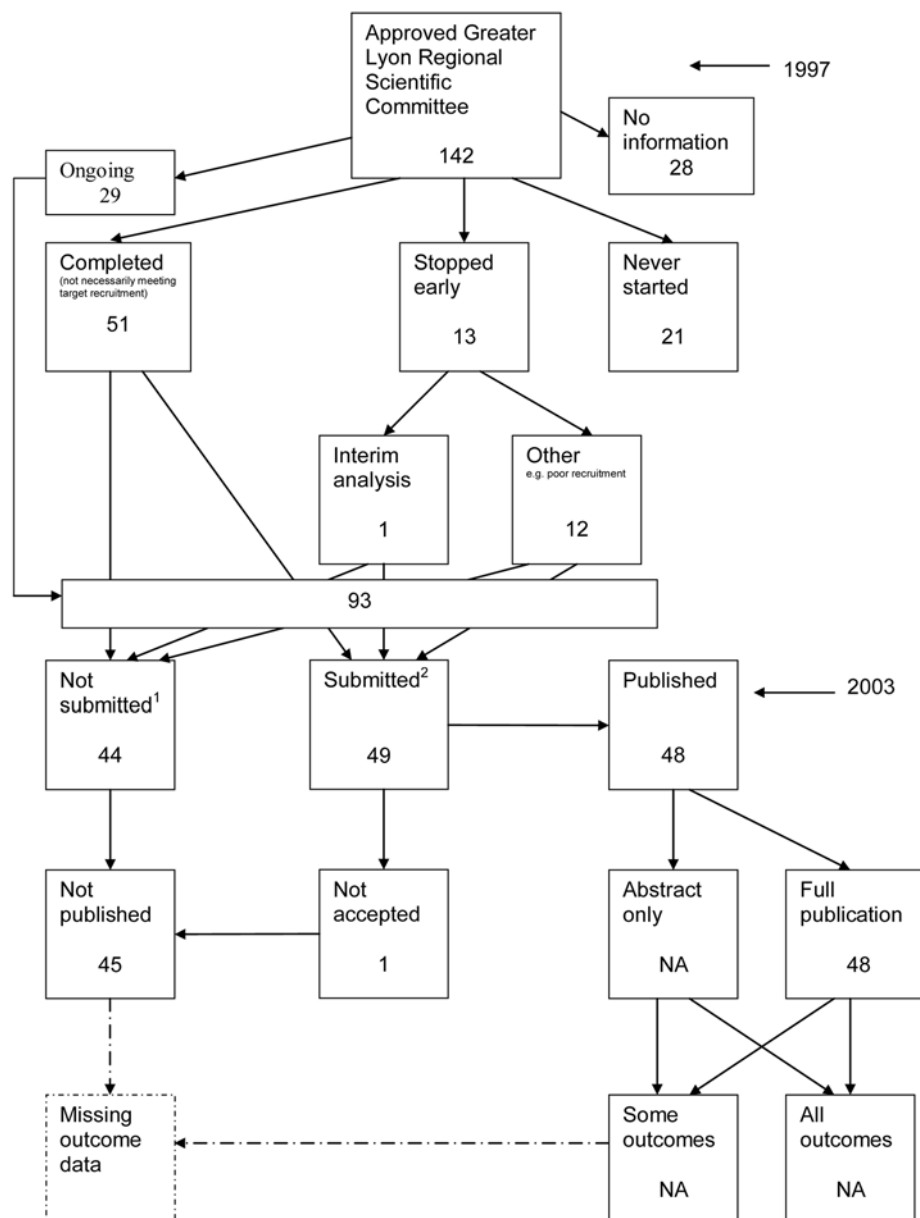
Pich et al [28] looked at whether studies in their cohort were completed and published; 64% (92/143) of initiated trials were finished in accordance with the protocol and 31% (38/123) were published (or in-press) in peer reviewed journals.

Seven empirical studies [3,21,26,27,29,30,32] described reasons why a study was not published as reported by the

trialists. Reasons related to trial results included: unimportant/null results; results not interesting; results not statistically significant.

**Outcome reporting bias.** The total number of studies published in each cohort varied from 37% to 67% (Table 3). However, none of the empirical studies investigating ORB considered the proportions of published trials with positive, negative, or null overall results.

Table 4 shows that three of the empirical studies [14,15,18] defined 'published' as a journal article; one empirical study [13]



1. The 44 "not submitted" consisted of 9 'stopped early', 17 'completed' and 18 'on-going'. These 44 contain the 11 in preparation/writing/submission in process and also the 2 with no article at all but with oral presentation.

