Outcome reporting bias in randomised trials:

Implications for systematic reviews

An-Wen Chan, MD

Hertford College

Doctor of Philosophy thesis

Trinity term 2003

Acknowledgements

Firstly, I would like to thank my supervisor, Professor Douglas Altman, who has been tremendously inspiring as a tutor and friend, and who provided constant guidance throughout my years in Oxford. I consider myself fortunate to have learned through his expertise, wisdom, and good humour.

This thesis is dedicated to my parents, Professor Wah-Chun and Yu-Chih Chan, who have offered unwavering support and encouragement throughout my education and development.

I am indebted to Jonathan Deeks, Mike Clarke, Diana Elbourne, and Mike Bradburn, all of whom provided valuable advice on various methodological aspects of my thesis. I would like to thank Peter Gøtzsche, Asbjørn Hróbjartsson, and Mette Haahr from the Nordic Cochrane Centre for all of their input and support; their efforts were crucial in making the Copenhagen study a success. I also wish to express my gratitude to Karmela Krleza-Jeric and Isabelle Schmid for their support at the Canadian Institutes of Health Research. At the Centre for Statistics in Medicine, Tracy Edwards and Ellen Storey provided invaluable administrative and moral support throughout my years in Oxford.

I am also grateful to the rest of my family - Eileen, Jean, Vivian, and Aaron - and my friends for their constant support and encouragement, and for helping me to enjoy a life outside of academics. Finally, I would like to thank the Rhodes Trust for funding my education at Oxford.

Abstract

Selective reporting of outcomes in randomised trials: Implications for systematic reviews

An-Wen Chan, MD
Hertford College

DPhil. thesis
Trinity term 2003

Background

Selective reporting of outcomes within a published study based on their nature or direction can result in systematic differences between reported and unreported data. Direct evidence of outcome reporting bias is limited to case reports.

Objective

To study empirically the nature of outcome reporting bias in randomised controlled trials (RCTs).

Methods

Three cohorts of RCTs were identified: PubMed-indexed RCTs published in December 2000; trial protocols approved by a Danish ethics committee from 1994-95; and trial protocols funded by a government agency in Canada from 1990-98. Data on reported and unreported outcomes were recorded from all trial publications and a survey of authors. An outcome was considered incompletely reported if insufficient data were presented for meta-analysis. Odds ratios relating the completeness of outcome reporting to statistical significance were calculated for each trial, and then pooled using a random

effects meta-analysis. Protocols and publications were also reviewed for discrepancies in primary outcome reporting.

Results

519 trials with 10,557 outcomes, 102 trials with 3613 outcomes, and 48 trials with 1390 outcomes were identified for the PubMed, ethics committee, and funding agency cohorts respectively. 22%-35% of outcomes per parallel group study were, on average, incompletely reported for meta-analysis. Fully reported outcomes had a two- to three-fold higher odds of being statistically significant compared to incompletely reported outcomes. The most common reasons given for omitting outcomes included a lack of clinical importance, lack of statistical significance, and space constraints. Major discrepancies between primary outcomes in protocols and publications were found in one half of trials.

Discussion and conclusions

The reporting of trial outcomes is frequently inadequate for meta-analysis; is biased to favour statistical significance; and is inconsistent with pre-specified protocol outcomes. Unacknowledged modifications to outcomes specified in trial protocols constitute scientific misconduct. Meta-analyses may therefore produce inflated and unreliable estimates of treatment effect.

Table of contents

List of tables

List of figures

Chapter 1 - Background and thesis rationale

1.1. Introduction

Randomised controlled trials have become the foundation for systematic reviews and evidence-based clinical practice. A search of the Cochrane Controlled Trials Register (Issue 3, 2002) yielded 17,074 clinical trial publications in the year 2000 compared to 14,725 in the year 1990 - an increase of 23%. Randomised trials over the past fifty years have helped to increase life expectancy by three years, while adding another five years of improved quality of life from chronic disease.[1]

The use of systematic reviews, particularly involving randomised trials, has become increasingly widespread in guiding evidence-based medical practice and policy.[2,3] Touted as being more rigorous, more objective, and less biased than traditional narrative reviews, systematic reviews have been ranked above randomised trials as the most reliable form of evidence for healthcare interventions.[4] Nonetheless, as with other study designs, reviews are open to potential bias at all stages of design, conduct, and reporting.

The reliability of a systematic review depends to a large extent on the identification and incorporation of all available evidence.[5] Reporting biases pose an obstacle to the reliability of systematic reviews by making available only select segments of the overall existing information. While the selective publication of entire studies (termed study publication bias) has received widespread recognition,[6] the selective reporting of outcomes within published studies (termed outcome reporting bias) has not undergone comprehensive

study. As a result of this lack of empirical assessment, questions remain as to the prevalence and extent of selective outcome reporting in the medical literature. This thesis examines the issue of bias in the selection of outcomes to report in publications, and aims to gather evidence for the nature and magnitude of such bias across cohorts of randomised trials.

This opening chapter will outline the existing evidence for publication biases at the between- and within-study levels. The following two chapters will describe the methodologies used to evaluate outcome reporting bias in randomised trials and the difficulties encountered in applying them. Chapters 4-7 will then present the thesis results and assess their validity. The concluding chapter will discuss the implications of these findings and suggest some methods to address the issues along with possibilities for future research.

1.2. Evidence of publication bias

Systematic reviews incorporating meta-analysis of available trials can provide an overall pooled estimate of the efficacy and safety of healthcare interventions. Although systematic reviews provide several advantages in terms of greater objectivity, transparency, and comprehensiveness compared to the traditional narrative review, many problems still plague their ability to provide reliable, unbiased evidence for healthcare interventions.

Discrepancies have been noted between the results of meta-analyses and large randomised controlled trials,[7-9] as well as between meta-analyses on the same topic.[10,11]

As with other types of research, each stage of a systematic review is susceptible to potential bias. One of the major problems stems from the inability to identify all of the data available on a particular intervention, resulting in potential selection bias. In relying on systematic reviews to guide evidence-based decisions, it must often be assumed that the medical literature is sufficiently representative of the entire body of knowledge available at the time. This assumption, however, has been put into question over the past two decades by strong evidence of biased under-reporting of medical research.[6]

Reporting biases pose some of the greatest problems to systematic reviews by creating an unrepresentative sample of outcomes from which conclusions are drawn. Such bias occurs when the dissemination of a study report depends systematically on the nature and/or direction of its results. Reporting bias encompasses a wide range of problems (Table 1-1). Statistically significant results are not only more likely to be published (publication bias), but also tend to be published sooner,[12,13] published in higher impact journals,[14,15] and cited more frequently by other papers.[16] Empirical research has shown that meta-analytic estimates of treatment effects tend to be decreased by the inclusion of studies that are unpublished, and inflated by incorporating studies that are reported in non-English language journals.[5]

'Publication bias' refers to the selective publication of research results based on their characteristics, resulting in systematic differences between published and unpublished data.[17] This can occur at two levels: selective publication of entire studies among a population of studies (study publication bias), and selective reporting of measured outcomes within a single published study

(outcome reporting bias). Selective reporting of analyses (analysis reporting bias), such as subgroup or per-protocol analyses, can also occur within published studies, but this type of bias will not be specifically addressed in the thesis.

Table 1-1. Types and definitions of reporting biases

Type	Occurs when:
Study publication bias[17]	Publication of studies is associated with the nature and/or direction of results
Outcome reporting bias[18]	Reporting of outcomes within published studies is associated with the nature and/or direction of results
Analysis reporting bias[19,20]	Reporting of analyses based on particular groups of participants within published studies is associated with the nature and/or direction of results
Time lag bias[18]	The time to publication of studies is associated with the nature and/or direction of results
Full publication bias[6]	Publication of initial abstracts or informal presentations as full papers is associated with the nature and/or direction of results
Language bias[18]	The language in which a study is published is associated with the nature and/or direction of results
Citation/reference bias[18]	The probability of citation by others is associated with the nature and/or direction of results
Grey literature bias[6]	Systematic differences exist between journal publications and studies reported in abstracts, dissertations, or working papers
Multiple publication bias[18]	The number of publications generated from single studies is associated with the nature and/or direction of results
Database/indexing bias[21]	Indexing of studies in particular databases is associated with the nature and/or direction of results
Place of publication bias[18]	The particular type of journal or other form of publication is associated with the nature and/or direction of results
Retrieval bias[22]	There is a difference between the average estimates based on all conducted studies compared to retrieved studies
Hot stuff bias[23]	The probability of publication for a study depends on whether the topic is 'hot'
Media attention bias[6]	The selection of studies to report by the media is associated with the nature and/or direction of results

1.2.1. Literature review methods

The terminology used to describe selective outcome reporting has not been standardised in the literature, which created difficulties in identifying appropriate keywords that were sufficiently specific and sensitive. As a result, a formal database search strategy was not used. Methodologists were contacted to request citations for relevant publications, and the reference lists of identified papers were reviewed. Furthermore, conference abstracts from the annual Cochrane Colloquia (1997-2001) were hand-searched. Finally, the 2003 Cochrane Library was searched using the terms 'selective', 'outcome reporting,' 'publication bias,' and 'bias'.

1.2.2. Study publication bias

Among the many reporting biases, publication bias between studies has received the widest recognition. It can be defined as the selective publication of entire studies based on the nature and/or direction of their results, resulting in systematic differences between published and unpublished studies. In particular, studies with statistically significant results are favoured over those with non-significant results. Both indirect and direct evidence supports the existence of selective publication of studies based on the statistical significance and/or direction of results. Indirect evidence consists of observations drawn from the characteristics of cohorts of published studies. Direct support for study publication bias is provided by comparisons between the results of published and unpublished studies.

1.2.2.1. Indirect Evidence

Indirect evidence consists of retrospective studies that observe an unreasonably large proportion of publications with 'positive' results in particular research areas (Table 1-2), as well as a negative association between sample size and the magnitude of effect estimates. With the absence of a comparison group of unpublished studies, this evidence is indirect and does not necessarily demonstrate an association between publication status and statistical significance - other factors may account for the observed associations.

<u>High proportion of 'positive' results</u>

Some of the earliest evidence of study publication bias was reported in 1959, when Sterling observed that 97% of studies in four psychology journals produced statistically significant results.[24] Since then, disproportionately high percentages of rejected null hypotheses have been found in reviews of publications in psychology (88-96%)[25-28] and biology (91%)[29], with a slightly lower percentage observed in medical journals (35% and 71-85%)[28,30-33] (Table 1-2). Although this suggests the presence of study publication bias, it does not constitute conclusive evidence since the percentage of unpublished studies with statistically significant results is unknown. There is also an issue of defining a 'positive' study when multiple outcomes are reported, which may explain some of the variation between previous reviews.

Table 1-2. Proportion of publications with significant findings in various study cohorts

Author (year)	Cohort	Definition of significant or 'positive'	% of studies with significant or 'positive' results
Hubbard and Armstrong[34] (1997)	Studies in three marketing/consumer journals	$p < 0.05$	92% (638/692)
Csada *et al.*[29] (1996)	Studies in 43 biological journals	$p < 0.05$ supporting main hypothesis	91% (1098/1201)
Mulward and Gøtzsche[30] (1996)	MEDLINE-indexed, double-blind, non-crossover randomised trials with 2 active treatment groups	$p < 0.05$	35% (136/386)
Sterling *et al.*[28] (1995)	Studies in three medical journals	$p < 0.05$	85% (270/316)
Sterling *et al.*[28] (1995)	Studies in eight psychology journals	$p < 0.05$	96% (538/563)
Moher *et al.*[31] (1994)	Trials in three general medical journals	Defined according to text or primary outcome	73% (281/383)
Moscati *et al.*[32] (1994)	Studies in emergency and general medical journals	$p < 0.05$ favouring new treatment	80% (142/177)
Davidson[33] (1986)	Trials in five general medical journals	Results favouring new treatment	71% (76/107)
Greenwald[25] (1975)	Studies in two psychology journals	$p < 0.05$	88% (175/199)
Bozarth and Roberts[26] (1972)	Studies in three psychology journals	$p < 0.05$	94% (841/895)
Smart[27] (1964)	Studies in four psychology journals	$p < 0.05$ for ≥50% of null hypotheses	91% (282/309)
Smart[27] (1964)	Studies in psychology conference and PhD abstracts	$p < 0.05$ for ≥50% of null hypotheses	75% (126/169)
Sterling[24] (1959)	Studies in four psychology journals	$p < 0.05$	97% (286/294)

If we optimistically assumed, however, that a) published studies were representative of the entire population of medical studies; b) the type I error rate was 5%; c) the power for all studies was a generous 80%; and d) all of the interventions examined were truly efficacious relative to the other study comparison groups; then one would expect that 20% of studies should fail to show any significant difference at the 5% level.[28] In fact, the power of randomised trials is often less than 80% in both general and specialty journals,[31,35-38] and interventions are not always more efficacious than controls

- meaning that many more than 20% of studies should demonstrate insufficient evidence to reject the null hypothesis. As shown consistently in numerous samples, the large proportions of studies with statistically significant results are thus highly suggestive of study publication bias.

Association between sample size and effect estimates

Another form of indirect evidence of study publication bias relies on the premise that although homogeneous studies may measure the same true treatment effect irrespective of sample size, estimates from smaller studies should vary more widely with larger confidence intervals due to chance. A funnel plot of precision versus treatment effect, therefore, should resemble a symmetric inverted funnel in the absence of bias. The presence of asymmetry may stem from study publication bias, as smaller negative studies will be absent.

A significant proportion of asymmetric funnel plots has been found among meta-analyses. Egger *et al.*[39] reviewed meta-analyses in four leading medical journals from 1993-96 as well as the 1996 Cochrane Database of Systematic Reviews (Issue 2). Using a linear regression approach, they found asymmetric funnel plots in 38% (14/37) and 13% (5/38) of meta-analyses respectively.

Using a quantitative model of the funnel plot, Berlin *et al.*[40] assessed three outcomes in a sample of 246 consecutively published cancer trials and found that small studies reported greater treatment effects compared to larger studies, with absolute differences of 17%, 41%, and 79% for each outcome. They concluded that the most likely explanation for this discrepancy was

publication bias. However, it is important to note that other sources of funnel plot asymmetry exist, including true heterogeneity, differences in quality between large and small studies, the choice of effect measure, and chance.[39,41]

1.2.2.2. Direct Evidence

Direct evidence of selective publication is provided by retrospective cohort studies that have assessed differences in publication rates between studies with statistically significant results and those with non-significant results. The cohorts have been defined as studies submitted for ethical or funding approval by specific bodies, or as studies conducted by researchers selected based on particular affiliations or publications.

Additional direct evidence consists of differences observed between the overall effect estimates derived from pooling of published and unpublished studies in meta-analyses. In contrast to indirect evidence, which relies on observations drawn from cohorts limited to published studies, these research findings directly support the existence of bias by examining the characteristics of both unpublished and published studies within defined cohorts.

<u>Studies approved by review or funding bodies</u>

Consistent evidence of study publication bias has been provided by retrospective follow-up of six cohorts of studies approved by ethics or institutional review boards, as well as by funding bodies (Table 1-3). Publication rates were determined for cohorts of studies approved by the Royal Prince Alfred Research Ethics Committee (Australia),[13] two Johns

Hopkins Health institutional review boards (United States),[42] and the Central Oxford Research Ethics Committee (United Kingdom).[14] Two cohorts of studies funded by the National Institutes of Health (United States) have also been assessed.[12,43] The cohort studies revealed that within the follow-up periods examined, projects with statistically significant results were more likely to be published than those with no significant differences between groups. A meta-analysis of five of the six cohort studies (excluding the HIV trial cohort) produced an overall adjusted odds ratio for study publication bias of 2.54 (95% CI 1.44 to 4.47).[44] Significant bias was detected among clinical trials in three of the four cohorts that reported results stratified by study design (Table 1-3).

Table 1-3. Summary of publication bias estimates and risk factors observed from follow-up of six cohorts of approved studies

Study cohort	Length of follow-up (years)	Odds ratio[a] (95% confidence interval)		Risk factors associated with increased bias
		All studies	Clinical trials	
Studies approved by Johns Hopkins institutional review board up to 1980 (Public Health)[42]	≥8	1.78 (0.94-3.39) n=172 studies	N/A	None identified
Studies approved by Johns Hopkins institutional review board up to 1980 (Medicine)[42]	≥8	3.38 (1.96-5.83) n=342 studies	N/A	None identified
Clinical trials funded by National Institutes of Health in 1979[43]	9	N/A	7.04 (1.90-26.16) n=198 trials	Single centre studies; Female principal investigator
Studies approved by Central Oxford Research Ethics Committee from 1984-1987[14]	3-6	2.96 (1.68-5.21) n=285 studies	2.10 (0.98-4.52) n=148 trials	Not randomised; No concurrent comparison group; Low importance; Sample size ≤ 20
Studies submitted to Royal Prince Alfred Hospital Ethics Committee from 1979-1988[13]	4-13	2.66 (1.32 to 5.35) n=218 studies	4.19 (1.71 to 10.32) n=130 trials	None reported
HIV trials funded by National Institute of Allergy and Infectious Diseases from 1986-96[12]	0-10	N/A	HR 3.7[b] (1.8-7.7) n=66 trials	None identified

[a]Odds ratio for publication of statistically significant versus null results

[b]Hazard ratio based on time from start of enrollment to publication

The range of follow-up periods in the six cohorts was 0-10 years (Table 1-3). Because of relatively short follow-up durations in some cohorts, it is difficult to differentiate the effects of study publication bias and time lag bias. Relative to the date of study approval or commencement of patient enrollment, trials with statistically significant outcomes are published significantly earlier than those with non-significant results (Table 1-4). A short follow-up period would thus decrease the probability that a non-significant study will have been published within the timeframes used in several of the cohort studies. It may be, however, that 'positive' and 'negative' studies would eventually be published in similar proportions given a sufficiently long follow-up. Regardless, the implications for systematic reviews are the same - at a given time point, studies with statistically significant results have a higher probability of appearing in the published literature.