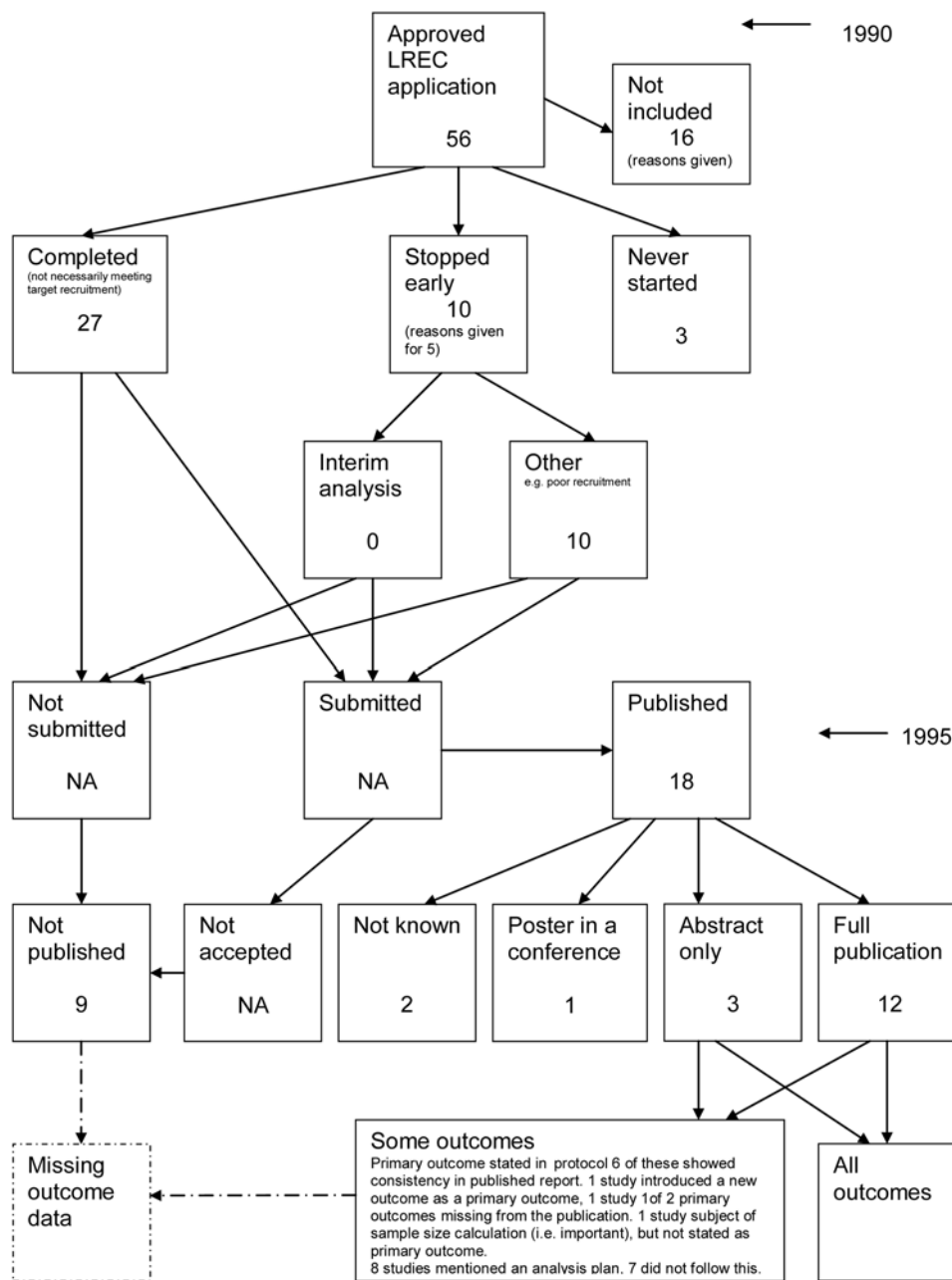
2. The 49 submitted consisted of 4 'stopped early', 34 'completed' and 11 'on-going'.

**Figure 14. Status of approved protocols for Decullier 2006 study [30].**  
doi:10.1371/journal.pone.0003081.g014

considered grey literature in their definition of 'published' although information on full publications and grey literature publications are separated (Figure 15). Although not considered in the definition of 'published', one empirical study [14] gave information on the grey literature or reports in preparation. Only two empirical studies [14,15] present results for the percentage of studies not submitted (31% to 56%), the percentage of studies submitted but not accepted (1 to 2%) by the time of analysis of the cohort and the percentage of studies not published that were not submitted (97% to 99%).

All four empirical studies [14,15,17,18] that examined the association between outcome reporting bias (outcome level bias) and statistical significance found that statistically significant outcomes were more likely to be completely reported than non-significant outcomes (range of odds ratios: 2.2 to 4.7 (Table 5)).

Five empirical studies [13–15,17,18] compared the protocol and the publication with respect to the primary outcome (Table 5). Only two empirical studies looked at the different types of discrepancies that can arise [14,15] and concluded that 40–62% of trials had major discrepancies between the primary



**Figure 15. Status of approved protocols for Hahn 2002 study [13].**  
doi:10.1371/journal.pone.0003081.g015

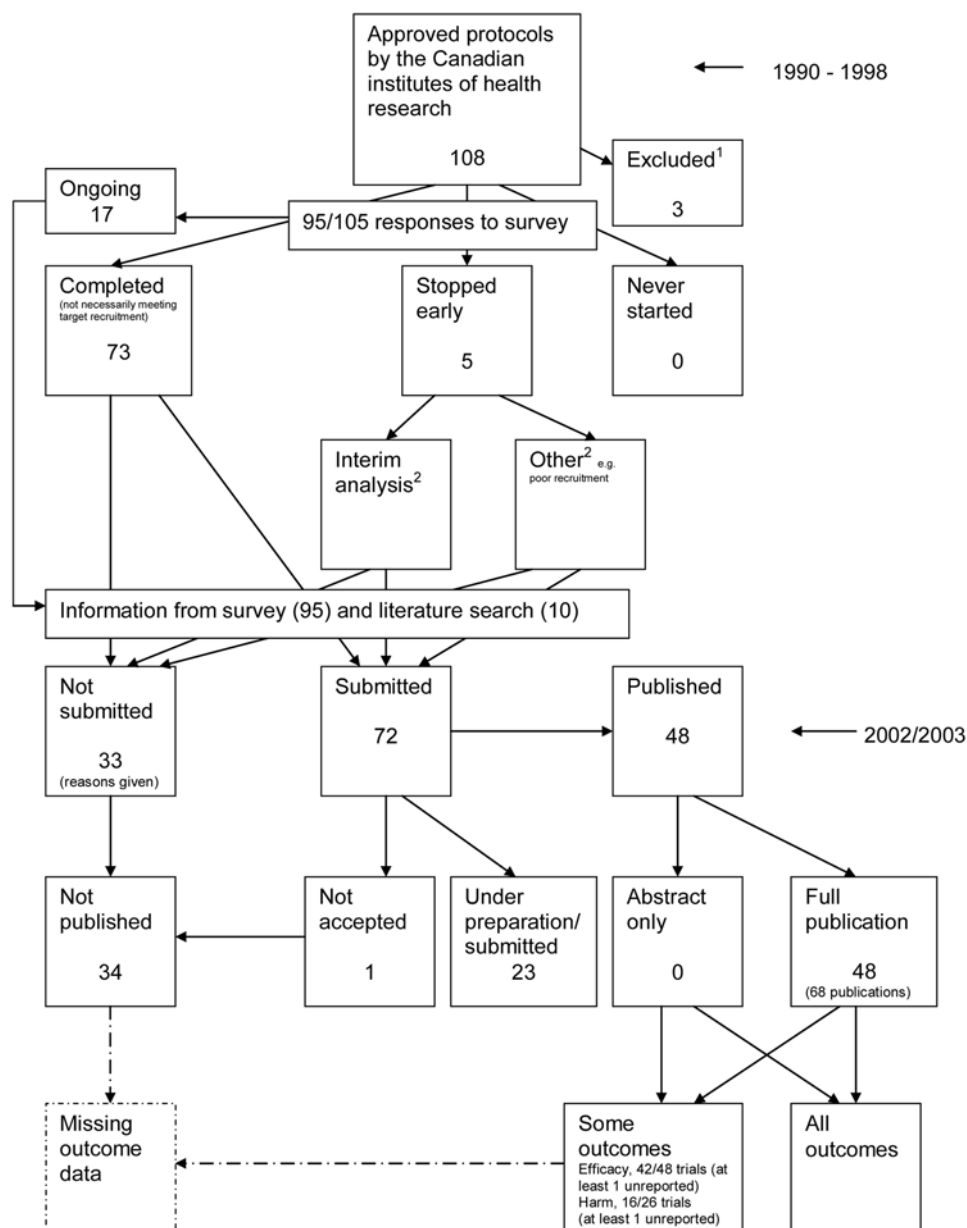
outcomes specified in protocols and those defined in the published articles. Four of the included empirical studies found that in 47–74% of studies the primary outcome stated in the protocol was the same as in the publication; between 13 and 31% of primary outcomes specified in the protocol were omitted in the publication and between 10 and 18% of reports introduced a primary outcome in the publication that was not specified in the protocol.

Chan et al also looked at efficacy and harm outcomes and in their Canadian empirical study [14] found that a median of 31% of efficacy outcomes and 59% of harm outcomes were incompletely reported and statistically significant efficacy outcomes had a higher odds than non significant efficacy outcomes of being fully reported (OR 2.7; 95% CI 1.5, 5). In their Danish empirical study

[15] they found that 50% of efficacy and 65% of harm outcomes per trial were incompletely reported and statistically significant outcomes had a higher odds of being fully reported compared with non significant outcomes for both efficacy (OR 2.4, 95% CI: 1.4, 4) and harm (OR 4.7, 95% CI: 1.8, 12) data.

von Elm et al [18] considered efficacy and harm outcomes as well as primary outcomes overall and found that 32% (223/687) were reported in the publication but not specified in the protocol and 42% (227/546) were specified in the protocol but not reported, however this is preliminary data.

Two empirical studies [14,15] describe the reasons why outcomes do not get reported but the study is published, these include lack of clinical importance and lack of statistical significance.

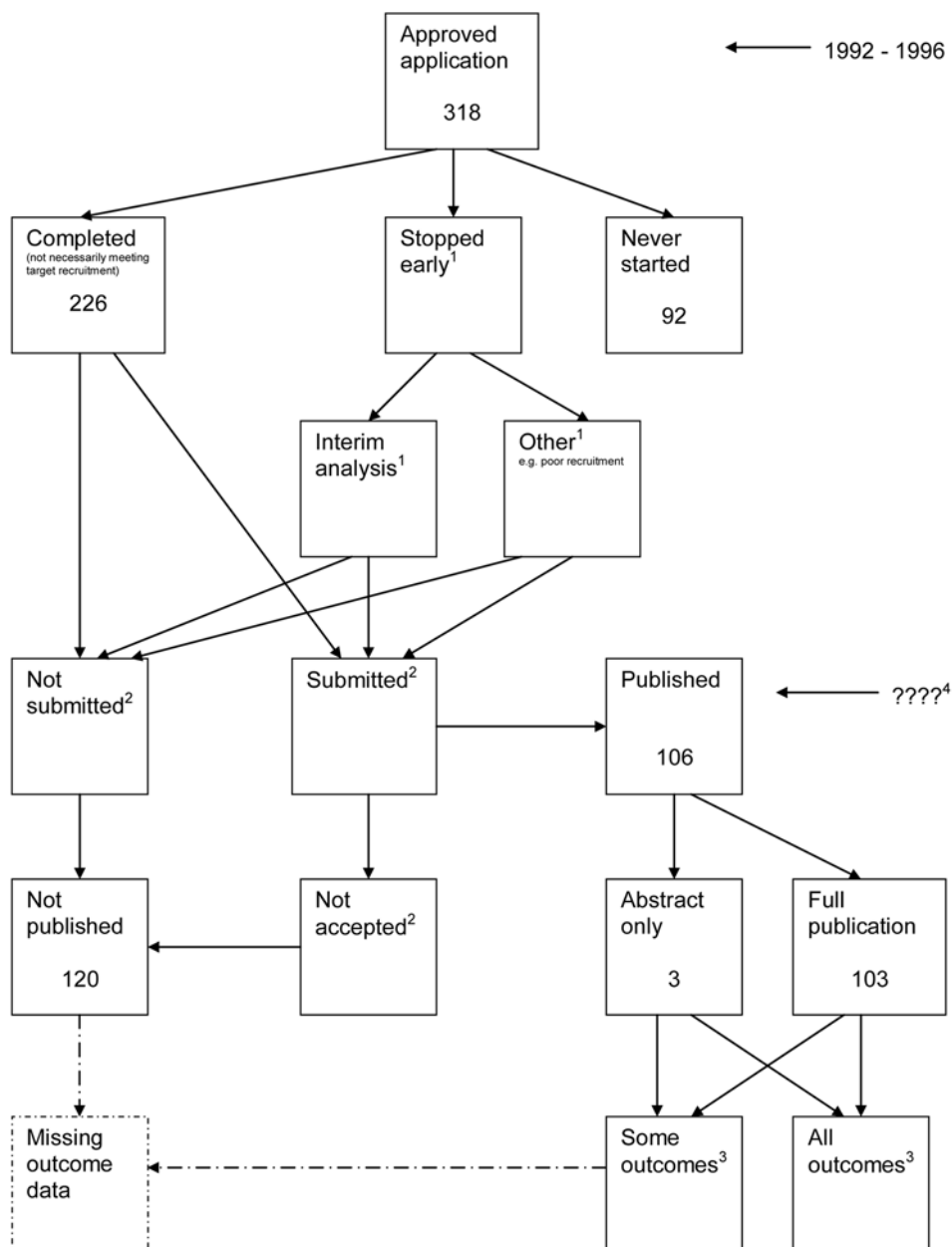


**Figure 16. Status of approved protocols for Chan 2004a study [14].**  
doi:10.1371/journal.pone.0003081.g016

## Discussion

Very few empirical studies examined both study publication bias and outcome reporting bias in the same cohort. However, 12 of the included empirical studies demonstrate consistent evidence of an association between positive or statistically significant results and publication. They suggest that studies reporting positive/statistically significant results are more likely to be published and that statistically significant outcomes have higher odds of being fully reported.