Table 1-4. Time to publication stratified by statistical significance of results for studies approved by ethics, funding, and editorial bodies

Study cohort	Time interval assessed	Median time to publication (95% confidence interval)	
		$p < 0.05$	$p \geq 0.05$
Trials submitted to Royal Prince Alfred Hospital Ethics Committee from 1979-88[13]	Approval date to publication	4.69 years (3.75 to 5.72) n=76 trials	7.99 years (7.02 to infinity) n=54 trials
HIV trials funded by National Institute of Allergy and Infectious Diseases from 1986-96[12]	Start of enrollment to publication	4.3 years (not given) n=31 trials	6.5 years (not given) n=35 trials
Studies on effects of passive smoking identified from survey of investigators funded by 89 agencies[45]	Funding start date to date of publication	3 years (3-5) n=33 studies	5 years (4-7) n=21 studies
Trials accepted for publication by *JAMA*[a] from 1996-99[46]	Submission date to publication	7.8 months (not given) n=78 trials	7.6 months (not given) n=51 trials

[a]Journal of the American Medical Association

Some of the follow-up studies also investigated risk factors for study publication bias (Table 1-3).[14,42,43] These factors consist of characteristics that

interact with statistical significance to affect the publication status of a study. Subgroups with significantly higher magnitudes of bias were identified in two cohorts. Single centre studies and the presence of a female principal investigator were found to be associated with higher degrees of publication bias in the National Institutes of Health trial cohort.[43] Observational or laboratory-based studies, as well as non-randomised trials, were identified as statistically significant risk factors in the Oxford ethics committee cohort.[14]

Studies identified by researcher affiliations and publications

Surveys of study authors and members of professional organisations have also provided some direct evidence of publication bias (Table 1-5). Researchers were asked about studies they had conducted, along with the nature of the results and the publication status. Publication and submission rates were higher for studies with statistically significant results compared to those with non-significant findings. However, the reliability of self-reported survey data is unknown, and response rates in the surveys varied from 49%-79%. In addition, one study used a non-random sampling method to select individuals to be surveyed, thus limiting its generalisability.[47]

Effect of incorporating unpublished studies into meta-analyses

Additional direct evidence of study publication bias is provided by assessing the impact of unpublished data on meta-analyses (Table 1-6). Reduced estimates of effect have been observed when registered or unpublished studies are included compared to only published studies, demonstrating that unpublished trials tend to have results that are closer to the null.

Table 1-5. Surveys of researchers to assess publication or submissions rates in four study cohorts

Author (year)	Survey population	Response rate	Definition of 'positive'	Crude publication/ submission rate	
				'Positive' studies	'Negative' studies
Dickersin *et al.*[47] (1987)	Authors of published trials	49% (156/318)	New treatment is superior	94% (423/449) published	69% (344/496) published
Sommer[48] (1987)	Society for Menstrual Cycle Research members	65% (91/140)	Clear effect in predicted direction	73% (22/30) published	54% (14/26) published
Coursol and Wagner[49] (1986)	American Psychology Association members	61% (609/1000)	Client improved	66% (85/129) published	22% (14/65) published
Greenwald[25] (1975)	Authors and reviewers of a psychology journal	79% (75/95)	$p < 0.05$	49%[a]	6%[a]

[a]Mean rating by researchers of the probability that they would submit their study for publication

Table 1-6. Effect of including registered or unpublished trials on pooled estimates of effect in meta-analyses

Author (year)	Intervention	Pooled estimate (95% confidence interval)	
		Published trial data only	Including registered or unpublished trial data
Simes[50] (1986)	Initial alkylating agent versus combination chemotherapy in ovarian cancer	Median survival ratio 1.16 (1.06-1.27) n=16 trials	Median survival ratio 1.05 (0.98-1.12) n=16 registered trials
Simes[50] (1986)	Initial alkylating agent plus prednisone versus combination chemotherapy in multiple myeloma	Median survival ratio 1.26 (0.93-1.70) n=6 trials	Median survival ratio 1.11 (0.96-1.29) n=11 registered trials
Stewart and Parmar[51] (1993)	Single non-platinum drugs versus cisplatin-based combination chemotherapy in ovarian cancer	Odds ratio for survival 0.70 (0.51-0.94) n=7 trials	Odds ratio for survival 0.75 (0.59-0.96) n=11 published and unpublished trials
MacLean CH *et al.*[52] (2003)	Non-steroidal anti-inflammatory drugs versus placebo	Odds ratio for dyspepsia 1.21 (0.81-1.81) n=15 trials	Odds ratio for dyspepsia 1.14 (0.86-1.53) n=26 published and unpublished trials

Using two cohorts of trials comparing combination chemotherapy to either alkylating agent in ovarian cancer or alkylating agent plus prednisone in multiple myeloma, Simes[50] compared meta-analysis results after incorporating

trials registered with the International Cancer Research Data Bank and the Compilation of Experimental Cancer Therapy Protocol Summaries. Although a survival advantage was found for combination chemotherapy in ovarian cancer when results from published trials were pooled, this outcome was no longer statistically significant when all registered trials were combined (Table 1-6). In the multiple myeloma review, a reduced survival advantage was also found from pooling of registered trial results compared to published trial results.

In a third review, individual patient data meta-analysis of cisplatin-based therapy in ovarian cancer demonstrated that the inclusion of data from unpublished trials reduced the magnitude of the pooled estimate of treatment effect (Table 1-6).[51]

Finally, a recent review found that the inclusion of unpublished trial reports submitted to the United States Food and Drug Administration had little effect on the odds ratio for dyspepsia caused by non-steroidal anti-inflammatory drugs (Table 1-6).[19] Similar quality scores were noted for the published and unpublished trials.

1.2.2.3. Responsibility for selective publication

It can be hypothesized that a strong driving force behind the selection process is the desire for 'positive' results. The definition of a 'positive' or 'interesting' result, however, is multi-factorial and context-dependent. Contributing factors include its statistical significance, the direction of effect in terms of favouring the new or sponsored product, and the novelty of the study findings. Among

these factors, statistical significance has undergone the most scrutiny, as it is both objectively defined and amenable to quantitative review.

Previous studies have demonstrated that the failure to publish 'positive' studies - particularly ones with non-significant results - may stem from the decisions of all those involved in the publication process: researchers, peer reviewers, journal editors, and sponsors.

Why do researchers fail to publish?

Surveys of researchers have shown that the primary reasons for studies remaining unpublished are a reflection of investigator attitudes and/or failure to submit manuscripts for publication as opposed to rejection by journals (Table 1-7). Common reasons for failing to publish studies include a lack of time or priority, methodological or logistical problems, anticipated rejection, and negative or uninteresting results. Only a small percentage of studies were found to be unpublished due to rejection by journals.

However, it is difficult to absolve editors and peer reviewers of substantial responsibility, as investigator attitudes are guided to a large extent by the perceptions and actions of journals. Authors of 'negative' studies may be reluctant to prepare and submit manuscripts because of anticipated rejection.[53] Among sixty-eight responders to a questionnaire sent to eighty authors of psychology studies, over 60% felt that non-significant results had little chance of being published.[54]

Table 1-7. Reasons provided in surveys of researchers for not publishing studies

Author (year)	Survey population	Response rate for providing reasons	Proportion of responses giving reasons for not submitting or not publishing studies[a]	
			Null results	Other
Timmer *et al.*[55] (2002)	Gastroenterology conference abstracts	39% (206/525)	7% (14/206 studies)	Lack of time (60%); design limitations (12%); unimportant results (9%); anticipated rejection (6%); publication not an aim (6%)
Misakian and Bero[45] (1998)	Investigators of studies on passive smoking funded by 89 agencies	Unclear	3.4% (2/59)	Lack of time (63%); ongoing study/analysis (56%); manuscript rejected (7%)
Weber *et al.*[53] (1998)	Authors of abstracts submitted to emergency medicine conference	81% (179/222)	4% (7/179)	Lack of time (41%); anticipated rejection (20%); unimportant results (12%); trouble with co-authors (9%); not worth the trouble (7%); similar findings to other papers (6%)
Rotton *et al.*[56] (1995)	Authors of articles in 75 psychology journals	63% (468/740)	60% (numbers not given)	Unfavourable reviews (33%); inexplicable results (22%); failure to replicate (5%); non-hypothesized results (5%)
Scherer *et al.*[57] (1994)	Ophthalmology conference abstracts	100% (32/32)	0	Lack of time (28%); manuscript rejected (19%); ongoing study (16%); study design problem (9%)
Dickersin and Min[43] (1993)	Clinical trials funded by National Institute of Health in 1979	100% (14/14)	0	Not interesting results/ no time (43%); Co-investigator or operational problems (38%); Analysis not complete (14%)
De Bellefeuille *et al.*[58] (1992)	Abstracts from 1984 meeting of American Society of Clinical Oncology	50% (41/82)	0	Lack of time/resources (32%); Low priority (22%); ongoing study (12%); manuscript rejected (10%)
Dickersin *et al.*[42] (1992)	Studies approved by Johns Hopkins institutional review board (Medicine) up to 1980	100% (65/65)	0	Results not interesting (40%); design/operational problems (26%); publication not the aim (12%); manuscript rejected (3%)
Dickersin *et al.*[42] (1992)	Studies approved by Johns Hopkins institutional review board (Public Health) up to 1980	100% (59/59)	0	Design/operational problems (39%); results not interesting (19%); publication not the aim (14%); manuscript rejected (7%)
Easterbrook *et al.*[14] (1991)	Studies approved by Central Oxford Research Ethics Committee from 1984-1987	53% (78/147)	15% (26/175 reasons)	Submitted/published elsewhere (19%); Design/logistic problems (12%); Sponsor controls data (11%); ongoing analysis (11%); manuscript rejected (9%); publication not the aim (7%); lack of time/interest (6%); unimportant results (6%)
Dickersin *et al.*[47] (1987)	Authors of published trials	82% (167/204)	35% (58/167 completed trials)	To be submitted/in progress (15%); lack of interest (14%); Poor recruitment (14%); Poor methodology (5%); Side effects (8%); Funding problem (6%); Controversy (3%)

[a]Reasons provided are not mutually exclusive

Table 1-8. Summary of studies examining peer review bias using fictional manuscripts

Author (year)	Fake manuscript used	Reviewers	Conclusions
Ernst and Resch[59] (1999)	2 versions of *in vitro* study using either mainstream or unconventional drug	Convenience sample of 291 medical doctors from list of conference participants	• No significant difference in quality ratings • Low inter-rater reliability
Abbot and Ernst[60] (1998)	4 versions of complementary medicine study: poor vs. high quality, and positive vs. negative result	200 authors of articles on MEDLINE	• No evidence of bias based on results • Poor quality version more likely to be rejected compared to high quality (55% vs. 16%)
Resch *et al.*[61] (1997)	2 versions of short report on obesity treatment using orthodox or homeopathic remedy	398 experts identified on MEDLINE	• Orthodox therapy received significantly higher quality and acceptance ratings compared to unconventional treatment
Ernst and Resch[62] (1994)	Study of transcutaneous nerve stimulation (TENS)	33 referees classified as pro-TENS or against TENS	• Poor inter-rater reliability • Judgement associated with referees' prior beliefs and preconceptions
Nylenna *et al.*[63] (1994)	2 short papers in either English or national language, both with methodological flaws	180 Scandinavian referees received both versions	• English version given higher quality rating than national-language version
Epstein[64] (1990)	2 versions of study with positive or negative results	Submitted to 146 social work journals	• No evidence of bias: 25% (4/16) of journals accepted the negative version vs. 35% (6/17) for the positive version
Mahoney[65] (1977)	4 versions of psychology study with positive, negative, mixed, or no results	75 blinded referees chosen such that they would support a positive study	• Negative results more likely to receive lower rating scores and more often rejected with major revisions recommended • Two manuscript errors were noticed by a significantly higher % of reviewers for negative studies (71% vs. 25%)

Subjectivity in peer review

Bias on the part of journal reviewers has been investigated in several experimental studies using fictional manuscripts (Table 1-8). Inter-rater reliability was shown to be poor, and versions of psychology manuscripts with positive results were favoured over those with negative results.[65] No evidence of such bias was observed in a study using fictitious social work manuscripts.[64]

Versions of reports in English[63] as well as those describing results that supported the reviewer's beliefs[62,65] were also favoured. Evidence of reviewer bias against unconventional therapies has been mixed.[59-61] One study selected thirteen psychology articles published by renowned researchers from prominent institutions, and re-submitted them to the same journals using fictional authors from unknown institutions. Ten of the thirteen studies were rejected, suggesting that peer review is unreliable and may possibly be influenced by institutional or author prestige.[66]

Anecdotal evidence has also been published describing correspondence from journal editors rejecting manuscripts on the basis of negative results.[28] In addition, some journal instructions to authors have implied a preference for positive outcomes.[67-70]

Surveys provide further evidence of bias on the part of editors. A survey of 1,000 psychologists with a response rate of 61% revealed a preference for 'positive' results at both the editorial and trialist levels.[49] 301 responders to another survey of 429 editors and advisory board members from nineteen social science/management journals revealed that several factors decreased the likelihood of acceptance for publication - including non-significant results, replications of previous studies, lack of original data, and previous presentation at meetings.[71] Factors found to increase the chance of publication included a strong author reputation, successful testing of the author's own hypothesis, and material differing from the journal's traditional content.

Finally, little evidence of editorial bias was found among studies submitted to two journals for publication. A retrospective study examined the association between institutional prestige (based on funding levels) and peer reviewer recommendations or editorial decisions for 147 brief reports and 258 major papers submitted to the *Journal of Pediatrics*.[72] The authors found that lower acceptance and recommendation rates were associated with institutional rank only for brief reports, but not for major papers. Evidence of bias at the editorial level based on statistical significance was not found in a recent prospective study of 745 manuscript submissions to the *Journal of the American Medical Association*.[73] The generalisability of these findings to other medical journals is unclear.

Conflict of interest with commercial sponsors

In addition to biased practices on the part of researchers, editors, and peer reviewers, the publication of study results may be influenced by sponsors. Industry-sponsored studies frequently remain unpublished. Despite similar quality scores between published and unpublished studies, 38% of pharmaceutical company trials submitted to licensing authorities in Sweden and Finland were unpublished.[74] In a separate cohort, drug trials registered with the Finnish National Agency for Medicines were classified according to whether they submitted final study reports.[75] Reporting rates for studies with positive results (38%, 42/111) were greater than for inconclusive (18%, 6/33) or negative (20%, 9/44) results. Another review of submissions to a Swedish drug regulatory agency found that industry-sponsored trials with significant primary outcomes were more likely to be published as stand-alone reports compared to non-significant trials (19/21 versus 6/21).[19]

Two recent systematic reviews found that published industry-sponsored trials were significantly more likely to favour the sponsored drug compared to trials without industry funding (odds ratios of 3.60 and 4.05).[76,77] Finally, a review of all fifty-six MEDLINE-indexed non-steroidal anti-inflammatory drug trials published from 1987-1990 provided indirect evidence of suppression of statistically non-significant studies. Comparison drugs were not found to be superior to sponsor drugs in any of the studies.[78] In addition, nineteen of the twenty-two trials that reported a difference in toxicity profiles favoured the sponsor drug.

Contractual agreements between academic institutions and commercial agencies often enable the sponsor to delay publication and/or edit portions of a manuscript to protect its commercial interests.[79] A survey of American academic centres revealed that trial agreements rarely ensured the investigators' participation in trial design, their right to access data from all study centres, and the freedom to publish results without editorial control by the sponsor.[80]

In some instances, pharmaceutical companies have pursued or threatened legal action in order to prevent publication of results that were unfavourable to their commercial interests.[81] A manufacturer-sponsored study of levothyroxine found similar efficacy between the generic and the sponsor's brands, prompting the company to take legal action to delay publication of the findings.[82] In a different case, a researcher conducting an industry-sponsored clinical trial to investigate a new therapy (deferiprone) for thalassemia in children published her concerns about the drug's safety, and was

subsequently sued by the manufacturer and dismissed by her hospital.[83] Other examples of industry interference in the publication of results include systematic reviews of statins[84] and hormone bovine somatotropin.[85]

It is therefore clear that the nature and inter-play of factors involved in ultimately determining a study's publication status are complex. The evidence implicating a preference for studies with statistically significant results is substantial, and has the potential to introduce significant bias in the results of systematic reviews. However, it is possible that the published evidence supporting the existence of study publication bias is itself a biased sample.

1.2.3. Outcome reporting bias

The superior value placed on statistically significant results relative to non-significant findings at the between-study level can be extended to the reporting of outcomes within published studies. Supported primarily by anecdotal evidence, the existence of selective reporting of measured outcomes has been widely suspected for years.[21,86-92] However, relative to study publication bias, limited empirical investigation has been conducted on this particular phenomenon.

Outcome reporting bias (also known as outcome variable selection bias or within-study reporting bias) refers to the selective reporting of measured outcomes within a published study based on their nature and/or direction, resulting in systematic differences between reported and unreported data. For example, a published trial measuring multiple outcomes may report only those which are found to be statistically significant, or a publication measuring an

outcome collected at various points in time may selectively report the time point corresponding to the maximum observed difference between groups.[93]

The completeness of reporting for a particular outcome in a trial publication can range from the omission of the outcome to the provision of sufficient data for meta-analysis. It is important to distinguish between two levels of incomplete reporting - outcomes that are omitted from publications and those that are reported with insufficient data for meta-analysis. The latter may have little impact when interpreting a single trial report, but can have a significant impact on systematic reviews. Outcome reporting bias arises if the occurrence of either level of incomplete reporting is associated with characteristics such as statistical significance.

Trials in which primary outcomes are not defined *a priori* have the potential for selective reporting after data analysis.[94,95] Unfortunately, reviews have found that 25%-94% of trial reports do not specify a primary outcome.[92,94,96,97] A pilot study reviewing fifteen ethics protocol submissions and their subsequent publications observed that two of six studies that had specified primary outcomes in their protocol did not define the same ones in their publication.[98]

As with study publication bias, there is both indirect and direct evidence for outcome reporting bias. Indirect evidence consists of discrepancies observed within and between published studies that demonstrate the existence of unreported outcomes. Such observations do not constitute direct evidence of bias because the nature of the unreported outcomes is unknown, and

therefore biased selection cannot be confirmed. Direct evidence consists of case examples of meta-analyses in which the inclusion of unreported outcomes tended to reduce the magnitude of pooled treatment effect estimates.

1.2.3.1. Indirect evidence

Within-study discrepancies

Indirect evidence of outcome reporting bias includes discrepancies within trials that suggest the existence of a number of measured but unreported outcomes. Tannock[94] reviewed thirty-two published reports of oncology trials and observed a large number of unreported but implied statistical comparisons that often outnumbered the reported comparisons (median reported outcomes, 32; reported and implied outcomes, 86). A review of 196 non-steroidal anti-inflammatory drug trials for rheumatoid arthritis[97] revealed that 49% did not report the same set of outcomes in the Methods and Results sections. 30% of the trials mentioned one or more outcomes only in the Methods but not the Results sections. Furthermore, among thirty-one separate trials that were published more than once, the total number of reported outcomes varied between the multiple publications of five separate trials.[99] Finally, an assessment of trial reports submitted to licensing bodies by pharmaceutical companies in Finland and Sweden found that adverse effects were reported more frequently in unpublished than published trials despite similar study quality scores.[74]

In the lone empirical study specifically examining the nature of outcome reporting across a broad sample of studies, a follow-up pilot project reviewed

submissions to an ethics committee in order to examine the consistency between outcomes specified in protocols and those reported in subsequent publications.[98] The use of protocols provided an objective outline of outcomes specified *a priori*. Fifty-six principal investigators who submitted ethics applications were contacted to request permission to review their protocols. Thirty-seven of forty responders gave consent, and eighteen of these responders had published their study findings. Publications or citations could not be obtained from three of these eighteen researchers. One particular difficulty observed by the authors was that protocols were often not specific in describing outcome measures and their analyses, thus enabling a wide range of practices to occur during the trial and decreasing the potential to make meaningful comparisons with publications.