In this review we focused on empirical studies that included RCTs since they provide the best evidence of the efficacy of medical interventions [35]. RCTs are prone to study publication bias, but it has been shown that other types of studies are more prone to study publication bias [26]. The main limitation of this review was that for eight of the 16 included cohorts, information on RCTs could not be separated from information on other studies. Due to this barrier, and variability across empirical studies in the time lapse between when the protocol was approved and when the data were censored for analysis, we felt it was not



1. Unclear if any trials were stopped early and if this can be split into interim analyses and other.
2. Unclear if all trials were submitted.
3. It is difficult to report this based on information we had.
4. The date the study began is unclear.

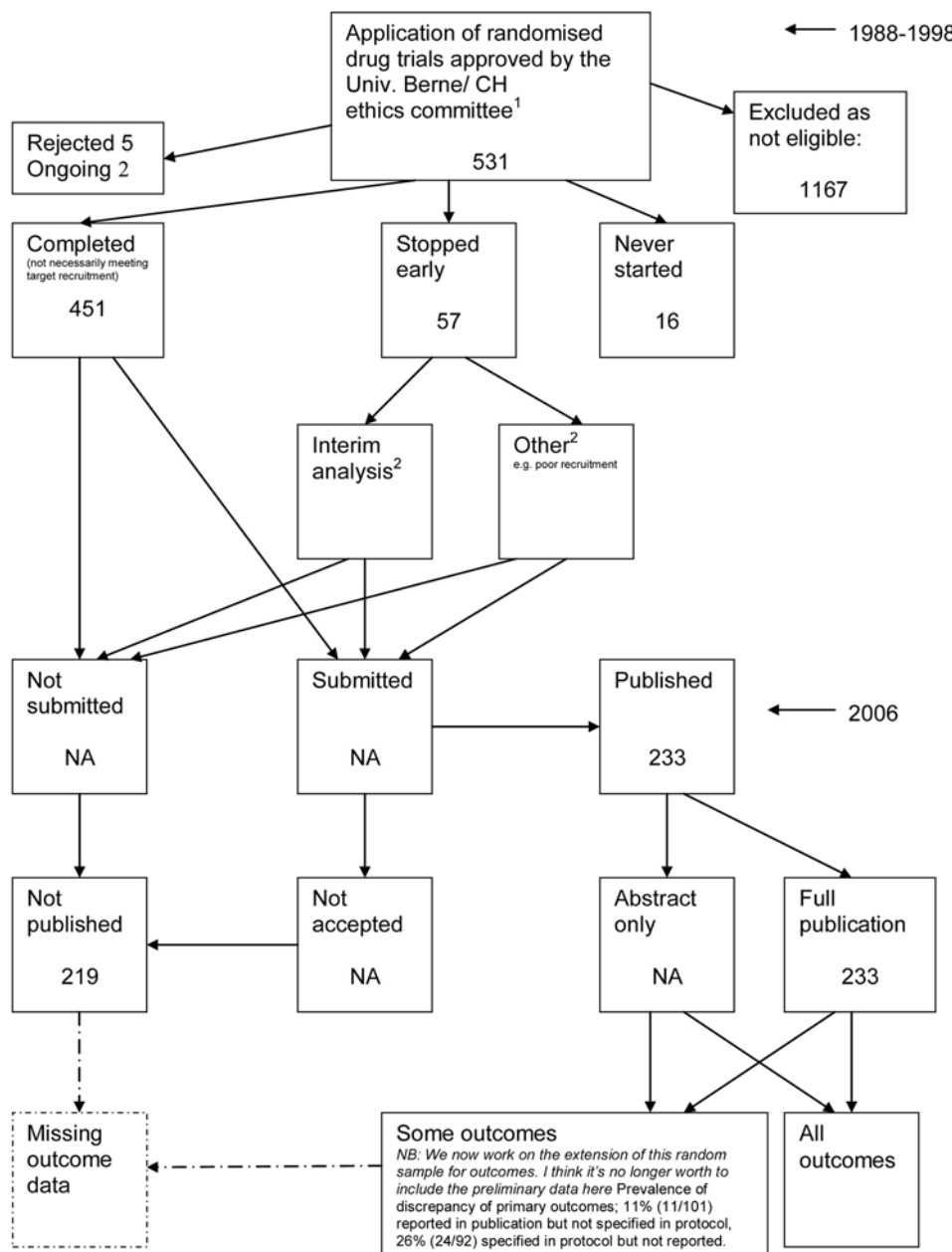
**Figure 17. Status of approved protocols for Ghera 2006 study [17].**  
doi:10.1371/journal.pone.0003081.g017

appropriate to combine statistically the results from the different cohorts. Also, the fact that in five empirical studies [3,4,26,29,30] follow-up of trials was less than 90% could mean that the problem of study publication bias is underestimated in these cohorts.

It is difficult to tell the current state of the literature with respect to study publication bias, as even the most recently published empirical evaluations included in the review, considered RCTs which began 10 years ago. Nevertheless, the empirical studies that

were published within the last eight years show that the total amount of studies published was less than 50% on average.

None of the empirical studies explored the idea of all outcomes being non-significant versus those deemed most important being non-significant. In the reasons given, it was not stated which outcomes/how many outcomes were non-significant. Some empirical studies imply that all results were non-significant although this is due to the way the reason was written i.e. no



1. Of initial sample of 1698 applications.
2. Studies stopped early were not analysed for whether decision was based on results of interim analysis. It was often unclear if formal interim analyses by a DSMB had led to termination.

**Figure 18. Status of approved protocols for von Elm 2008 study [18].**  
doi:10.1371/journal.pone.0003081.g018

significant results; but it is not explained whether this means for all outcomes, or primary and secondary, harm and efficacy etc. This implies a potential ambiguity of ‘no significant results’. It is not clear whether studies remain unpublished because all outcomes are non-significant and those that are published are so because significant results are selectively reported. This is where study publication bias and outcome reporting bias overlap.

Dubben et al [38] looked at whether study publication bias exists in studies which investigate the problem of study publication bias. Although they found no evidence of study publication bias, it

is interesting to note that two of the included cohorts in this review have not been published [17,21]. The study conducted by Wormald et al [21] concluded that ‘there was limited evidence of study publication bias’ whereas the authors of the other study [17] have not as yet had time to submit the study for publication. The inclusion and exclusion criteria were applied by only the first author, and there may be other unpublished studies of study publication bias or outcome reporting bias that were not located by the search, however contact with experts in the field reduces the likelihood of these issues introducing bias.



**Table 3.** Publication and trial findings.

Study ID	Total published (percentage)	Positive (percentage)	Negative (percentage)	Null (percentage)
Easterbrook, 1991 [26]	138/285 (48%)	93/154 (60%)	12/34 (35%)	33/97 (34%)
Dickersin, 1992 [27]	390/514 (76%)	260/314 (83%)	130/200 (65%)	NI
Dickersin, 1993 [3]	184/198 (93%)	121/124 (98%)	63/74 (85%)	NI
Stern, 1997 [4]	189/321 (59%)	153/232 (66%)	13/37 (35%)	23/52 (44%)
Cooper, 1997 [32]	38/121 (status known for 117/121) (31%)	-	-	-
Wormald, [21]	30/61 (status known for 39 completed trials) (49%)	14/15 (93%)	15/21 (71%)	NI
Ioannidis, 1998 [5]	36/66 (55%)	20/27 (74%)	16/39 (41%)	NI
Pich, 2003 [28]	26/123 (21%)	NI	NI	NI
Cronin, 2004 [31]	28/70 (40%)	NI	NI	NI
Decullier, 2005 [29]	205/649 (32%) (status known for 248 <sup>1</sup> )	129/188 (67%)	3/16 (19%)	14/44 (32%)
Decullier, 2006 [30]	48/93 (status known for 47/51 completed trials) (52%)	26/37 (70%)	6/10 (60%)	NI
Hahn, 2002 [13]	18/27 (67%)	NI	NI	NI
Chan, 2004a [14]	48/105 (46%)	NI	NI	NI
Chan, 2004b [15]	102/274 (37%)	NI	NI	NI
Gherzi, 2006 [17]	103/226 (46%)	NI	NI	NI
Von Elm, 2008 [18]	233/451 (52%)	NI	NI	NI

<sup>1</sup>Analysis restricted to 248 completed, non confidential, with hypothesis tests and direction of results.

NI No information, this study does not look at this.

- Not able to work out values.

Status implies positive or negative findings.

doi:10.1371/journal.pone.0003081.t003

Submission is an important aspect of investigating study publication bias as it will provide information on whether reports are not being published because they are not submitted or they are submitted but not accepted. Obviously those studies that are not submitted are not published and it was found by Dickersin et al [36] that non-publication was primarily a result of failure to write up and submit the trial results rather than rejection of submitted manuscripts. This is confirmed for the cohorts identified here with the percentage of studies not published due to not being submitted ranging from 63% to 100%. Olson et al [37] also found that there was no evidence that study publication bias occurred once manuscripts had been submitted to a medical journal. However, this study looks at a high impact general journal, which is unlikely to be representative for specialist journals that publish the majority of clinical trials.

Ten studies assessed the impact of funding on publication; this was done in several ways. Three studies found that external funding lead to a higher rate of publication [4,27,30]. von Elm et al [18] found that the probability of publication decreased if the study was commercially funded and increased with non commercial funding. Easterbrook et al [26] found that compared with unfunded studies, government funded studies were more likely to yield statistically significant results but government sponsorship was not found to have a statistically significant effect on the likelihood of publication and company sponsored trials were less likely to be published or presented. Dickersin et al [3] found no difference in the funding mechanism grant versus contract and Ioannidis et al [5] found no difference in whether data was managed by the pharmaceutical industry or other federally sponsored organisations. Chan 2004b et al [15] found that 61% of the 51 trials with major discrepancies were funded solely by industry sources compared with 49% of the 51 trials without discrepancies. Gherzi [17] did examine the effect of funding in terms of reporting and discrepancies of outcomes but no

information about the results is currently available. Hahn et al [13] compared the funder stated in protocol to publication. These studies indicate that funding is an important factor to consider when investigating publication bias and outcome reporting bias, however more work needs to be done to examine common questions before conclusions regarding the relationship between funding and outcome reporting bias can be drawn.

Our review has examined inception cohorts only, however, other authors have investigated aspects of study publication bias and outcome reporting bias using different study designs, with similar conclusions. The Cochrane review by Scherer et al [6] investigating the full publication of results initially presented in abstracts found that only 63% of results from abstracts describing randomized or controlled clinical trials are published in full and 'positive' results were more frequently published than non 'positive' results. Several studies investigated a cohort of trials submitted to drug licensing authorities [22,23,42] and all found that many of these trials remain unpublished, with one study demonstrating that trials with positive outcomes resulted more often in submission of a final report to the regulatory authority [22]. Olson et al [37] conducted a prospective cohort study of manuscripts submitted to JAMA and assessed whether the submitted manuscripts were more likely to be published if they reported positive results. They did not find a statistically significant difference in publication rates between those with positive and negative results. None of the inception cohorts addressed the question as to whether the significance determined whether a submitted paper was accepted or not, with the exception of one inception cohort [5] that found that "positive" trials were published significantly more rapidly after submission than "negative" trials. Finally, a comparison of the published version of RCTs in a specialist clinical journal with the original trial protocol found that important changes between protocol and published paper are common; the published primary outcome was