In their final sample of fifteen protocols with publications, eight studies (including two randomised trials) had discrepancies between the outcomes specified in their protocol and those reported in the publications. All eight reports introduced outcomes in the publication that had not been specified in the original protocol, most of which were statistically significant. Six reports also omitted outcomes that had been specified in the protocols.

One possible explanation is that the discrepancies reflect a tendency to report interesting, statistically significant results over non-significant findings. However, without knowing the reasons behind the inconsistencies in outcome reporting or the characteristics of the unreported outcomes, it is not possible to determine the exact nature or the magnitude of any potential bias. Alternative explanations include legitimate changes to study protocols that were made

after ethical approval, or the addition and removal of outcomes occurring independently from results - such changes would not necessarily produce bias.

A major weakness in the methodology of this pilot study was the requirement for consent from each researcher in order to examine the study protocol. If biased reporting of outcomes exists, then trialists who selectively suppress non-significant outcomes would have a clear interest in not consenting to have their protocols scrutinised. It is thus probable that those who provided consent would be systematically different from those who did not, thus undermining the scientific validity of the study and likely underestimating the extent of misreporting. Conclusions from such a biased sample may not be reliable.

Between-study discrepancies

Discrepancies between trials included in systematic reviews also provide indirect evidence of selective outcome reporting.[100] In a systematic review of anthelmintic drugs for treating worms in children,[101] it was noted that some trialists conducted several analyses and reported those with statistically significant results. For example, one study described three outcomes in the Methods section but then reported only one, stating that that "selective data [are] presented so as to reduce the number of tables."[102] Sensitivity analyses using a selection model revealed that the conclusions of the systematic review could easily be changed depending on assumptions made about the unreported outcomes.[101]

The authors of another systematic review of antidepressants in schizophrenia and depression observed that none of the trials defined a primary outcome,

and that few trials reported binary outcomes.[103] The reviewers commented that this raised the possibility of selective reporting.

Tinnion and Hanlon[104] found that data for four types of adverse events were collected but unreported in several trials of acellular vaccines for whooping cough in children. Some of the reports stated qualitatively that data for these results were omitted due to a lack of difference between treatment groups. Studies that presented complete data for these outcomes generally reported a significant inter-group difference. The reviewers also noted a tendency for some trialists to selectively report adverse event data at the time point when a maximum difference was observed.

In a systematic review of massage for promoting growth and development in infants, Vickers *et al.*[105] expressed concern about selective reporting when a clinically important outcome (length of hospital stay) was reported by only one of fourteen studies. Raw data obtained for one of the non-reporting studies revealed that the outcome had been measured but was found to be statistically non-significant. The single study that reported data for length of stay showed a significant difference between groups, but failed to present data for another outcome with non-significant results.

Vickers and Smith[106] conducted a review of homeopathic Oscillococcinum for influenza. They suggested that the lone study which reported individual symptom scores may have selected them from a larger set of scores based on each of their *p*-values being less than 0.05; the total number of symptoms did not differ significantly between groups.

Finally, selective reporting of subgroup analyses was suspected for two of five trials in a systematic review of malaria chemoprophylaxis in pregnancy.[20] The two trials reported overall results for outcomes without providing subgroup results stratified by gravidity. Although the review concluded that treatment efficacy was limited to primigravidae, sensitivity analyses with imputed data revealed that this conclusion may have been unreliable and that the pooled estimate of treatment effect may have been over-estimated due to the exclusion of the two trials.

These case examples suggest the possible occurrence of selective outcome reporting in a variety of settings. They are based, however, on unverified assumptions about the nature of the unreported outcomes, and their generalisability to other studies is limited. Furthermore, they provide no indication of the prevalence of outcome reporting bias across journals and specialty fields.

1.2.3.2. Direct evidence

Direct evidence of outcome reporting bias is sparse and limited to case examples in which statistically significant outcomes were reported preferentially over non-significant outcomes. The majority of cases have been noted when systematic reviewers were able to obtain individual patient data or unreported data, allowing for comparisons of results with and without the inclusion of unreported outcome data.

In a systematic review of artemisinin therapy for malaria, it was observed that non-significant subgroup results reported in the unpublished manuscript of one

of the trials were excluded from its final journal publication.[20] In a meta-analysis of second-line drugs for rheumatoid arthritis,[107] the reviewers noted that some published trials measured up to ten efficacy outcomes but reported numerically only those that were statistically significant; the non-significant outcomes were either omitted or described qualitatively as showing no significant difference between groups.[21] These examples demonstrate the tendency to fully report outcomes that show significant differences between comparison groups over those that do not.

An example of the potential impact of selective outcome reporting comes from two published analyses of the same trial investigating amoxicillin in otitis media.[108,109] The conclusions differed for the two analyses because of variations in the importance given to some of the measured outcomes. It was revealed that the primary investigators in the report favouring amoxicillin had received considerable financial support from the drug manufacturers.[110]

Table 1-9. Effect of incorporating unreported outcomes in meta-analyses of randomised trials

Author (year)	Outcome	Pooled effect estimate (95% CI)			
		n	Reported outcomes only	n	Reported and unreported outcomes
West RR & DA Jones[111] (1997)	Total mortality after myocardial infarction	8	Relative risk 0.65 (0.46-0.91)	11	Relative risk 0.73 (0.53-1.00)
McCormack K *et al.*[112] (2001)	Persisting groin pain after hernia surgery (mesh versus non-mesh methods)	3	Odds ratio 5.31 (1.37-20.51)	13	Odds ratio 0.60 (0.42-0.84)
McCormack K *et al.*[112] (2001)	Persisting groin pain after hernia surgery (laparoscopic versus open mesh methods)	2	Odds ratio 2.28 (0.58-8.92)	7	Odds ratio 0.65 (0.52-0.81)

Additional direct evidence of outcome reporting bias illustrates the potential impact of unreported outcomes on meta-analyses (Table 1-9). A systematic review comparing psychological rehabilitation with usual care after myocardial infarction found a statistically significant result when the eight published trials reporting total mortality were pooled (relative risk 0.65; 95% CI 0.46 to 0.91).[111] However, this result lost its statistical significance when unreported mortality data from the three remaining trials were included (relative risk 0.73; 95% CI 0.53 to 1.00). A subsequent large trial observed no difference in mortality (relative risk 1.01; 95% CI 0.75 to 1.37).[111]

A final example involves a systematic review of hernia surgery trials[112] in which individual patient data were obtained for several published trials. Outcomes that were measured but not reported in the trial publications were identified from the individual patient data. For the outcome of persisting groin pain, the inclusion of unreported data increased the number of contributing trials from three to thirteen for the comparison of mesh versus non-mesh methods, and two to seven for the comparison of laparoscopic versus open mesh methods. Incorporating these additional data into the meta-analyses dramatically changed the direction and statistical conclusions of the results.

These examples illustrate the tendency to preferentially report statistically significant outcomes in trials identified for systematic reviews, while results that are closer to the null tend to be suppressed. The potential impact on conclusions drawn from systematic reviews is substantial. However, as with most of the indirect evidence for outcome reporting bias, the case reports

constitute a relatively weak and unreliable form of evidence with limited generalisability to the overall medical literature.

It should be noted that in addition to the selective reporting of outcomes, researchers can selectively report analyses based on different populations of participants, such as particular pair-wise comparisons within multi-arm studies, subgroup analyses, and intention-to-treat or per-protocol analyses.[19,20] For example, a recent review of submissions to a Swedish drug regulatory agency for approval of selective serotonin reuptake inhibitors found that pooled effect estimates based on all forty-two submitted studies were generally smaller than those based on thirty-eight published studies.[19] These differences were attributed to the selective reporting of the more favourable per-protocol analyses in publications. Both intention-to-treat and per-protocol analyses were reported in most trials submitted to the regulatory agency, but only a quarter of publications reported the former analyses. Although this type of selective reporting is important, it is not specifically included in our evaluation of outcome reporting bias.

1.2.4. Summary

Selective reporting of outcomes within published studies has the potential to bias the results of systematic reviews, particularly if important outcomes are not specified *a priori*. The impact of favouring statistically significant results within published studies would be in addition to that of publication bias between entire studies, resulting in spuriously inflated effect estimates in meta-analyses with potentially negative consequences for patients and researchers. Although outcome reporting bias has long been suspected by researchers,

little direct evidence exists other than case reports, which may themselves be subject to publication bias.

1.3. Thesis rationale

The evidence available on the biased reporting of research findings, both between- and within-studies, highlights a number of crucial issues relating to the reliability of the medical literature. Although study publication bias has been studied extensively and recognised as a significant threat to systematic reviews, the question remains as to whether a similar phenomenon occurs with regards to the selective reporting of outcomes within published studies, and if so, the extent and impact of such bias.

Evidence for such a phenomenon is lacking - no previous investigation has directly assessed the nature and prevalence of selective outcome reporting across a broad population of studies. Evidence from a representative sample of studies is required to establish whether the biased outcome reporting demonstrated in individual examples can be generalised to medical research overall. In addition, the potential for bias through *post hoc* changes to primary outcomes requires further examination to determine their prevalence and impact. Because randomised trials rank highly in the hierarchy of medical evidence and form the foundation for reliable systematic reviews of healthcare interventions, it seems logical that initial investigations focus on this particular study design. Firm evidence of reporting biases among trials could help to ensure that preventative measures will be adopted to improve the reliability of trial reports and systematic reviews.

This thesis presents the results of two approaches to examine outcome reporting bias in randomised controlled trials: a review of published trials indexed on PubMed supplemented by a survey of the trialists; and a comparison of protocols with subsequent publications. The contrasting approaches - the former looking backward from the publication stage, and the latter looking forward from the protocol stage - offer complementary evidence for selective outcome reporting in trials and the potential for bias in systematic reviews. Both studies aimed to establish the prevalence of poor outcome reporting, as well as the magnitude of association with statistical significance. The access to trial protocols also enabled a comparison of outcomes specified *a priori* to those reported in publications.

Chapter 2 - Methods used to examine outcome reporting practices in randomised trials

2.1. Introduction and objectives

Chapter 1 reviewed the background literature and presented case examples that illustrate the potential for selective outcome reporting in medical research. It is apparent, however, that empirical evidence of this form of bias is lacking. This chapter will outline the objectives and describe the methodologies used in this thesis.

The general objectives of this thesis were to examine empirically the occurrence of inadequate outcome reporting in publications of randomised trials, and to determine whether this poor reporting is associated with bias. Specifically, this thesis has four primary aims:

- To assess the prevalence of incomplete outcome reporting in journal publications of randomised trials;
- To determine if the selection of outcomes to fully report in publications is associated with their statistical significance;
- To determine the magnitude of this bias across various trial characteristics; and
- To evaluate the consistency between outcomes specified in trial protocols and those reported in publications.

These objectives will be achieved through two separate approaches applied to three trial cohorts. The first approach identified a sample of trials from publications indexed on PubMed, and obtained supplemental information on unreported outcomes from trialists by means of a survey. The second approach

began with two cohorts of trial protocols, which provided information on pre-specified outcomes, followed by a survey of authors and review of relevant journal publications.

2.2. Trial cohorts

Three retrospective cohorts of published randomised trials were identified: a) Trials whose primary report was published in December 2000 and indexed on PubMed (PubMed trials); b) Trials approved from 1990-98 for funding by the Canadian Institutes of Health Research (CIHR trials), the national medical research agency of Canada; and c) Trials approved from 1994-95 by the Copenhagen and Frederiksberg Research Ethics Committee (Ethics trials), the largest ethics approval body for clinical research in Denmark.

2.2.1. Definition of study samples

2.2.1.1. PubMed cohort

The PubMed cohort was defined as randomised trials whose primary reports were published in December 2000 and indexed in PubMed. PubMed is a free search and retrieval system that provides access primarily to MEDLINE, as well as some additional life sciences journals. MEDLINE encompasses studies from over 4,600 journals published in 71 countries.[113]

From a practical perspective, the month of December 2000 was selected to define a relatively recent cohort while allowing adequate time for trials to be indexed on PubMed. An up-to-date cohort was desirable because contact details for surveying authors would be more accurate than information from older

publications. Also, recent data would be more readily accessible to trialists, which would presumably increase the response rate to surveys. Related to this desire for recent trials, PubMed was chosen over the "best single source of published trials for inclusion in systematic reviews"[10] - The Cochrane Controlled Trials Register - because the register lags behind PubMed in indexing trials due to its reliance on other electronic databases (including MEDLINE) as well as hand-searching of journals.

2.2.1.2. Ethics cohort

The Ethics sample consisted of published randomised trials that received approval from the Copenhagen and Frederiksberg Research Ethics Committee (Denmark) from 1994-95. Committee approval is required for any clinical research conducted in the Copenhagen and Frederiksberg regions. The timeframe was chosen to define the most recent sample that would enable a high publication rate. This was based on previous observations that trials approved by an ethics committee in Australia required a median time from ethical approval to publication of up to 8 years.[13]

2.2.1.3. CIHR cohort

The CIHR cohort was defined as published randomised trials that were funded through grant competitions held from 1990-1998 by a government research funding agency - formerly the Medical Research Council of Canada - known now as the Canadian Institutes of Health Research (CIHR).

Funded trials were selected through a highly competitive application process. They were expected to be of high methodological quality with the ability to

provide definitive answers to questions of clinical importance. Full operating grants were available, as well as partnership grants for co-sponsorship with commercial or private sources.

The widest possible range of years was chosen to maximise the number of trials available. 1990 was used as the earliest inclusion year because files prior to this time could not be retrieved from archives. The final cut-off year of 1998 was imposed to allow sufficient time for trials to be published by 2003.

2.2.2. Inclusion criteria

2.2.2.1. For trials

A relatively broad definition of randomised controlled trials was used for all three cohorts. They were defined as prospective studies assessing the therapeutic, preventative, adverse, or physiological effects of one or more healthcare interventions in either human subjects or body parts that were to be replaced in humans (eg. transplant organs), and allocating participants to study groups using a random method. Studies were included if they simply claimed to allocate randomly, or if they described a truly random sequence of allocation. Pseudo-random methods of allocation, such as alternation or the use of date/case numbers, were deemed inadequate for inclusion. Trials must also have included at least one statistical comparison between randomised groups.

For the PubMed cohort, trials were included only if their primary report was published in December 2000. The primary report of a completed trial was defined as the first full-length publication of main outcomes. If more than one trial

was reported in a single publication, only the first eligible study appearing in the report was used to avoid any clustering effect.

2.2.2.2. For publications

Trial reports were included if they were full-length journal publications that reported final outcome comparisons between randomised groups. A language restriction to English and French reports was imposed based on the investigator's linguistic abilities. Relevant sections of Danish publications for the Ethics cohort were translated by a member of the Nordic Cochrane Centre.

2.2.3. Exclusion criteria

2.2.3.1. For trials

Economic studies, with cost outcomes as their primary interest, were excluded in order to focus on clinical outcomes. The methodology and reporting of such outcomes for meta-analysis has not been fully established. *Studies of diagnostic test properties* were excluded for similar reasons. Trials assessing the effect of a particular diagnostic test on clinical outcomes were included.

2.2.3.2. For publications

Reports of interim analyses, defined as publications reporting preliminary analyses that were subject to change upon study completion, were excluded in order to focus on final outcomes. *Abstracts* and *conference proceedings* were also excluded.

2.2.4. Process of sample selection

The three trial cohorts were established primarily though literature searches and surveys of authors (Table 2-1). Index trials were initially identified to define each cohort. For the PubMed sample, index trials consisted of studies with a PubMed-indexed primary report published in December 2000. The Ethics and CIHR cohorts were defined by protocols of published index trials that were approved in particular years. A large amount of time and effort were required before gaining access to confidential trial protocols for the Ethics and CIHR cohorts; efforts in the United Kingdom were eventually abandoned. Ethical and legal obstacles were central among the difficulties encountered over an 14 month period - the lengthy process is outlined briefly in Chapter 3.

Table 2-1. Methods used to identify index trials and subsequent publications for the study cohorts

Cohort	Identification of index trials	Identification of publications
PubMed	PubMed search	• Outcomes Survey • Literature search
Ethics	Manual search	• Publications Survey • Literature search
CIHR	CIHR internal database search	• Publications Survey • Literature search

2.2.4.1. PubMed cohort

PubMed-indexed randomised controlled trial reports published in December 2000 were identified using the optimal search strategy developed for the Cochrane Collaboration.[114] Phase 1 of the Cochrane strategy was able to identify 98% of all controlled trials in a gold standard database.[114]

The search was first conducted in March 2001, and repeated every few months to identify newly-indexed trials. The initial search strategy was based on Phase 1

of the original Cochrane strategy published in 1994 (Appendix 1).[115] In October 2001, two new terms - 'cross-over studies[mh]' and 'multicenter study[pt]' - were added to our search strategy for increased sensitivity based on preliminary recommendations from a poster presentation at the 9th International Cochrane Colloquium.[116] A final modification occurred in June 2002, when the term 'controlled clinical trial [pt]' was added based on the publication of a revised Cochrane strategy.[114] The final search strategy used is shown in Appendix 2. The final search was conducted on July 17, 2002.

Studies identified in the PubMed search were screened in two phases. Abstracts were initially reviewed to exclude any obvious non-trials such as reviews or observational/laboratory studies, as well as articles in languages other than English or French. If any uncertainty existed, or if the abstract was not available from PubMed, the study was retained. Full-text articles were then obtained for remaining studies. The cohort of randomised trials was determined after inspection of the full text of all trial reports.

Once the cohort was defined by the index primary trial reports published in December 2000, subsequent publications after the year 2000 were identified from the Outcomes Survey responses and literature searches. This ensured that all publications were reviewed for data extraction from each trial in order to reflect the conduct of systematic reviews. The Outcomes Survey (to be detailed in Section 2.3.4) solicited information about unreported outcomes, and additional trial reports were identified when trialists stated that outcomes were reported in other publications. For Outcomes Survey non-responders, as well as responders stating the existence of unreported outcomes, literature searches of PubMed,

EMBASE, the Cochrane Controlled Trials Register, and PsychINFO were conducted to identify further publications. Investigator names (first, second, and last authors, as well as contact authors) and study keywords were used as search terms. The final literature search for these secondary publications was conducted in January 2003.