**Table 4.** Aspects of study level publication bias.

Study	Definition of published article employed by eligible studies	Time since protocol approved	% Not submitted <sup>1</sup> (95% CI)	% Submitted but not accepted <sup>2</sup> (95% CI)	% Studies not published that were not submitted (95% CI)
Easterbrook, 1991 [26]	Acceptance by a journal (published or in press), but not book chapters or published meeting abstracts or proceedings.	3 years	9 (6, 12)	6 (3, 9)	63 (49, 77)
Dickersin, 1992 [27]	Reported in one or more journal articles, monographs, books, or chapters in books, were available from medical libraries or were in documents available from a published archive, e.g., the National Technical Information service	8 years	23 (19, 27)	1.5 (0, 3)	95 (91, 99)
Dickersin, 1993 [3]	Abstracts, journal articles, book chapters, proceedings, letter to the editor, or other material.	9 years	7 (3, 11)	0	100
Stern, 1997 <sup>3</sup> [4]	As an article in a peer reviewed journal	4 plus years	33 (28, 38)	13 (8,17)	80 (73, 87)
Cooper, 1997 [32]	NI	NI	NI		
Wormald, 1997 [21]	NI	2 to 34 years	21 (8,34)	0	100
Ioannidis, 1998 [5]	Peer-reviewed journal articles	12% completed less than 1 year ago, 20% completed over 1 year ago.	32 (21, 43)	20 (8, 32)	70 (54, 86)
Pich, 2003 [28]	Peer reviewed journal articles	3 years	*		
Cronin, 2004 [31]	NI	1 to 5 years	*		
Decullier, 2005 <sup>3</sup> [29]	Published as a scientific paper. Grey literature – published in other formats than scientific papers, that is generally not accessible through libraries (internal reports, theses, abstracts, posters)	6 years (2% not ready)	58 (54,62)	2 (0, 4)	99 (98, 100)
Decullier, 2006 <sup>3</sup> [30]	Published as a scientific paper	6 years	47 (37,57)	2 (0, 6)	98 (95, 100)
Hahn, 2002 [13]	Published article or a written study report	5 years	*		
Chan, 2004a <sup>3</sup> [14]	Journal article	4 years	31 (22,40)	1 (0, 3)	97 (91, 100)
Chan, 2004b <sup>3</sup> [15]	Journal article	8 years	56 (50, 62)	2 (0, 5)	99 (97, 100)
Ghersi, 2006 [17]	NI	NI	NI		
Von Elm, 2008 [18]	Journal article	8 years	*		

<sup>1</sup>Denominator is the total number of studies with known publication status in the cohort i.e. published plus unpublished plus in preparation plus in press.

<sup>2</sup>Denominator is the number submitted.

<sup>3</sup>Denominator includes ongoing studies.

\*Question not asked.

NI No information available.

doi:10.1371/journal.pone.0003081.t004

**Table 5.** Comparisons of primary outcome stated in protocol and in publication.

Study	Study level		Outcome level		Primary outcome stated in the same as in the publication	Primary outcome stated in protocol is downgraded to secondary in the publication	Primary outcome omitted from the publication	A non primary outcome in the protocol is changed to primary in the publication	A new primary outcome that was not stated in the protocol (as primary or secondary) is included in the publication	The definition of the primary outcome was different in the protocol compared to the publication	Completeness of reporting	Overall	Definition of disagreement/ discrepancy
	Are studies with statistically significant or positive results, more likely to be published than those finding no difference between the study group?		Primary outcome stated in the same as in the publication	Primary outcome stated in protocol is downgraded to secondary in the publication									
Easterbrook, 1991 [26]	OR 2.32, 95% CI: 1.25, 4.28.	*	*	*	*	*	*	*	*	*	*	*	*
Dickersin, 1992 [27]	OR 2.54, 95% CI: 1.63, 3.94	*	*	*	*	*	*	*	*	*	*	*	*
Dickersin, 1993 [3]	OR 12.30, 95% CI: 2.54, 60	*	*	*	*	*	*	*	*	*	*	*	*
Stern, 1997 [4]	HR 2.32, 95% CI: 1.47, 3.66, $p = 0.0003$	*	*	*	*	*	*	*	*	*	*	*	*
Cooper, 1997 [32]	$p < 0.0001$ (submission only)	*	*	*	*	*	*	*	*	*	*	*	*
Wormald, 1997 [21]	RR 4, 95% CI: 0.6, 32, $p = 0.10$	*	*	*	*	*	*	*	*	*	*	*	*
Ioannidis, 1998 [5]	*	*	*	*	*	*	*	*	*	*	*	*	*
Pich, 2003 [28]	*	*	*	*	*	*	*	*	*	*	*	*	*
Cronin, 2004 [31]	OR 0.53, 95% CI: 0.25, 1.1 $p = 0.1$	*	*	*	*	*	*	*	*	*	*	*	*
Decullier, 2005 <sup>2</sup> [29]	OR 4.59, 95% CI: 2.21, 9.54	*	*	*	*	*	*	*	*	*	*	*	*
Decullier, 2006 [30]	OR 1.58, 95% CI: 0.37, 6.71	*	*	*	*	*	*	*	*	*	*	*	*
Hahn, 2002 [13]	*	40% (6/15) stated which outcome variables were of primary interest and 4 of these (67%) showed consistency in the reports	*	17% (1/6)	*	17% (1/6)	*	17% (1/6)	*	*	*	*	NA

Table 5. cont.

Study	Study level	Outcome level					The definition of the primary outcome was different in the protocol compared to the publication	Completeness of reporting	Overall	Definition of disagreement/discrepancy
		Are studies with statistically significant or positive results, more likely to be published than those finding no difference between the study group?	Primary outcome stated in protocol is the same as in the publication	Primary outcome stated in protocol is downgraded to secondary in the publication	Primary outcome stated in the protocol is omitted from the publication	A non primary outcome in the protocol is changed to primary in the publication				
Chan, 2004a <sup>1</sup> [14]	*	67% (32/48)	23% (11/48)	13% (6/48)	9% (4/45)	18% (8/45)	36 trials reported a power calculation based on a particular outcome in their publications; 2 used an outcome that differed from the one used in the protocol and another introduced a power calculation that had not been mentioned in the protocol	Primary outcomes were incompletely reported in 7 (16%) of 45 trials that defined such outcomes in their publications	40% (19/48) of the trials contained major discrepancies in the specification of primary outcomes between the protocols and publications. None of the publications stated that an amendment had been made to the protocol.	The failure to report a pre-specified primary outcome; reporting of a prespecified primary outcome as secondary or as neither primary nor secondary in the publication; the introduction of new primary outcomes in the publication; and changing the outcome specified for the power calculation.
Chan, 2004b <sup>1</sup> [15]	*	47% (36/76)	34% (26/76)	26% (20/76)	19% (12/63)	17% (11/63)	38 trials reported a power calculation, but 4 calculations were based on an outcome other than the one used in the protocol. In another 6 cases, there was a power calculation presented in a published article but not in the protocol	Primary outcomes were specified for 63 of the published trials, but for 17 (27%) of these trials at least one primary outcome was incompletely reported.	62% (51/82) of trials had major discrepancies between primary outcomes specified in protocols and those defined in the published articles (at least 1 primary outcome that was changed, introduced or omitted).	A prespecified primary outcome was reported as secondary or was not labelled as either; a prespecified primary outcome was omitted from the published articles; a new primary outcome was introduced in the published articles and the outcome used in the power calculation was not the same in the protocol and the published articles.