2.2.4.2. Ethics cohort

This study required collaboration with local investigators at the Nordic Cochrane Centre in Copenhagen in order to obtain access to ethics committee files. A confidentiality agreement was signed. Clinical studies approved by the ethics committee were identified manually by retrieving boxes archived by year of study approval. A list of approved studies for the relevant years was not available. However, the archived boxes and files were consecutively numbered, thus enabling the identification of missing files. The original approved protocols and application forms were reviewed to identify randomised trials. Relevant sections of protocols written in Danish were extracted and translated by a researcher from the Nordic Cochrane Centre.

Once the initial study cohort was defined by the index protocols, publications reporting final results for each trial were identified through an electronic mail (e-mail) survey of the principal investigator as well as literature searches. This survey will be referred to as the Publications Survey to distinguish it from the Outcomes Survey described later. Publications Survey questionnaires asked trialists to list any publications, and requested information about funding sources (Appendix 3). If published, they were also asked whether any unreported outcomes existed. If unpublished, the survey asked for the status of the study as

well as reasons for not publishing. The cover letters were sent in Danish on behalf of the ethics committee secretary. Questionnaires remained in English.

Literature searches were conducted in February 2003 to identify publications for all trials. Investigator names and study keywords were used as search terms in PubMed, EMBASE, and the Cochrane Controlled Trials Register. Publications identified for survey non-responders were deemed to be associated with the corresponding protocols if they were conducted after the date of protocol approval, and if they were consistent in at least five of the following seven characteristics: a) individuals listed as authors or participating investigators; b) study sites; c) trial design; d) study groups; e) sample size (attained sample size within the range of 10% below or 20% above the planned sample size); f) study duration; and g) other distinguishing features. Searches for a random sample of 100 trials without survey responses were repeated in March 2003 as part of the quality control. No new publications were found.

A requirement imposed by the committee was that e-mail questionnaires be sent only from their on-site computer. This introduced a practical barrier to sending reminders from Oxford. A Danish researcher was thus recruited to send a single reminder to non-responders after six weeks.

2.2.4.3. CIHR cohort

Collaboration with the CIHR Clinical Trials Unit was required to obtain access to their files, and a confidentiality agreement was signed. Successful applications submitted to the CIHR clinical trials committee from 1990-98 were identified from their computer database. The database contains administrative information on

funding applications (successful or unsuccessful) submitted to the CIHR. Files for funded studies were obtained, and study protocols were reviewed to identify eligible randomised trials.

Publications reporting final trial results were identified using the same Publications Survey as for the Ethics cohort. Questionnaires were sent as individualised e-mail attachments along with an explanatory message (Appendix 4) in December 2002. Both English and French versions were included as required by CIHR. All correspondence was sent from a CIHR e-mail address on behalf of the unit director. In the absence of, or at the request of the main researcher, co-investigators were contacted. Questionnaires for the Publications Survey were sent by fax if a correct e-mail address was not available. Reminders were sent by e-mail approximately every two weeks, with phone calls being made after eight weeks.

Literature searches were conducted in February 2003 for non-responders using the same method described for the Ethics cohort.

2.3. Data Collection

2.3.1. Overview

The methods used to identify protocols and publications were described in the previous section. These protocols and publications, as well as the Outcomes Survey, constituted the three primary sources of trial data (Table 2-2). Data for the PubMed sample were recorded from publications and the Outcomes Survey, while the Ethics cohort extracted data from protocols and reports. Information for

the CIHR trials was collected from all three sources. Data extraction and entry were accomplished in a single step, with data being entered directly onto electronic forms in a Microsoft Access® database.

Table 2-2. Type of data collected from protocols, publications, and the Outcomes Survey for the trial cohorts

Cohort	Type of data collected from each source		
	Protocols	**Publications**	**Outcomes Survey**
PubMed	N/A	• Number and characteristics of reported outcomes • Number of unreported outcomes • Trial characteristics	• Statistical significance of unreported outcomes • Characteristics of unreported outcomes • Reasons for not reporting
CIHR	• Number and characteristics of specified outcomes • Trial characteristics	• Number and characteristics of reported outcomes • Number of unreported outcomes • Trial characteristics	• Statistical significance of unreported outcomes • Reasons for not reporting
Ethics	• Number and characteristics of specified outcomes • Trial characteristics	• Number and characteristics of reported outcomes • Number of unreported outcomes • Trial characteristics	N/A

2.3.2. Trial protocols

CIHR and Ethics protocols were reviewed to assess eligibility and to record trial and outcome characteristics. Formal amendments that were submitted to the CIHR were reviewed. However, submitted amendments for Ethics trials could not be identified due to the filing system of the committee. For Ethics trials, relevant portions of protocols written in Danish were translated and extracted by experienced researchers at the Nordic Cochrane Centre in Copenhagen.

2.3.2.1. Trial characteristics

Various trial characteristics were felt to be important for describing the study cohort, while some were also hypothesised to be relevant to outcome reporting practices. The trial design, type of intervention, specialty field, number of

randomised groups, planned sample size, number of data collection sites, funding sources, and use of blinding were recorded on an electronic form (Appendix 5). Operational definitions for each characteristic are described in Appendix 6. The specialty field was subjectively-defined, as trials were often cross-disciplinary.

2.3.2.2. Trial outcomes

In addition to extracting trial data, protocols were reviewed to record each study outcome and its characteristics (efficacy versus safety, data type, specification, and planned statistical test for analysis) using a standardised electronic form (Appendix 5). An outcome was defined as a variable intended to be assessed in all study participants for the purpose of comparing the effects of interventions between randomised study groups. It was defined as a specific measure with the time point of its measurement. This definition did not include participant characteristics collected solely at baseline, data intended to be collected only in a subgroup of participants, and data collected for administrative purposes. The assessment of whether a particular variable constitutes an outcome can be difficult; a detailed operational definition was therefore established and is described in Chapter 3.

The outcome type was classified as *efficacy* or *safety*. An *efficacy* outcome was defined as one that was used to measure any intended or desirable effects of the intervention, while a *safety* outcome was used to assess any harmful or undesirable effects. The data type was classified as *continuous*, *binary*, *ordinal*, *categorical* (> 2 categories), or *survival* (time-to-event) data based on the description of the outcome or its planned statistical analysis. For example,

mortality could be defined as a binary outcome or a time-to-event outcome, and its classification was determined by its description or its planned analysis as binary or survival data. The planned statistical test was also noted for each outcome.

Pre-specification of outcomes was classified as *primary*, *secondary*, or *unspecified* according to the protocol. A *primary outcome* was one described explicitly as the main or primary outcome; in the absence of such a statement, outcomes reported in the power calculation were used; if neither of the above was available, then outcomes specifically stated in the main study objectives were recorded. *Secondary outcomes* were those described explicitly as such, while *unspecified outcomes* were those not described as primary, main, or secondary.

2.3.2.3. Quality control

Protocol data on Ethics trial characteristics were reviewed twice by separate individuals, as well as by comparison to publications. Twelve Ethics protocols for which outcomes data extraction was deemed difficult were reviewed twice by different individuals to correct any inaccuracies. In addition, the first 15 protocols translated from Danish were double-checked to correct for any mis-interpretations due to the initial lack of familiarity with extracting outcomes data. Disagreements were resolved by consensus.

2.3.3. Trial publications

Trial reports were identified for all three cohorts. Using standardised forms (Appendix 7), data regarding trial characteristics, methodology, and outcomes were recorded.

2.3.3.1. Trial and journal characteristics

The same eight trial characteristics that were previously described for protocols were also recorded from trial reports. In addition, two journal publication characteristics were extracted from primary trial reports: the type of journal (general medical or specialty) and the type of report (brief or full length). Operational definitions are provided in Appendix 6.

2.3.3.2. Reporting of methodological details

The quality of reporting of various methodological factors was recorded from primary trial reports. Five factors were specified as possible surrogate markers for study quality: reporting of a power calculation, specification of primary outcomes, reporting of methods of random sequence generation and allocation concealment, and reporting of attrition (Appendix 8). The latter three markers of quality were only recorded in the PubMed cohort, after which a study was published indicating that the assessment of these parameters based on trial reports was unreliable.[117]

2.3.3.3. Trial outcomes

Each outcome listed in trial publications was recorded using the same definitions as for trial protocols. Data on all reported and some unreported outcomes were available from trial reports. Unreported outcomes were identified when they were

mentioned in the trial report (usually the Methods section), but no data were provided in the Results section. The outcome type (efficacy or safety), data type, and specification as primary, secondary, or unspecified were recorded for both reported and unreported outcomes, using the definitions described previously for protocols. For reported outcomes, the statistical significance, level of reporting, and statistical test used for analysis were recorded. The planned statistical test was noted only for the CIHR and Ethics trials, as the PubMed cohort did not have information on protocol-specified analyses for comparison.

Level of reporting and statistical significance

The two main outcome characteristics for evaluating selective reporting were its statistical significance and level of reporting. Statistical significance was defined as $p<0.05$ in any statistical comparison between randomised groups, including subgroup analyses. This value was chosen based on standard practice in the research community. Although the emphasis on *p*-values and the dichotomisation of results based on arbitrary thresholds is inappropriate,[92,118] it was felt that any bias based on statistical significance would act according to such a cut-off point.

Outcome reporting was defined at four primary levels based on the completeness of data presented in the results section of publications (Figure 2-1). A *fully reported outcome* was one with all of the data necessary for its inclusion in a meta-analysis. The nature and amount of data required to meet this criterion vary depending on the data type (Table 2-3). *Partially reported outcomes* had some but not all of the data necessary for meta-analysis, while *qualitatively reported outcomes* had no useful data except for a statement regarding statistical

significance or a *p*-value. Finally, u*nreported outcomes* were those for which no data were provided in the publication despite the outcome being defined in either the Methods section of publications or the trial protocol. Examples of the primary levels of reporting are shown in Table 2-4.

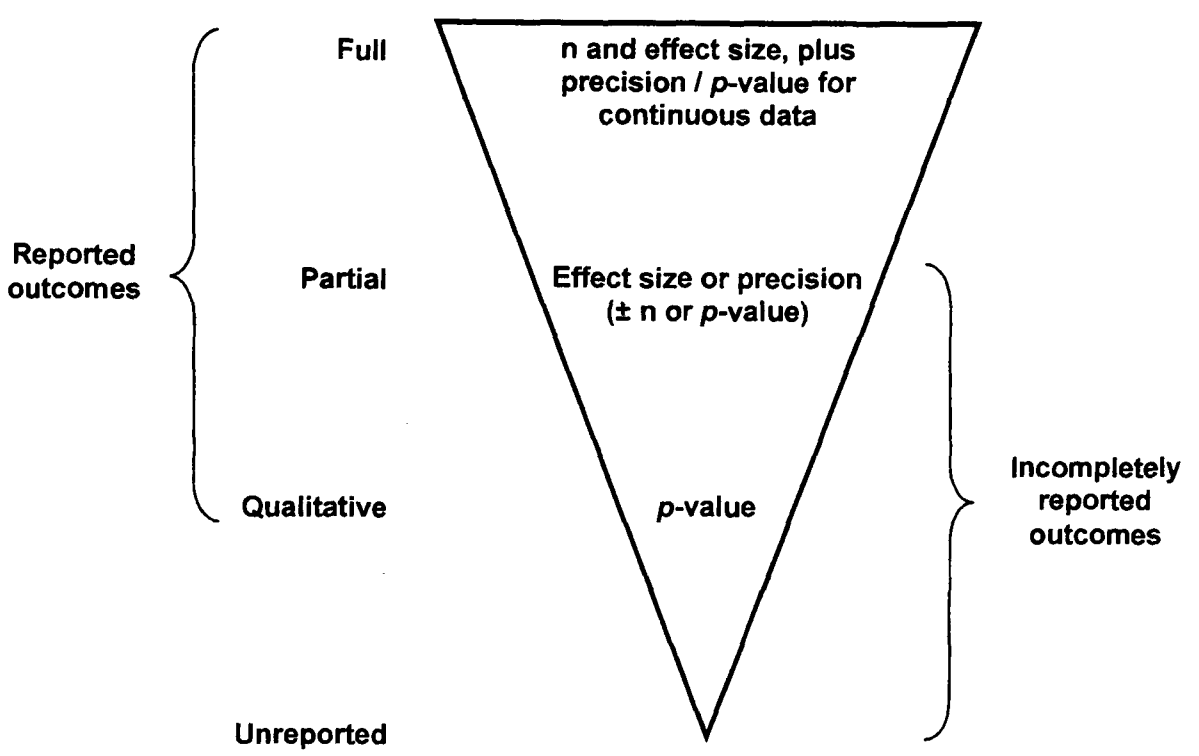


Figure 2-1. Hierarchy of the levels of outcome reporting (n = number of participants per group)

Table 2-3. Amount of data required for meta-analysis of fully reported outcomes

Type of outcome data	Data required for meta-analysis
Unpaired continuous data	• Group numbers; • Magnitude of treatment effect (group means/medians or difference in means); and • Measure of precision (confidence interval, standard deviation, or standard error for means; range for medians) or the precise *p*-value
Unpaired binary data	• Raw numbers or event rates in each group
Paired continuous data	Either • Mean difference between groups and a measure of its precision/exact *p*-value; or • Raw data for each participant
Paired binary data	• Paired numbers of participants with and without events
Survival data	Either • Kaplan-Meier curve with numbers of patients at-risk over time; or • Hazard ratio with a measure of precision

Table 2-4. Examples of the four primary levels of reporting for the continuous, unpaired outcome of mean diastolic blood pressure

Level of reporting	Example
Full	Diastolic blood pressure at 1 year was significantly higher in group A [mean 87.8 (19.2), n=63] compared to group B [mean 80.5 (19.6), n=60][a] or Diastolic blood pressure at 1 year was significantly higher in group A [mean 87.8, n=63] compared to group B [mean 80.5, n=60] (p=0.039).
Partial	Diastolic blood pressure at 1 year was significantly higher in group A [mean 87.8] compared to group B [mean 80.5].
Qualitative	Mean diastolic blood pressure at 1 year was significantly higher in group A compared to group B (p<0.05).
Unreported	No data provided

[a]Numbers represent mean (standard deviation) and number of participants (n)

It is important to define two additional terms that were used to describe relevant composite levels of reporting (Figure 2-1). *Reported outcomes* referred to those that were reported to some degree in the publications, and thus consisted of the highest three primary levels of reporting (full, partial, and qualitative). *Incompletely reported* outcomes referred to those that were inadequately reported for meta-analysis, and consisted of the lowest three primary levels collectively (partial, qualitative, and unreported). These terms will be used throughout the thesis.

2.3.3.4. Quality control

Trial characteristics were double-checked for each publication. For the PubMed and CIHR cohorts, the two reviews were conducted several months apart by the same individual. For the Ethics cohort, the reviews were conducted twice by two individuals. In cases where particular trial characteristics (trial design, number of data collection sites, funding sources, number of study groups) conflicted

between protocols and publications, the publications took priority as they were considered to be the more updated account of trial methodology. The quality of reporting was only assessed using information from publications.

For PubMed trials, the outcomes data were verified in a second review of publications several months apart by the same individual. Outcomes data extraction from the CIHR and Ethics cohorts was double-checked when protocols and reports were noted to be difficult, and for a random subset of 19 Ethics trials. Outcomes were also double-checked to confirm any major discrepancies between protocols and publications (to be described in Chapter 7). Disagreements were resolved by consensus.

2.3.4. Outcomes Survey

In addition to publications (all cohorts) and protocols (CIHR and Ethics cohorts), data on unreported outcomes were collected from trial investigators in the PubMed and CIHR cohorts (Table 2-2). Trialists were asked to complete a pre-piloted questionnaire with information about the nature and number of unreported outcomes in their trial. This survey is referred to as the Outcomes Survey in order to differentiate it from the Publications Survey that solicited publications for CIHR and Ethics trials.

Due to practical considerations, the Outcomes Survey could not be conducted for the Ethics cohort. The ethics committee required that all e-mail correspondence be sent from their on-site computer, and therefore questionnaires and reminders

could not be sent from Oxford. A Danish assistant could not be recruited within a reasonable timeframe to accomplish this task.

2.3.4.1. Development of the questionnaire

The original questionnaire was developed for the PubMed cohort, as this was the first sample examined. The following outline of its development therefore relates primarily to the Outcomes Survey for PubMed trials. A modified version of this questionnaire was used for the CIHR cohort, and the changes will be described at the end of this section.

A preliminary questionnaire was developed and reviewed by two statistician methodologists for face and content validity (Appendix 9). There was a need to balance the quantity of information solicited with the time and effort required to complete the questionnaire. Several findings from a systematic review on methods to increase postal questionnaire response rates were applied.[119] These included the use of shorter, 'user-friendly', and personalised questionnaires, as well as an emphasis on university sponsorship by including a colour seal from the University of Oxford in the cover letter. Follow-up contact was also planned. In March 2001, a pilot survey was conducted on a convenience sample of fifty PubMed-indexed randomised controlled trial reports published in January 2001. Questionnaires were faxed if a fax number was provided in the report; otherwise they were sent by post. A response rate of 35/50 (70%) was obtained over two months without any reminders being issued. This encouraging response rate was likely higher than what would be expected from a more comprehensive and less recent sample of trials, but it demonstrated that the survey method was feasible.

Following the pilot study, several modifications were made to the questionnaire. It was decided that a customised list of outcomes mentioned in each trial report would be sent with the questionnaire in order to help define and provide examples of outcomes for trialists in the PubMed cohort. In addition, 'Journal space limit' was divided into journal- and author-imposed space restrictions in order to determine the party responsible for omitting outcomes due to lack of space. Also, in order to identify secondary publications more easily, trialists who chose 'Reported elsewhere' as a reason for not reporting outcomes were requested to provide the study citations.

To further complete our data, a question about the number of study centres was added for trialists who failed to specify this information in their publication. Finally, it was noted that a large proportion of piloted trial reports (29/50, 58%) failed to indicate their sources of funding. A question asking about the source of funding was therefore added.

Final versions of the cover letter (Appendix 10), questionnaire (Appendix 11), and customised lists of outcomes were sent by e-mail, fax or post to the contact authors. In the absence of, or at the request of this author, co-authors were contacted. A first reminder was sent four to six weeks after the initial mailing. A second reminder was sent after another four to six weeks.

Modified versions of the cover letter (Appendix 12) and Outcomes Survey questionnaire (Appendix 13) were used for the CIHR cohort. Unreported outcomes identified from comparing protocols to publications were listed in the table so that trialists could simply fill in the data for each outcome provided, with

the opportunity to add additional unreported outcomes as well. The columns for 'specification' and 'clinical importance' were deleted because an indication of clinical importance was already available from the pre-specification of outcomes in trial protocols. The question regarding funding sources was moved from the Outcomes Survey to the Publications Survey questionnaire in order to reduce the number of questions, while the query about study sites was deleted because this information was available for every trial from protocols or publications. Because contact details were readily available, telephone reminders were conducted as a last resort for the CIHR cohort. Such reminders were considered impractical for the PubMed trialists due to lack of phone details for trialists located worldwide.

2.3.4.2. Outcomes Survey data

The Outcomes Survey elicited information regarding measured but unreported outcomes:

Number of unreported outcomes

For the PubMed cohort, trialists were asked to list all outcomes that were measured in the trial but not reported in the December 2000 publication. CIHR trialists were requested to add similar information to the list of unreported outcomes that were already provided for them.

Statistical significance

For each unreported outcome, trialists were asked to indicate whether $p < 0.05$ in any comparison between randomised study groups.

Clinical importance (PubMed cohort only)

Trialists were asked to score the clinical importance on an ordinal scale of 1 to 3 for each unreported outcome (1=low, 2=moderate, 3=high importance). The purpose of this question was to attempt to distinguish clinical importance from statistical significance as a reason for not reporting outcomes in the PubMed trials.

Pre-specification (PubMed cohort only)

An outcome was to be classified as *primary* or *secondary* if it was pre-specified as such, or as *unspecified* otherwise. This variable was included for the PubMed trials as another measure of the pre-defined importance of each unreported outcome.

Primary reason(s) for not reporting

Trialists were asked to provide their main reason(s) for not reporting each outcome. A checklist of common reasons was provided to simplify responses. These included journal- or author-imposed space limitations, lack of statistical significance or clinical importance, and reported elsewhere in a separate publication. 'Lack of statistical significance' was purposely placed in the middle of the list to avoid suggesting that it was the primary reason under investigation for not reporting outcomes. When outcomes were reported elsewhere, trialists were requested to provide citations so that these could be reviewed for additional outcomes data in an identical manner to primary reports.

Two additional questions were included in the Outcome Survey for PubMed trials:

Source of funding

Source of funding was deemed to be potentially important as a risk factor for outcome reporting bias, and was frequently not reported in trial publications. This question was also included in the Publications Survey questionnaire for CIHR and Ethics cohorts.

Number of study centres

For trial reports in which it was unclear whether the study was single or multicentre, a question was added to the survey to help complete our data.

2.3.4.3. Quality control

The reliability of survey responses was assessed by comparing characteristics of responders to non-responders, as well as comparing information provided in survey responses to that available from publications (see Section 2.5.2.1).

2.3.5. Summary of data collection

For each cohort, data were collected from at least two sources: trial reports (all cohorts), Outcomes Surveys (PubMed and CIHR cohorts) and/or protocols (CIHR and Ethics cohorts). Publications provided information on trial characteristics and reported outcomes. Some unreported outcomes could also be recorded from trial reports if they were mentioned in the Methods but not the Results sections. However, by definition, no details about statistical significance were available from the publications for these unreported outcomes. The Outcomes Survey of

trialists in the PubMed and CIHR cohorts served as the more comprehensive source of information for unreported outcomes, as they provided further details including statistical significance, clinical importance, and reasons for not reporting. Trial protocols were used to compare pre-specified outcomes to those reported in publications, enabling the identification of unreported outcomes as well as discrepancies in outcome characteristics such as specification or planned statistical analyses.

2.3.6. Sub-study: Randomised study of two e-mail survey methods

2.3.6.1. Rationale

It is unknown whether different methods of sending reminders for e-mail questionnaires result in higher response rates.

2.3.6.2. Objective

To assess which of two methods for sending e-mail reminders would yield the higher response rate at study completion and within six weeks.

2.3.6.3. Methods

Initial Outcomes Survey questionnaires were sent by fax or post for the PubMed cohort. E-mail addresses for non-responders were identified from publications or internet searches. Trials for which a contact e-mail address could not be identified were excluded. Using a computer-generated list of random numbers, trials were randomised to two reminder methods: e-mails with or without the original questionnaire and cover letter attached as a Microsoft Word® document. For e-mails with the attachment, a standard reminder text was used (Appendix

14a). For e-mails without attachments, a two-stage process was introduced. The reminder message was modified to request that trialists respond with a simple "Yes" if they were willing to complete the survey (Appendix 14b), and the original questionnaire was then sent to them upon request.

The primary outcome was the final response rate at the end of the PubMed study, six months after the first reminders were sent. The secondary outcome was the immediate response rate within six weeks of sending the first reminders. The six weeks corresponded to the time at which second reminders were sent.

Response rates were compared using 95% confidence intervals for the difference in proportions. The main analysis was conducted as intention-to-treat; per-protocol analyses were performed as a sensitivity analysis.

2.3.6.4. Results

The flow of study participants is shown in Figure 2-2. 110 trials were randomised to have no attachment, while 106 were assigned have attachments sent. Twenty trialists (12 without attachments, 8 with attachments) did not actually receive the reminder by e-mail due to incorrect e-mail addresses that were returned undelivered. Reminders for these trialists were sent within one day by fax (n=11 trials) if available, and by post (n=9 trials) otherwise.

Two trialists in the 'no attachment' group were unable to provide information about their trial due to lack of time or industry-control of data. The final response rates were 53/110 (48%) and 50/106 (47%) for reminders sent without and with attachments respectively (risk difference 1.0%, 95% confidence interval -12% to

14%). Response rates within six weeks of sending the reminders were 37% and 31% for reminders sent without and with attachments respectively (risk difference 6%, 95% confidence interval -7% to 19%). Per-protocol analyses produced similar results (49/98 [50%] vs. 48/98 [49%] for final response rates; 38/96 [40%] vs. 33/98 [34%] for response rates within six weeks).

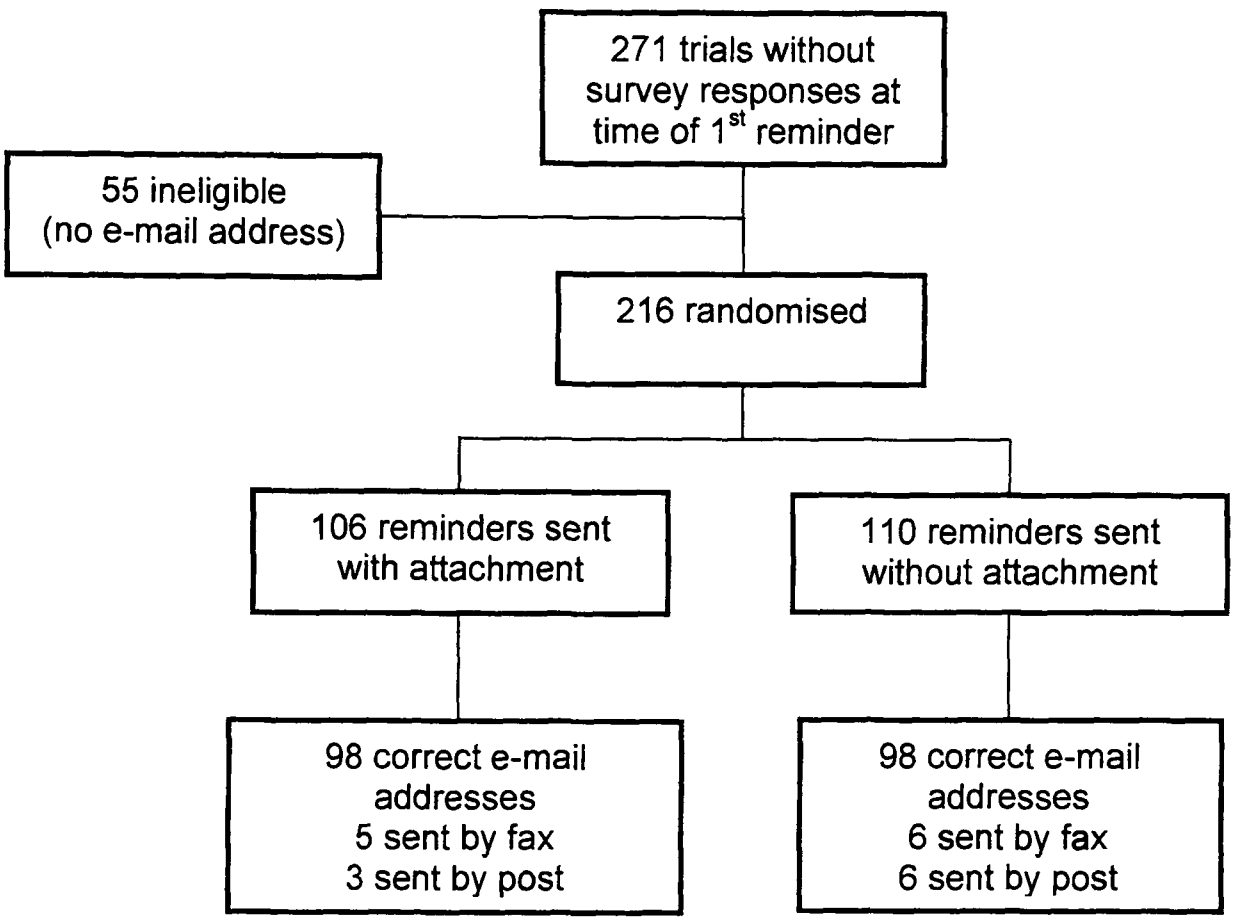


Figure 2-2. Flow chart of participants in the randomised trial of e-mail reminder methods

2.3.6.5. Discussion and conclusions

The widespread use of e-mail has introduced an economical, convenient, and rapid method of conducting surveys in an international sample. In the PubMed cohort, there was no evidence of a difference in short-term or final response rates to first reminders sent by e-mail with or without the questionnaire attached. Both survey methods had their potential drawbacks. Reminders sent with attachments may have been deleted for fear of computer viruses from an unrecognised sender, while reminders sent without attachments introduced an additional step

which may have facilitated or encouraged trialists not to make the effort to reply favourably to the request. Based on greater practical convenience when preparing e-mails without attachments, all second reminders were sent to PubMed trialists by e-mail without the questionnaire attached.

2.4. Preparation of final data sets

2.4.1. PubMed and CIHR cohorts

Data from the trial reports and the Outcomes Survey were merged to produce the final data sets for the PubMed (Figure 2-3) and CIHR (Figure 2-4) cohorts. Duplicate outcomes resulting from merging of the data sets were deleted. In the event that conflicting information was obtained from reports and surveys, the trial reports took precedence based on the assumption that publications would have been more carefully and reliably scrutinised than the survey responses.

All trials were to be included in the analyses regardless of whether responses to the Outcomes Survey were available. Missing data on unreported outcomes were solely the result of survey non-response. Non-responders whose trial reports provided no evidence of unreported outcomes were analysed conservatively as though no such outcomes existed. For the non-responders whose trial reports did provide evidence of unreported outcomes, these outcomes were included in the total number of trial outcomes to calculate the prevalence of poor outcome reporting. However, because the statistical significance of these unreported outcomes was not available from the survey, they could not be included in the analysis for outcome reporting bias (Figure 2-3 and Figure 2-4).

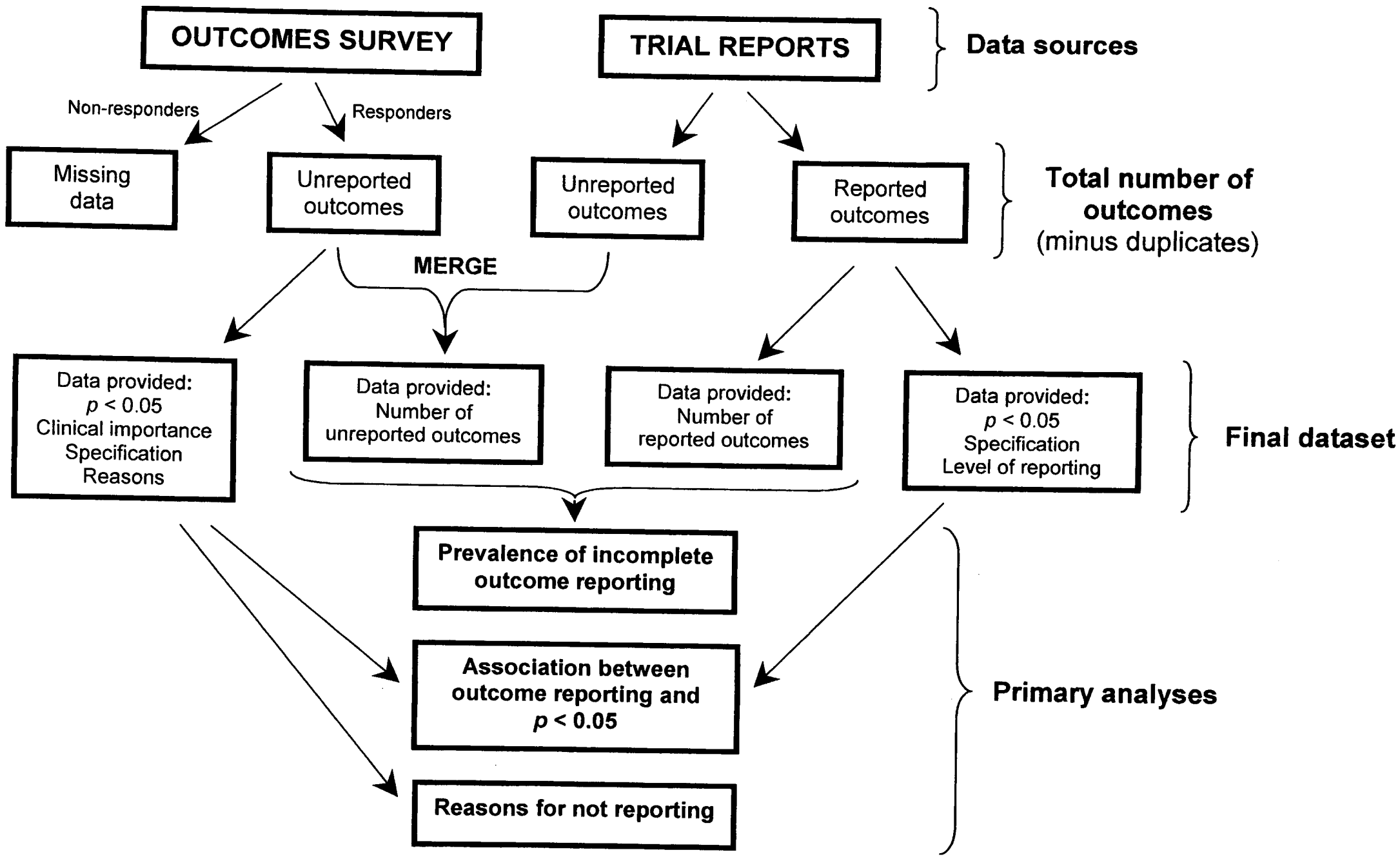


Figure 2-3. Origins of the final dataset for analysis of trial outcomes in the PubMed cohort

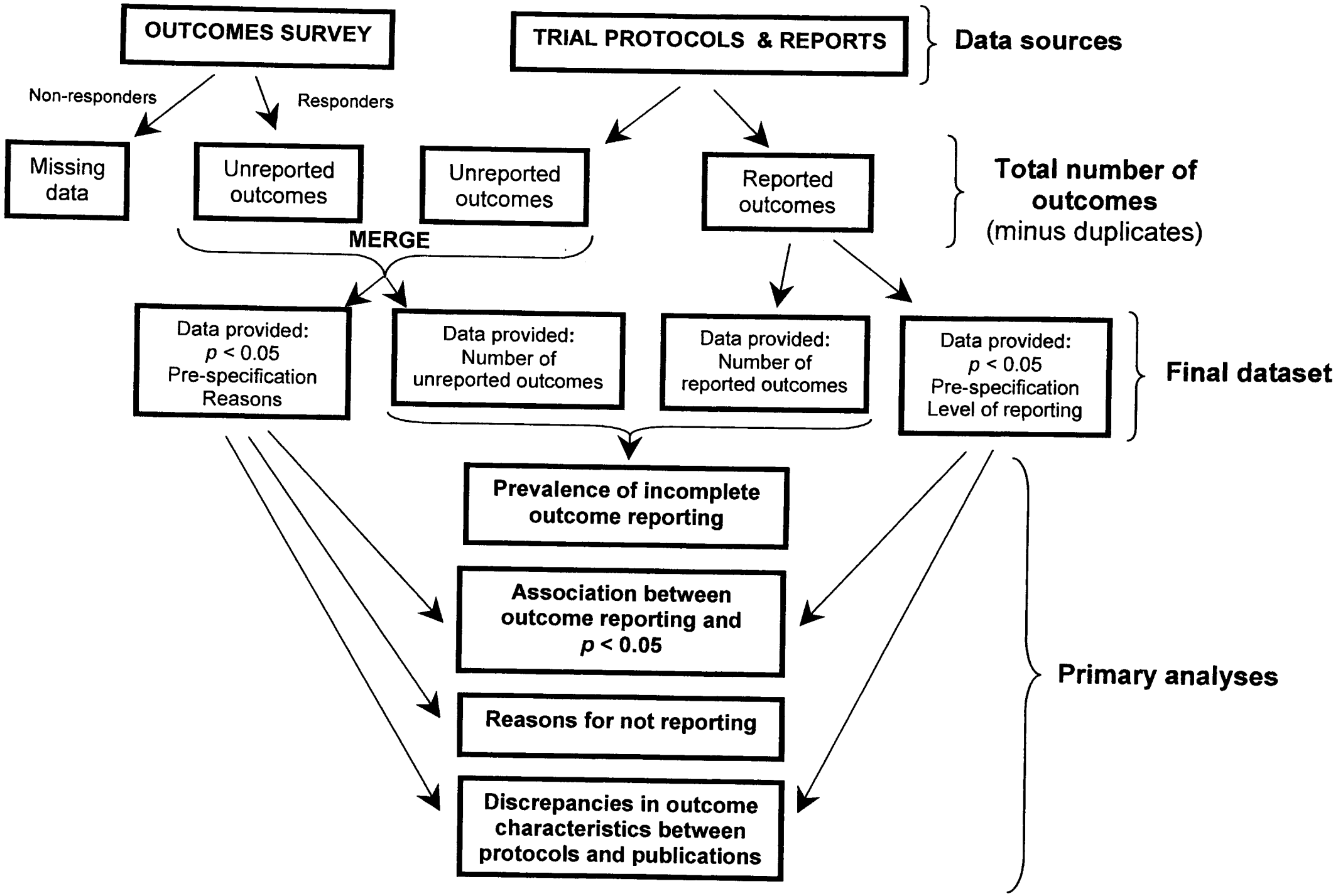


Figure 2-4. Origins of the final dataset for analysis of trial outcomes in the CIHR cohort