Table 5. cont.

Study	Study level	Outcome level					Overall	Definition of disagreement/ discrepancy		
	Are studies with statistically significant or positive results, more likely to be published than those finding no difference between the study group?	Primary outcome stated in the same as in the publication	Primary outcome stated in protocol is downgraded to secondary in the publication	Primary outcome stated in the publication	A non primary outcome in the protocol is changed to primary in the publication	A new primary outcome that was not stated in the protocol (as primary or secondary) is included in the publication	The definition of the primary outcome was different in the protocol compared to the publication	Completeness of reporting	Overall	
Ghersi, 2006 [17]	*	74%	*	31%	*	10%	60% definition the same, 31% unable to judge and 8% discrepant	The odds of fully reporting a comparison were greater if the result was statistically significant, (OR 2.2, 95% CI: 0.9–5.3, $p = 0.08$ )	*	NA
Von Elm, 2008 [18]	*	*	*	26% (24/92)	*	11% (11/101)	*	*	*	NA

NA Not Applicable.

\*The study did not investigate this.

<sup>1</sup>Gives results for trials with discrepancies for  $\geq 1$  primary outcome.<sup>2</sup>Results obtained from 248 of the completed studies.

doi:10.1371/journal.pone.0003081.t005

exactly the same as in the protocol in six out of 26 trials (23%) [43].

We recommend that researchers use the flow diagram presented in this work as the standard for reporting of future similar studies that look at study publication bias and ORB as it clearly shows what happens to all trials in the cohort.

Reviewers should scrutinise trials with missing outcome data and ensure that an attempt to contact trialists is always made if the study does not report results. Also, the lack of reporting of specified outcome(s) should not be an automatic reason for exclusion of studies. Statisticians should be involved for the data extraction of more complex outcomes, for example, time to event. Methods that have been developed to assess the robustness of the conclusions of systematic reviews to ORB [44,45] should be used. Meta-analyses of outcomes where several relevant trials have missing data should be seen with extra caution. In all, the credibility of clinical research findings may decrease when there is wide flexibility in the use of various outcomes and analysis in a specific field and this is coupled with selective reporting biases.

The setting up of clinical trials registers and the advance publication of detailed protocols with an explicit description of outcomes and analysis plans should help combat these problems. Trialists should be encouraged to describe legitimate changes to outcomes stated in the protocol. With the set up of online journals, where more space is available, trialists should be encouraged to write up and submit for publication without selection of results.

For empirical evaluations of selective reporting biases, the definition of significance is important as is whether the direction of the results is taken into account, i.e. whether the results are significant for or against the experimental intervention. However, only one study took this into account [5]. The selective publication preference forces may change over time. For example, it is often seen that initially studies favouring treatment are more likely to be published and those favouring control suppressed. However, as time passes, contradicting trials that favour control may become

attractive for publication, as they are 'different.' The majority of cohorts included in this review do not consider this possibility.

Another recommendation is to conduct empirical evaluations looking at both ORB and study publication bias in RCTs to investigate the relative importance of both i.e. which type of bias is the greater problem. The effects of factors such as funding, i.e. the influence of pharmaceutical industry trials versus non pharmaceutical trials, should also be factored in these empirical evaluations.

## Supporting Information

### Appendix S1 Search Strategy.

Found at: doi:10.1371/journal.pone.0003081.s001 (0.03 MB DOC)

### Text S1 Explanation of flow diagram.

Found at: doi:10.1371/journal.pone.0003081.s002 (0.03 MB DOC)

## Acknowledgments

The authors would like to thank Sally Hopewell for kindly providing her reference list, Simone Menzel for further information on their study and the referees who reviewed this work for their helpful comments.

## Author Contributions

Conceived and designed the experiments: KD DGA CG PW. Performed the experiments: KD. Analyzed the data: KD. Contributed reagents/materials/analysis tools: JAA JB AWC EC ED PE EvE CG DG JPAI JS PW. Wrote the paper: KD DGA. Performed the literature search and data extraction, contacted all authors of included studies and wrote the first draft: KD. Contributed to the development of the protocol: KD DGA CG PW. Provided comments on the manuscript: DGA ED EvE CG JPAI JS PW. Gave permission to include data from their study in the review, responded to queries and provided extra information when needed and available: JAA JB AWC EC ED PE EvE DG JPAI.