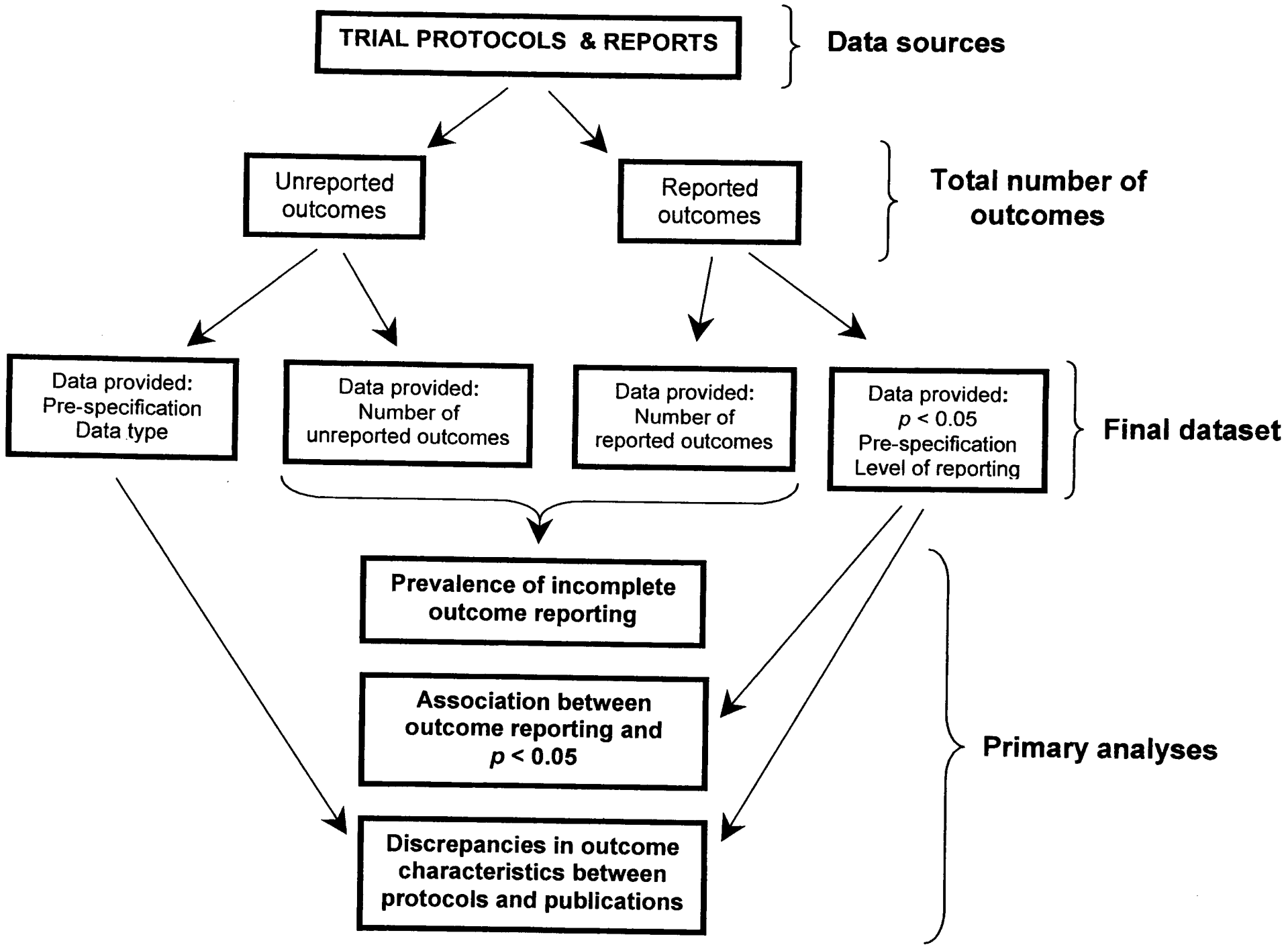


Figure 2-5. Origins of the final dataset for analysis of trial outcomes in the Ethics cohort

2.4.2. Ethics and CIHR cohorts

Data regarding trial characteristics were extracted from trial protocols, Publications Surveys (funding information only), and publications. These data were merged to produce the final data set (Figure 2-4 and Figure 2-5). In the case of conflicting data on trial characteristics from different sources, the primary publications were taken to be correct. Outcome data extracted from protocols and publications were kept separate for later comparison.

2.5. Data Analysis

The final datasets were analysed according to the main thesis objectives. Trial and outcome characteristics were summarised to provide a descriptive overview of the three cohorts. The prevalence of incomplete outcome reporting, its association with statistical significance, and risk factors for this bias were then determined. Finally, discrepancies between protocols and publications were examined in relation to outcome characteristics and analyses. All analyses were conducted using the statistical package Stata 7 (Stata Corporation, College Station, Texas, USA).

2.5.1. Cohort characteristics

Descriptive summary statistics were calculated to outline the trial characteristics of each study cohort. For unpublished trials in the CIHR and Ethics cohorts, reasons given in the Publications Survey for not publishing were tabulated.

2.5.2. Analysis of trial outcomes

Primary and secondary analyses are listed in Table 2-5. Efficacy and safety outcomes were analysed separately because the reporting of safety outcomes is known to be poor relative to efficacy outcomes.[74,120] Summary statistics were calculated to describe the nature and number of trial outcomes overall.

Table 2-5. Primary and secondary analyses for the evaluation of trial outcome reporting

Primary analyses	Secondary analyses
a) Median proportion of incompletely reported outcomes, and proportion of trials with at least one incompletely reported outcome b) Description of reasons given by investigators for not reporting outcomes (PubMed/CIHR only) c) Association between $p > 0.05$ and incomplete outcome reporting d) Proportion of trials with each of the following major discrepancies between protocols and publications (CIHR/Ethics only): • Introducing a new primary outcome that was not specified in the protocol • Omitting a pre-specified primary outcome • Changing the primary outcome specification to secondary or unspecified • Changing the outcome used in the power calculation • Changing the pre-specified analysis of primary outcomes	a) Median proportion of incompletely reported outcomes, and proportion of trials with at least one incompletely reported outcome, stratified by trial characteristics b) Description of clinical importance and pre-specification of unreported outcomes (PubMed only) c) Meta-regression to determine significant associations between trial characteristics and outcome reporting bias (PubMed only) d) Association between new outcomes and statistical significance (CIHR/Ethics only)

2.5.2.1. Reliability of Outcomes Survey responses

Characteristics of responders and non-responders were compared descriptively for the PubMed and CIHR cohorts. Differences were not formally tested as specific hypotheses were not defined. Concordance between survey responses and trial reports was also assessed; of particular interest were the number of surveys that stated no unreported outcomes despite the publications or protocols indicating otherwise.

2.5.2.2. Prevalence of incompletely reported outcomes

Incompletely reported outcomes were defined as those which presented insufficient data for meta-analysis (ie. partially reported, qualitatively reported, and unreported outcomes). The median proportion of incompletely reported outcomes per trial was calculated for all three cohorts, as well as the proportion of trials with at least one incompletely reported outcome. These proportions were also assessed after stratifying by journal type and funding source.

In merging the number of outcomes in protocols with the number of outcomes in publications for Ethics and CIHR trials, there was some uncertainty with regards to the subset of potential outcomes that were not clearly specified in the protocol and not subsequently mentioned in the trial reports. Although these variables were not recorded as outcomes for our analyses, their description was vague to the extent that they may have been intended for inter-group comparisons. Including and excluding such outcomes would result in over- and under-estimates of the true number respectively (Figure 2-6). A sensitivity analysis including these outcomes was thus conducted to provide worst-case estimate of the proportion of incompletely reported outcomes per trial.

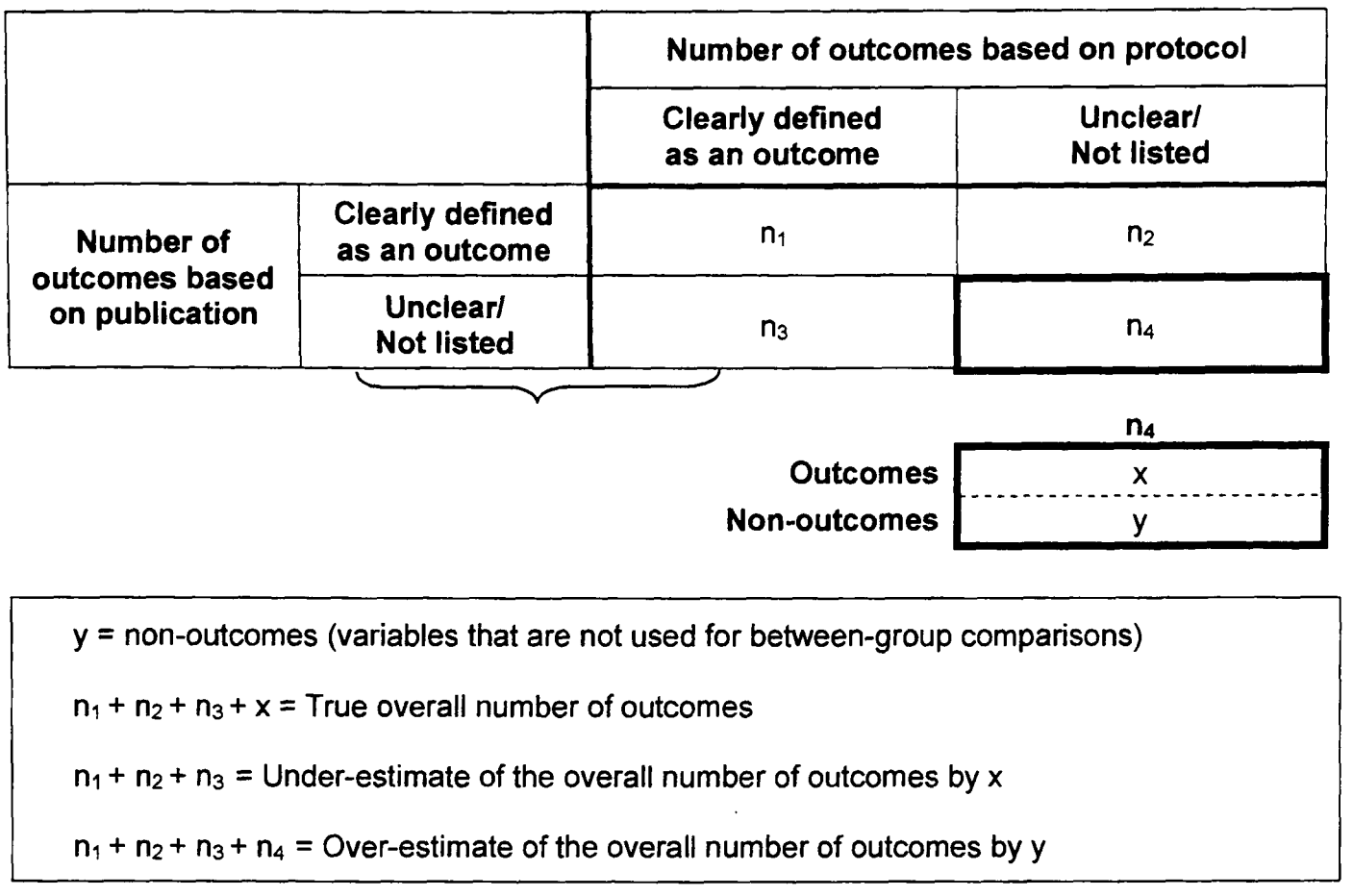


Figure 2-6. Inaccuracy in the overall number of outcomes analysed per trial

2.5.2.3. Statistical significance and outcome reporting

From the Outcomes Survey responses, reasons for not reporting outcomes were tabulated and described (PubMed and CIHR cohorts), as were the pre-specification and clinical importance of unreported outcomes (PubMed cohort only). A 2x2 table relating statistical significance/non-significance to fully reported/incompletely reported outcomes was created for each trial. An odds ratio for outcome reporting bias, defined as the odds for a fully reported outcome being statistically significant divided by the odds for an incompletely reported outcome being significant, was calculated for each trial. The individual odds ratios were then pooled across trials using a random effects model[121] to produce an overall estimate of bias with its 95% confidence interval.

Studies with no outcomes in adjacent cells of the 2x2 table did not contribute useful information to this analysis of bias, as meaningful odds ratios could not be calculated. Trial and outcome characteristics for these excluded studies were compared descriptively to included studies.

The odds ratio was used because of its favourable mathematical properties and the interchangeability of the 'event' and 'non-event' coding. A random effects model was chosen based on the assumption that a population of true values of bias existed among trials in a normally distributed manner, rather than assuming a single true value of bias between trials.

Sensitivity analyses were conducted by excluding trials without survey responses, as well as excluding physiological and pharmacokinetic studies. The impact of using a different cut-off point for dichotomising the level of reporting (fully/partially reported versus qualitatively/unreported outcomes) was also examined in a third sensitivity analysis.

2.5.2.4. Trial characteristics and outcome reporting bias

Evaluation of risk factors for outcome reporting bias was conducted in an exploratory analysis using meta-regression. The between-studies variance was estimated based on restricted maximum likelihood.[122] The analysis was performed only for efficacy outcomes in the PubMed cohort, as there was a sufficient number of trials to explore numerous parameters. Statistical methods and the problems with meta-regression are outlined in Appendix 15.

Univariate meta-regression was used to determine the effect of thirteen factors on the odds ratio for outcome reporting bias (Table 2-6). These covariates were felt to be potentially associated with study quality and selective outcome reporting. Sample size was log transformed to correct for a non-linear relationship. Backward stepwise meta-regression was used to assess any association between the thirteen covariates and outcome reporting bias in a multivariate model (Table 2-6).

Table 2-6. Thirteen exploratory variables examined using meta-regression

Covariate	Data type	Coding[a]
Trial design	Dichotomous	0 = parallel 1 = other
Type of intervention	Dichotomous	0 = drug 1 = other
Journal type	Dichotomous	0 = general 1 = specialty
Type of journal report	Dichotomous	0 = short/letter 1 = full length
Log(sample size)	Continuous	Continuous
Type of funding	Dichotomous	0 = full industry 1 = other
Number of sites	Dichotomous	0 = single centre 1 = multicentre
Specification of primary outcome(s)	Dichotomous	0 = not specified 1 = specified
Power calculation	Dichotomous	0 = described 1 = not mentioned
Reporting of random sequence generation	Dichotomous	0 = adequate 1 = inadequate
Reporting of allocation concealment	Dichotomous	0 = adequate 1 = inadequate
Use of blinding	Dichotomous	0 = adequate 1 = inadequate
Reporting of attrition	Dichotomous	0 = adequate 1 = inadequate

[a]Defined in Appendices 6 and 8 for each trial and reporting characteristic

2.5.2.5. Discrepancies in outcome characteristics between protocols and publications

Outcomes listed in protocols and publications were compared for CIHR and Ethics trials. The number of reported outcomes that were vaguely described in the protocol was noted, and the proportion of trials with major discrepancies

were calculated. Major discrepancies of interest were listed in Table 2-5. They included the introduction of new primary outcomes, failure to report a pre-specified primary outcome, changing the specification of the pre-defined primary outcome, changing the outcome used in the power calculation, and changing the analysis plans of primary outcomes. The proportion of trials with new outcomes was also calculated for primary and non-primary outcomes combined. The odds ratio for a new outcome being statistically significant was determined for each trial, and then pooled using a random effects model. Finally, the proportion of trial reports that mentioned any amendments to the original protocol was determined.

A major discrepancy in outcome analysis was recorded if the statistical analysis in the report used a different data type from that specified in protocols. Data types were grouped as continuous/ordinal, binary/categorical, and survival data based on the type of statistical test used. Changes between these three categories were considered to be major discrepancies only when a new unspecified data type was reported in place of the original specified data type. For example, in the case of a variable that was described in the protocol as a continuous outcome to be analysed with a t-test, a major discrepancy was recorded if the publication only reported a dichotomised variable analysed using a chi-square test. However, it was not considered a major change if the publication reported separate analyses of the outcome as both continuous and binary data. This would constitute an additional analysis and outcome rather than a change in analysis. In addition to changes in data type, a major change was recorded if the analysis plan was altered from a descriptive tabulation to a hypothesis test.

The existence of major discrepancies between primary outcomes in protocols and publications was confirmed by at least two individuals who re-verified the raw outcome data under consideration. Disagreements were resolved by consensus.

2.6. Chapter summary

Data on reported and unreported outcomes were collected from trial reports, protocols, and surveys. Trial reports primarily provided information about reported outcomes; some unreported outcomes could also be identified if outcomes were mentioned in the Methods section but not subsequently reported in the Results section. The availability of protocols provided an objective indication of the number of trial outcomes intended for measurement in each trial, enabling a more accurate estimate of the number of unreported outcomes in the Ethics and CIHR cohorts. Protocols also enabled the identification of discrepancies in outcome specification and analysis. Finally, the Outcomes Survey provided details about unreported outcomes, particularly their statistical significance. These three complementary sources of data allowed for a comprehensive assessment of inadequate and selective reporting of outcomes.

Chapter 3 - Problems with the use of trial protocols in the evaluation of outcome reporting

3.1. Introduction

Evaluating biased practices among researchers presents several difficulties. The potentially incriminating nature of this type of investigation renders it less likely that accurate information about study conduct can be obtained *a posteriori* from those who are most familiar with what actually took place - the researchers themselves. In the case of collecting information about trial outcomes, it is important to use data sources that are independent from both study results and reporting. Such independence involves two components - the data source should have been compiled prior to study initiation, and access to it should not be dependent on those who are aware of the study results or reporting practices.

Trial protocols fulfill these requirements when access to them is not dependent on consent being granted by the authors or sponsors. Protocols constitute an objective account of what was planned *a priori* for a study. If properly written, they should provide a reliable estimate of planned trial outcomes along with an unbiased description of outcome characteristics and analysis plans.

The previous chapter outlined methods of assessing outcome reporting practices using protocols of randomised trials approved by the Canadian Institutes of Health Research (CIHR) and a Danish ethics committee. Methodological aspects surrounding the use of study protocols have not previously been explored nor refined to a large extent. Three previous reviews

of confidential study protocols have been published. One study examined deficiencies in the study design and analysis of 75 protocols and 33 publications using biomedical studies approved by the University of California San Francisco Committee on Human Research in 1974, 1978, and 1980.[123] A second review was a pilot study that compared the outcomes in 15 ethics protocols to those reported in their publications.[98] Finally, a recent review examined selective reporting of analyses based on protocols submitted to a Swedish drug regulatory agency.[19]

With the paucity of prior reviews of confidential protocols, the methods applied to the CIHR and Ethics cohorts provided an opportunity to identify issues and obstacles to the reliable assessment of outcome reporting practices using trial protocols. This chapter will outline two aspects of the methodology that proved to be particularly challenging: acquiring access to trial protocols without consent from researchers, and dealing with vagueness in the descriptions of trial outcomes and analyses.