## References

- Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ (2000) Publication and related biases. *Health Technol Assess* 4(10).
- Rothstein H, Sutton AJ, Borenstein M (2005) *Publication bias in meta-analysis: prevention, assessment and adjustments*. Chichester: Wiley.
- Dickersin K, Min YI (1993) NIH clinical trials and publication bias. *Online Journal of Curr Clin Trials* doc no 50.
- Stern JM, Simes RJ (1997) Publication bias: evidence of delayed publication in a cohort of clinical research projects. *BMJ* 315: 640–645.
- Ioannidis JPA (1998) Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA* 279: 281–6.
- Scherer RW, Langenberg P, von Elm E (2007) Full publication of results initially presented in abstracts. *Cochrane Database of Systematic Reviews* 2007, Issue 2. Art. No.: MR000005. DOI: 10.1002/14651858.MR000005.pub3.
- Ioannidis JPA, Lau J (2001) Evolution of treatment effects over time: Empirical insight from recursive cumulative meta analyses. *PNAS* 98 (3): 831–836.
- Trikalinos TA, Churchill R, Ferri M, Leucht S, Tuunainen A, et al. (2004) Effect sizes in cumulative meta-analyses of mental health randomized trials evolved over time. *J Clin Epidemiol* 57: 1124–1130.
- Hutton JL, Williamson PR (2000) Bias in meta-analysis due to outcome variable selection within studies. *Appl Stat* 49(3): 359–370.
- Williamson PR, Gamble C, Altman DG, Hutton JL (2005) Outcome selection bias in meta-analysis. *Stat Methods Med Res* 14: 515–524.
- Kane JL, Wang J, Garrard J (2007) Reporting in randomised clinical trials improved after adoption of the CONSORT statement. *J Clin Epidemiol* 60: 241–249.
- Tannock IF (1996) False-Positive Results in Clinical Trials: Multiple Significance Tests and the Problem of Unreported Comparisons. *J Natl Cancer Inst* 88 (3/4): 206–207.
- 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; 3: 353–359.
- Chan AW, Krijeza-Jeric K, Schmid I, Altman DG (2004a) Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health research. *CMAJ* 171 (7): 735–740.
- Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG (2004b) Empirical Evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 291 (20): 2457–2465.
- Chan AW, Altman DG (2005) Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ* 330: 753–759.
- Ghersi D (2006) Issues in the design, conduct and reporting of clinical trials that impact on the quality of decision making. Thesis (Ph. D.)—School of Public Health, Faculty of Medicine, University of Sydney.
- Von Elm E, Rollin A, Blümle A, Huwiler K, Witschi M, et al. (2008) Selective reporting of outcomes of drug trials; Comparison of study protocols and published articles. *Swiss Med Wkly* 138 (13–14): 197–203.
- Contopoulos-Ioannidis DG, Alexiou GA, Gouvas TC, Ioannidis JPA (2006) An empirical evaluation of multifarious outcomes in pharmacogenetics: beta-2 adrenoceptor gene polymorphisms in asthma treatment. *Pharmacogenet Genomics* 16(10): 705–711.
- Olsen KL, Hopewell S, Dickersin K, Clarke M, Oxman AD (2001) Publication bias in clinical trials. (Protocol) *Cochrane Database of Methodology Reviews* 2001, Issue 3. Art. No.: MR000006. DOI: 10.1002/14651858.MR000006.
- Wormald R, Bloom J, Evans J, Oldfield K (1997) Publication bias in eye trials. In: 2<sup>nd</sup> International Conference Scientific Basis of Health Services & 5<sup>th</sup> Annual Cochrane Colloquium, 1997 5–8 October; Amsterdam, The Netherlands.
- Bardy AH (1998) Bias in reporting clinical trials. *Br J Clinical Pharmacol* 46: 147–150.
- Melander H, Ahlqvist-Rastad, Meijer G, Beermann B (2003) Evidence based medicine – selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ* 326: 1171–1173.
- Menzel S, Uebing B, Hucklenbroich P, Schober O (2007) Evaluation of clinical trials following an approval from a research ethics committee. *Dtsch Med Wochenschr* 132(44): 2313–7. German.
- Misakian AL, Bero LA (1998) Publication bias and research on passive smoking. *JAMA* 280: 250–3.
- Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR (1991) Publication bias in clinical research. *Lancet* 337: 867–72.

27. Dickersin K, Min YI, Meinert CL (1992) Factors influencing publication of research results: follow up of applications submitted to two institutional review boards. *JAMA* 267: 374–8.
28. Pich J, Carne X, Arnaiz JA, Gomez B, Trilla A, et al. (2003) Role of a research ethics committee in follow-up and publication of results. *Lancet* 361: 1015–1016.
29. Decullier E, Lheritier V, Chapuis F (2005) Fate of biomedical research protocols and publication bias in France: retrospective cohort study. *BMJ* 331: 19–24.
30. Decullier E, Chapuis F (2006) Impact of funding on biomedical research: a retrospective cohort study. *BMC public health* 6: 165.
31. Cronin E, Sheldon T (2004) Factors influencing the publication of health research. *Int J Technol Assess Health Care* 20: 351–5.
32. Cooper H, DeNeve K, Charlton K (1997) Finding the missing science: the fate of studies submitted for review by a human subjects committee. *Psychol Methods* 2 (4): 447–452.
33. Dickersin K (1997) How important is publication bias? A synthesis of available data. *AIDS Educ Prev* 9: 15–21.
34. Haidich AB, Ioannidis JP (2001) Effect of early patient enrollment on the time to completion and publication of randomized controlled trials. *Am J Epidemiol* 154(9): 873–80.
35. Egger M, Smith GD, Altman DG (2001) Systematic reviews in health care; meta-analysis in context. *BMJ Publishing Group*.
36. Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H JR (1987) Publication bias and clinical trials. *Contol Clin Trials* 8: 343–353.
37. Olson C, Rennie D, Cook D, Dickersin K, Flanagan A, et al. (2002) Publication bias in editorial decision making. *JAMA* 287 (21): 2825–2828.
38. Dubben HH, Beck-Bornholdt HP (2005) Systematic review of publication bias in studies on publication bias. *BMJ* 331: 433–434.
39. Huwiler-Muntener K, Juni P, Junker C, Egger M (2002) Quality of reporting of randomized trials as a measure of methodologic quality. *JAMA* 287 (21): 2801–2804.
40. Soares H P, Daniels S, Kumar A, Clarke M, Scott C, et al. (2004) Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group. *BMJ* 328: 22–24.
41. Soares H P, Kumar A, Daniels S, Swann S, Cantor A, et al. (2005) Evaluation of new treatments in radiation oncology, are they better than standard treatments? *JAMA* 293 (8): 970–978.
42. Hemminki E (1980) Study of information submitted by drug companies to licensing authorities. *BMJ*. pp 833–836.
43. Chappell L, Alfievic Z, Chien P, Jarvis S, Thornton JG (2005) A comparison of the published version of randomized controlled trials in a specialist clinical journal with the original trial protocols. *Proceedings of the 5<sup>th</sup> International congress on peer review and biomedical publication*. Chicago (unpublished).
44. Williamson PR, Gamble C (2005) Identification and impact of outcome selection bias in meta-analysis. *Stat Med* 24: 1547–1561.
45. Williamson PR, Gamble C (2007) Application and investigation of a bound for outcome reporting bias. *Trials* 8: 9.

Evidence of the personal communications can be provided upon request.