3.2. Access to protocols

Obtaining comprehensive access to trial protocols proved to be a long and difficult process. The primary prohibitive issue revolved around the question of whether permission was needed from protocol authors and sponsors. There remains, however, a strong ethical and scientific argument for allowing protocols to be reviewed anonymously for methodological research. Institutional review boards and funding agencies have an ethical obligation to ensure that that their results are fully disseminated, and that clinical studies are carried out using the same methodology described in the protocols upon

which approval was based.[124-126] Although a policy of study monitoring may be unfeasible given the resource constraints for ethical and scientific review bodies, the opportunity should at least be granted for independent researchers to examine the issue of consistency between protocols and publications. Such a review would benefit institutions and the scientific community, either by highlighting any deficiencies to provide evidence of the need for improvements, or by demonstrating that biased practices do not generally exist and that monitoring mechanisms may thus be unnecessary.

To achieve an unbiased sample, comprehensive access to all protocols in a given time period is essential, meaning that access should not be dependent on consent being granted by the protocol authors and/or sponsors. In three previous reviews of confidential study protocols, the San Francisco study[123] (carried out more than 20 years ago) and the recent Swedish study[19] proceeded without the need for consent from each researcher, while the recent British study required consent from investigators.[98] Despite recognising the importance of the type of research proposed, and despite being given assurances that anonymity would be preserved, various institutions in the United Kingdom (UK) and Canada refused to grant comprehensive access to protocols on the grounds that confidentiality of the documents had to be preserved.

3.2.1. Experiences in the United Kingdom, Canada, and Denmark

To briefly summarise the chronology of events, applications to local and multicentre research ethics committees in the UK over 14 months were unsuccessful in obtaining access to approved study protocols without explicit

consent from investigators and sponsors. Efforts to access protocols funded by the British Medical Research Council were also unproductive. Through personal contacts, permission was obtained in August 2002 to review protocols from a university research ethics committee in Canada. Unfortunately, despite having received formal ethical approval, the study was terminated after six weeks by the faculty administration on the advice of university lawyers. Finally, through personal contacts once again, access to protocols from both the CIHR and an ethics committee in Copenhagen was obtained in October 2002.

3.2.1.1. Unsuccessful efforts in the United Kingdom

An initial informal request to access trial protocols was submitted to the administrator of a local Clinical Research Ethics Committee in the UK in March 2001. The Committee recognised that this was "an interesting study which would be well worthwhile," but expressed concern regarding confidentiality and stated that consent from investigators would be required. Following consultation with The Oxford Centre for Ethics and Communication in Health Care Practice, a formal application was made to the same committee in November 2001. The proposal outlined the ethical and scientific rationale for allowing a review of trial protocols to proceed without obtaining consent from each investigator. The Committee once again "recognise[d] the importance of the study, [but] felt that it would not be appropriate ... to approve this application because it involve[d] information that [it] keeps." The Committee decided to refer to the Central Office for Research Ethics Committees (COREC). In April 2002, the Committee informed us that COREC had

suggested that an application should be made to a Mulitcentre Research Ethics Committee (MREC).

A separate application was thus made to an MREC in May 2002. The MREC also consulted with COREC, which advised that consent was required not only from investigators, but also from sponsors. The primary justification was that the protocols had been submitted in confidence, and it would not be possible for the committee to even release contact details for the investigators. The MREC suggested that a letter requesting consent to release contact details and protocols be drafted and re-submitted with a revised application. We decided not to pursue this route for reasons of feasibility and scientific validity, which will be discussed later.

A final effort was made to conduct the study using protocols of trials funded by the UK Medical Research Council (MRC). However, it was explained that the MRC does not obtain trial protocols for most of its applications, even for studies that are funded. There were also concerns regarding confidentiality and data protection legislation. Efforts to conduct the review of protocols in the UK were thus abandoned.

3.2.1.2. Unsuccessful efforts with a Canadian ethics committee

The use of university ethics committee protocols in Canada received informal approval in August 2002 from the Director of the Office of Medical Bioethics, who also agreed to act as the local co-investigator. Formal approval from the ethics committee was obtained in September 2002, and consent from investigators or sponsors was not felt to be necessary. The committee was

responsible for all clinical research in the surrounding geographical region. Titles of potential clinical trials approved from 1993-95 were obtained from a computer database, and relevant protocols were retrieved from archives. In October, with no advance notice, the study was unexpectedly terminated by the university administration and legal department. At the time of termination, over 140 eligible trials had already been identified from 275 study protocols, and outcomes data had been extracted from 143 trial protocols.

The university lawyer explained that the study could not proceed in its approved form because of a contract between the university and pharmaceutical companies that prohibited the release of their protocols to those not involved with the conduct or ethical review of the study. It was felt that the research project did not fall under the normal mandate of the ethics committee.

The lawyer further explained that the only way the study could proceed was if explicit consent from each study's sponsor was obtained. However, because public access to information was limited to study titles and investigator names, a multi-stage process would have been required. Study investigators would have to be contacted to identify sponsors. For non-responders, the only way to identify sponsors for particular studies would have been to contact every pharmaceutical company in Canada to enquire whether they had conducted studies in the region from 1993-95, and if so, whether they could provide a list of such studies. If a positive response was received, then additional permission to access each study protocol would have to be obtained. For

reasons of compromised scientific validity and feasibility, this approach was not pursued any further.

3.2.1.3. Successful efforts in Canada and Denmark

Through personal contacts, positive responses were received from both the head of the randomised trials unit of the CIHR and the secretary of the regional ethics committee in Copenhagen. Both institutions felt that it was their duty to monitor the conduct and reporting of trials that they approved. Each study was therefore adopted as a joint venture, with administrative support from the respective institutions. Confidentiality agreements were signed, and comprehensive access to protocols was granted.

3.2.2. The case supporting requirements for consent

The primary concern with waiving consent revolves around the confidentiality of study protocols. With regards to ethics committees, applications were submitted confidentially for the specific purposes of the ethical review. Although it can be argued that monitoring of trial reporting should constitute part of the review process, this is generally not part of the current activities of most ethics committees. The issue of ownership of study protocols has also been raised with regards to commercial and intellectual property. This has implications for determining which individuals have the right to waive confidentiality when granting permission to review protocols. However, the same Danish ethics committee which had approved the original trials did not feel that consent was necessary from investigators or sponsors to conduct our study. For the CIHR cohort, the agency was the sponsor and our study did not require ethical approval as no patient information was involved.

3.2.3. The case against requirements for consent

The procedures for obtaining consent to review study protocols, as required by ethics committees in the UK and a university legal department in Canada, would clearly lead to a biased sample, and are impractical and labour-intensive.

3.2.3.1. Compromise to scientific rigour

Of primary scientific concern is the bias introduced by the consent process. The requirement for consent from each sponsor would undermine the validity of the study results due to response bias. Sponsors and researchers who suppress or change outcomes would be less likely to give consent to have their protocol scrutinised - particularly if the comparison of protocols to publications would reveal poor research practices. Industry sponsors in particular would have little commercial interest in releasing their confidential protocols. There is thus a higher likelihood that those who provide consent would be systematically different from those who do not. Conclusions from such a biased sample would not be sufficiently reliable.

3.2.3.2. Lack of feasibility

A further concern involves the feasibility of soliciting consent from sponsors and investigators. It would not be possible to identify all randomised trials from the list of project titles available in the public domain. To be comprehensive, representative, and unbiased, investigators for each clinical study would have to be contacted regardless of presumed design, which is laborious and impractical. Given that only a small percentage of protocol submissions will describe a randomised trial, that investigators' contact details will have

changed from the time of protocol submission several years earlier, and that the response rate to surveys requesting consent (whether positive or negative) would not be 100%, the yield would be low for a massive amount of effort.

3.2.3.3. Lack of alternatives to address an important study question

As outlined previously, the assessment of outcome reporting bias using study protocols has direct societal significance in terms of medical decision-making and health policy. The importance of the project was not disputed by any of the committees that were approached. By examining whether the medical literature presents a biased set of outcomes for healthcare interventions, the proposed study would provide an indication of the reliability of systematic reviews and the evidence upon which the medical field relies. The project also fulfills an ethical and scientific obligation of ethics and funding agencies to ensure that trials are conducted and reported in a manner consistent with the protocol upon which original approval was based. If bias were identified, then methods for improvement and monitoring can be investigated in the future.

Unfortunately, there is no equally-reliable alternative to the use of study protocols. The use of the PubMed cohort presented one alternative methodology, but complete reliance on self-report data is less reliable than protocols that are written *a priori*. A prospective study would not be appropriate, as trialists may be influenced by enrollment in the study. It would also take several years for trial publication to occur.

3.2.3.4. Ethical justification

The ideal situation from an ethical perspective would involve obtaining consent from every investigator and sponsor to access his/her protocol. However, according to the UK Department of Health[127] and Medical Research Council (MRC)[128] guidelines, there are instances in which the need for such consent can be outweighed by the importance of the study and practicability.

With regards to epidemiological studies of medical records, the Department of Health states that "there will be occasions when a researcher would find it difficult or impossible to obtain consent from every individual and the ethics commitee will need to be satisfied that the value of such a project outweighs, in the public interest, the principle that individual consent should be obtained."[127] A similar statement has been issued by the MRC. The MRC also adds that there must be "no practicable alternatives of equal effectiveness", as well as "no intention to take decisions that affect [the individuals involved]."[128] By fulfilling the above criteria, the use of study protocols is analogous to the circumstance of 'no consent, low risk' research involving patient records. There is thus a strong ethical rationale for allowing the review of trial protocols without obtaining consent from investigators or sponsors.

Given that a) a comparison of outcome reporting in protocols and publications has important societal implications and has rarely been conducted; b) consent is impractical; c) the study could not be conducted if individual consent was required; d) there are no equally-reliable alternative methodologies; and e) anonymity would be maintained for reporting of results, the proposed thesis

methodology fulfills the criteria outlined by the Department of Health and the MRC for 'no consent' research.

Unfortunately, ethics committees in the UK appear to disagree with this viewpoint. Even when an ethics committee in Canada upheld this argument, the study was terminated by the legal department. In contrast, the Copenhagen ethics committee as well as the CIHR funding agency were very supportive of this type of research. It is evident that the issue of confidentiality versus the overall benefit to society remains controversial with regards to the use of protocols. Hopefully as more research of this nature is conducted, the justification for comprehensive access to protocols will not only become widely-accepted, but will also prompt institutions to support this type of research as a joint venture to monitor studies that they approve.

3.3. Vagueness in trial protocols

As defined briefly in Chapter 2, an outcome was defined as a variable intended for comparison between randomised groups at a defined time point. Although this definition may appear to be relatively straightforward, outcome descriptions were frequently noted to be vague when reviewing the convenience sample of 50 PubMed reports for the pilot survey. This observation prompted the development of strict guidelines for defining specific outcomes. Vagueness in defining outcomes in the Methods section of either protocols or publications has implications when determining the number of outcomes measured in a trial. The responses to Outcomes Surveys can help to clarify the status of vague variables as outcomes or non-outcomes.

A vague outcome can be defined as a variable whose status as an outcome is unclear. Vagueness can arise from four situations that are not mutually exclusive. In the first scenario, a variable may be listed without sufficient details to determine its data type (binary, continuous, ordinal, categorical, or survival) for analysis. For example, a variable listed as 'mortality' may represent a single binary or time-to-event outcome, or both types of outcomes. The second type of vague outcome involves the failure to define specific time points of clinical interest for analysing the final data. Even when the time points of measurement are described, it often remains unclear as to whether comparisons will be made at all or some of the time points, whether changes from baseline or absolute values will be analysed at certain time points, or whether a global summary comparison will be made, such as repeated measures or area-under-the-curve analyses. The third situation in which vagueness may be observed involves composite outcomes, where it is unclear whether each of the sub-components will also be compared between randomised groups. Finally, it can be unclear whether a variable listed for measurement post-baseline is actually an outcome intended for inter-group comparison, or whether it is a non-outcome variable collected for monitoring or descriptive purposes.

Vagueness may arise for several reasons. Protocol authors may not be aware that specific details are required in order to rigorously define outcomes and analyses *a priori*. In particular, descriptions of statistical analyses may suffer from a lack of detail when statistical expertise is not readily available. Another possibility is less innocuous - that investigators are intentionally vague in order to avoid committing their study to specific outcomes and analyses. This

approach would create the opportunity for data dredging and selective reporting of interesting findings. As advised by the International Conference for Harmonisation guidelines, "it is critical to specify in the protocol the precise definition of the primary variable as it will be used in the statistical analysis."[129] This should apply to all trial outcomes unless they are explicitly described as being exploratory in nature.

3.3.1. Operational protocol for recording vague outcomes

A standardised method of dealing with vague outcomes was required to minimise potential subjectivity and bias in the recording of outcomes. For each type of vague outcome, assumptions could be made that would result in either conservative estimates or over-estimates of the actual number of outcomes in a trial. A decision was made to aim for conservative estimates by assuming that vague variables were non-outcomes unless there was evidence suggesting otherwise. By recording outcomes in this manner, inflated estimates of the number of outcomes per trial could be avoided.

The conservative approach would also minimise the misclassification of non-outcome variables as outcomes, thus reducing the risk of over-estimating the prevalence of unreported outcomes. Such over-estimation might otherwise occur if several non-outcome variables were recorded as outcomes, and were subsequently not reported in the publications. One might have concluded that these were unreported outcomes when in fact they were intended for purposes other than inter-group comparisons.

3.3.1.1. Unclear outcome type

Variables may have been listed as study outcomes, but without sufficient details about the specific type of data to be used in analysis. For these outcomes, the data type was recorded as 'unclear' based on the protocol. Information reported in the publication was then used to define the outcome. For example, if the vague protocol outcome was 'Death due to stroke at two years', this could be a binary or survival outcome. If the publication only reported the 'time to stroke-related deaths at two years', then this single outcome was counted. If the publication reported stroke-related deaths as both survival and binary (death rate) data, then this would recorded as two separate outcomes.

3.3.1.2. Multiple time points

Data collection at multiple time points in a trial does not necessarily mean that each time point is clinically important. Multiple visits may be conducted to monitor compliance or safety, and to obtain regular measurements for use in 'last observation carried forward' analyses should the participant be lost to follow-up. It would thus be inaccurate to automatically count each time point as a separate outcome. Doing so would potentially over-estimate the true number of outcomes. On the other hand, it is plausible that multiple comparisons are in fact made at each time point, with selective reporting of the time points demonstrating the most desirable results. Not counting each time point as a separate outcome may thus lead to an under-estimate of the true number of outcomes.

A conservative approach was adopted for recording such outcomes from trial protocols and publications. Variables measured at multiple time points were counted as separate outcomes, provided that they were each analysed separately. Global comparisons across time points, such as repeated measures analyses, were counted as a single outcome.

3.3.1.3. Composite variables

Composite outcomes combine several different variables into a single outcome. There are generally two types of composite outcomes. In the case of binary variables, a composite outcome usually consists of several possible events. A participant is said to have had the outcome if at least one of the component events occurred. The second type of composite outcome consists of continuous data, and most commonly involves multiple scale measurements. An overall summary score can often be derived from several individual subscale scores.

A conservative approach was used to record composite outcomes from trial protocols and publications. Subscales and individual components of composite outcomes were counted separately only if they were each analysed individually.

3.3.1.4. Unclear outcome status

A variable may have been listed for measurement in the study without being specified as an outcome intended for between-group comparisons. The majority of these variables were safety outcomes. An example would include standard laboratory tests, which were often measured in industry-sponsored

drug trials. The intended use of these variables may be simply descriptive or regulatory in order to fulfill the requirements of licensing agencies.

A conservative approach to recording such outcomes from trial protocols and publications was adopted. These vague variables were counted as outcomes only if they were analysed in inter-group comparisons. Unsolicited or spontaneous adverse events were grouped as a single outcome. Solicited adverse effects were counted separately if they were intended for comparison.

		Number of outcomes based on protocol	
		Clearly defined as an outcome	**Unclear/ Not listed**
Number of outcomes based on publication	**Clearly defined as an outcome**	n_1	n_2
	Unclear/ Not listed	n_3	n_4

n_4

Outcomes	x
Non-outcomes	y

y = non-outcomes (variables that are not used for between-group comparisons)

$n_1 + n_2 + n_3 + x$ = True overall number of outcomes

$n_1 + n_2 + n_3$ = Under-estimate of the overall number of outcomes by x

$n_1 + n_2 + n_3 + n_4$ = Over-estimate of the overall number of outcomes by y

Figure 3-1. Vagueness in the overall number of outcomes analysed per trial

3.3.2. Implications for data analysis

The effects of vagueness on data analysis are illustrated in Figure 3-1, which was first introduced to describe a sensitivity analysis in Chapter 2. n_4 represents the number of vague variables in a trial, as described in protocols and publications. A certain number of these vague variables will be true outcomes (denoted by x), while the remainder will be non-outcome variables

(denoted by y). The uncertainty in the total number of outcomes in a trial arises from the fact that the values of x and y can never be truly known from reviewing protocols and publications, provided that $n_4>0$. The number of outcomes recorded will either over-estimate the true number by y, or under-estimate the true number by x, depending on the decision to include or exclude the vague variables respectively. Supplemental information from the Outcomes Survey can help to distinguish true outcomes from non-outcome variables.

3.4. Chapter summary

Although a comparison of protocols to publications is simple in theory, practical issues arise when applying the methodology in practice. Legal and ethical barriers to comprehensive protocol access, as well as vagueness in outcome definitions, constitute major obstacles to the reliable assessment of outcome reporting bias based on trial protocols and publications. Pragmatic solutions to these issues include conducting the study as a joint venture with ethics or funding institutions, adopting conservative measures to record vague outcomes, and using Outcomes Surveys of authors to clarify the status of vague variables as outcomes or non-outcomes.

Chapter 4 - Epidemiology of trial cohorts

4.1. Introduction

Randomised controlled trials have become the foundation for systematic reviews and evidence-based clinical practice. Although a search of the Cochrane Controlled Trials Register (Issue 3, 2002) yielded 17,074 citations with publication year 2000, little is known about the epidemiology of this important population of studies. This chapter describes the process of sample selection and the characteristics of three trial cohorts that were used to examine outcome reporting bias: PubMed-indexed trials whose primary reports were published in December 2000; published trials funded by the Canadian Institutes of Health Research (CIHR) from 1990-98; and published trials approved by a Danish ethics committee from 1994-95. A discussion of sampling biases and their potential effects on the validity of the three cohorts in assessing outcome reporting bias will conclude the chapter.

With regards to evaluating outcome reporting bias, it is important to examine the characteristics defining each sample in order to assess the external and internal validity of conclusions drawn from their analyses. From a global perspective, the cohorts also provide a unique opportunity to describe the overall composition of various populations of randomised trials. To our knowledge, a broad review of trials across study designs, specialties, and journals has not been conducted since 1980.[96] An updated panoramic snapshot would help to guide current perceptions and future methodological research.

4.2. Process of trial selection

4.2.1. Identification of index trials and publications

4.2.1.1. PubMed cohort

Primary reports published in December 2000 were sought as index studies to define the PubMed cohort. Although the PubMed search was first conducted in March 2001, subsequent searches revealed that studies published in December 2000 continued to be added to the database for more than one year post-publication. Results of these searches are shown in Table 4-1. The terms 'cross-over studies[mh]' and 'multicenter study[pt]' were added in September 2001. Forty-nine newly-indexed studies were retrieved in the PubMed search from September 2001 to March 2002, and an additional 34 citations were retrieved from March to July 2002. In July 2002, a final modification was made with the addition of 'controlled clinical trial[pt]' to the search strategy. No further searching for primary reports was carried out after July 2002.

Table 4-1. Results from repeated PubMed searches from March 2001 – July 2002

Date	Total citations retrieved	Search strategy
March 27, 2001	767	Original Cochrane strategy (Phase 1)
April 2, 2001	822	Original Cochrane strategy (Phase 1)
April 9, 2001	873	Original Cochrane strategy (Phase 1)
Sept. 25, 2001	1350	Original Cochrane strategy (Phase 1) plus 'cross-over studies[mh] OR multicenter study [pt]'
March 10, 2002	1399	Original Cochrane strategy (Phase 1) plus 'cross-over studies[mh] OR multicenter study [pt]'
July 17, 2002	1433	Original Cochrane strategy (Phase 1) plus 'cross-over studies[mh] OR multicenter study [pt]'
July 17, 2002	1542	Modified Cochrane strategy (Phase 1, including 'controlled clinical trial [pt]') plus 'cross--over studies[mh] OR multicenter study [pt]'

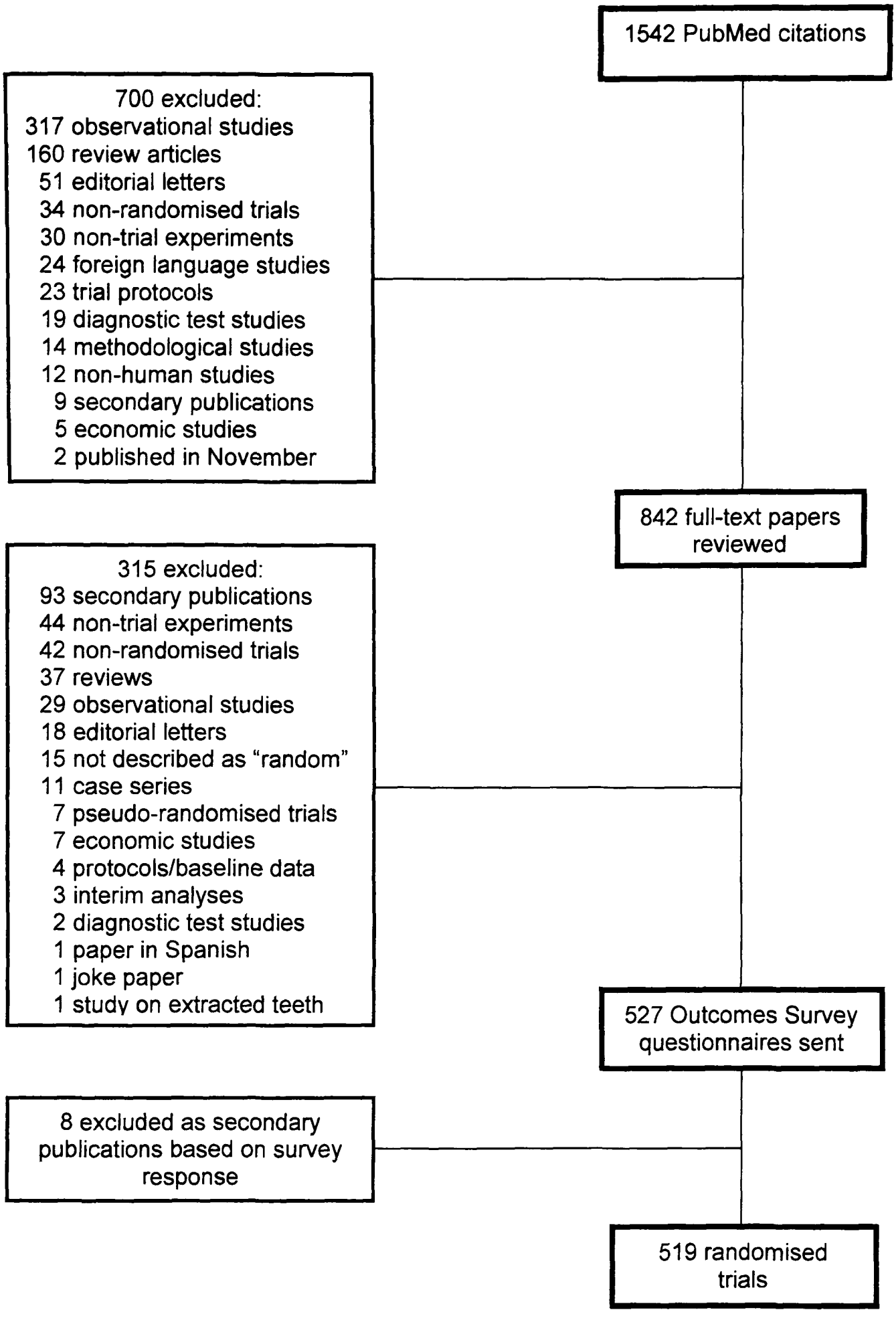


Figure 4-1. Identification of randomised trials whose primary reports were indexed on PubMed and published in December 2000 (PubMed cohort)

The flow of studies through the cohort selection process is shown in Figure 4-1. Of the 1542 citations retrieved from PubMed, 700 were excluded based on the abstract alone. The most common reason for exclusion was an inadmissible study design - primarily observational studies (n=317) and reviews (n=160). Thirty-four trials did not use random allocation, while 30 non-trial experiments did not assess healthcare interventions as their primary objective. Twenty-five trials (24 at this stage and 1 after receiving the full text paper) were excluded solely based on language, meaning that their full-text articles would have been reviewed had they been written in English or French. However, it is uncertain as to what proportion of these publications would have qualified for inclusion in the study cohort.

Of the 842 full-text articles reviewed, 315 did not meet the inclusion and exclusion criteria. 30% of these exclusions were secondary publications of randomised trials, while the remainder were mostly ineligible based on study design. Seven studies purported to be "randomised" but were subsequently found not to be properly randomised based on the description in the full report. Of 527 trials that were included in the Outcomes Survey, 8 were excluded as secondary publications after their survey responses revealed main reports published prior to December 2000.

Thirty-four post-December 2000 multiple publications for 29 trials were identified from survey responses (n=14 trials) and literature searches (n=15 trials). Of the 15 trials with post-December 2000 publications identified through literature searches, 11 were survey non-responders and 4 were responders who had stated that unreported outcomes existed. Two of the 4 responders had explained

that a second manuscript was to be submitted for publication to report additional outcomes.

The final PubMed cohort, therefore, consisted of 553 publications for 519 trials. 490 trials produced one final publication, 24 trials yielded two reports each, and 5 trials produced three publications each. 552 reports were published in English, while one was written in French. Five investigators were each listed as contact authors for 2 primary reports; contact authors for the remaining 509 primary publications appeared only once.

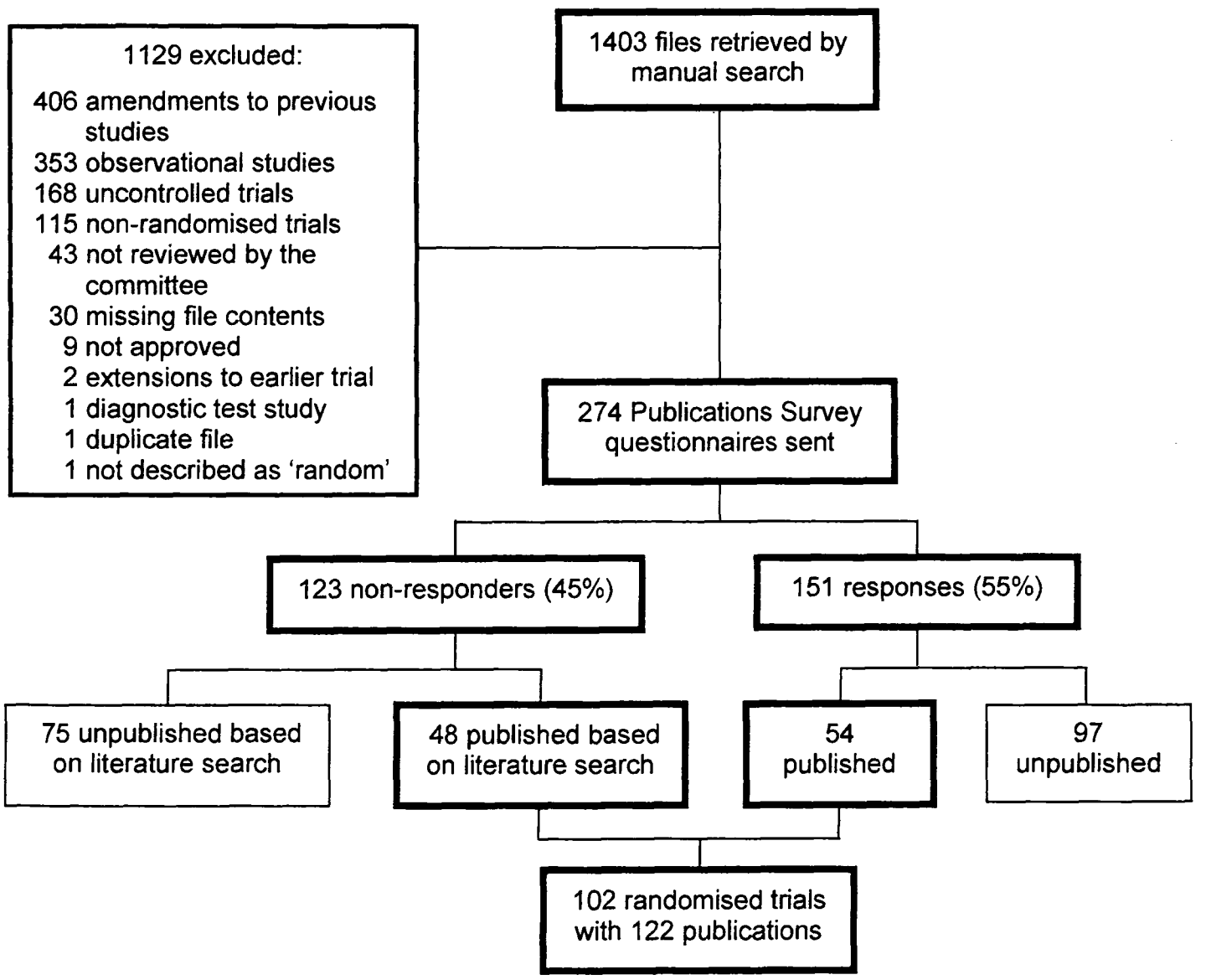


Figure 4-2. Identification of published randomised trials approved by the Copenhagen and Frederiksberg Research Ethics Committee from 1994-95 (Ethics cohort)

4.2.1.2. Ethics cohort

The process of identifying index trials and their publications for the Ethics sample is depicted in Figure 4-2. Manual searches retrieved 1403 files labelled as clinical trials submitted to the Copenhagen and Frederiksberg Research Ethics committee in 1994-95. 1129 were excluded as amendments to previous applications (n=406), observational studies (n=353), uncontrolled trials (n=168), and non-randomised trials (n=115).

Investigators for the remaining 274 trials were surveyed to identify publications (Figure 4-2). As described in Chapter 2, all trials also underwent literature searching for reports indexed on PubMed, EMBASE, PsychINFO, and the Cochrane Controlled Trials Register. 55% (151/274) of trialists responded to the survey to reveal 54 published studies. 82 of the remaining 97 responses provided reasons for not publishing (Table 4-2). The most common included not having started the trial (29%), low patient recruitment rates (22%), logistical difficulties (20%), and ongoing preparation of the manuscript (16%). Publications for 7 trials were identified by survey alone, 48 by literature search alone, and 47 by both. Of the 7 trials that were missed on literature search, 6 were published in PubMed-indexed journals.

The final cohort consisted of 102 trials with 122 publications, yielding a publication rate of 37%. 87 trials produced a single publication, while 12 trials each presented results in two reports. One study yielded three publications, and another 2 trials produced four publications each. Twenty-one protocols were submitted in English, and 81 were in Danish. Among trial reports, 117 were published in English, while 5 were in Danish. Four of the 5 Danish reports

presented the same information as their corresponding English publications. The remaining Danish report was the sole publication for that trial. 91 different investigators appeared as contact authors in the cohort. 83 published a single trial, while 6 were listed as contact authors for 2 trials. One trialist was the contact author for 3 trials, while another trialist published 4 separate studies. Danish researchers were listed as the contact authors for 62 (61%) trials.

Table 4-2. Reasons given by trialists for not publishing their studies in the CIHR and Ethics cohorts

Reason[a]	Number of trials (%)	
	Ethics cohort (n=82 trials providing data)	CIHR cohort (n=50 trials providing data)
Trial never started	24 (29%)	0
Low recruitment of participants	18 (22%)	6 (12%)
Logistical problems/No time	16 (20%)	3 (6%)
Manuscript under preparation	13 (16%)	18 (36%)
Results not statistically significant	7 (9%)	1 (2%)
Trial terminated early	4 (5%)	5 (10%)
Submitted for publication	3 (4%)	4 (8%)
Rejected for journal publication	2 (2%)	1 (2%)
Ongoing trial	2 (2%)	17 (34%)
Analysis not yet complete	0	1 (2%)

[a]Reasons are not mutually exclusive

4.2.1.3. CIHR cohort

The selection of CIHR trials is outlined in Figure 4-3. From the CIHR computer database of funding applications, 150 successful applications submitted to the Clinical Trials Unit from 1990-98 were identified. Upon review of the complete files, including protocols and applications, 45 studies were excluded due to ineligible study designs (n=23), lack of approval (n=9), original approval prior to 1990 (n=8), deletion from archives (n=3), and use of animal subjects (n=2).

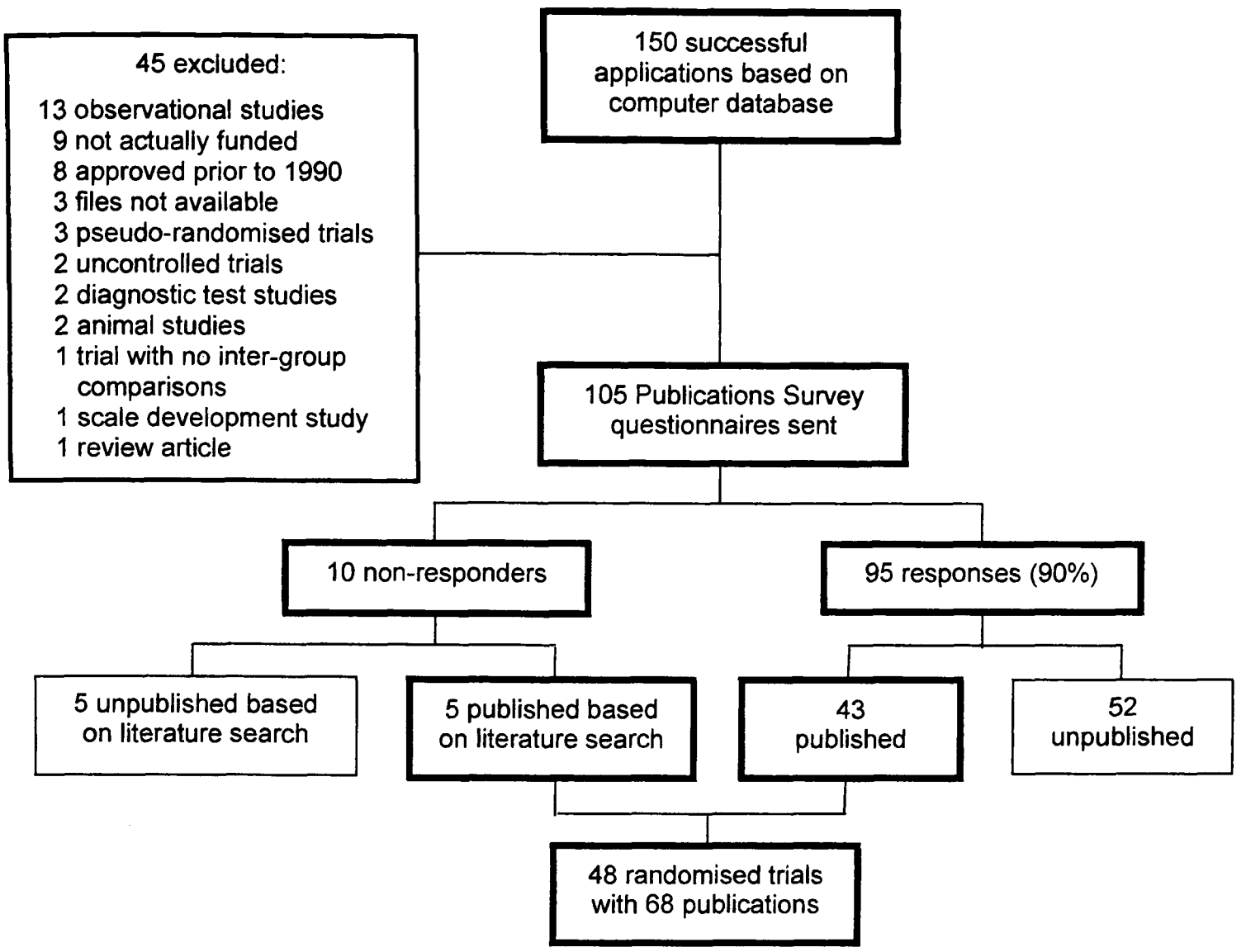


Figure 4-3. Identification of published randomised trials with successful funding applications submitted to the CIHR Clinical Trials Unit between 1990-98 (CIHR cohort)

Questionnaires for the Publications Survey were sent for the remaining 105 trials (Figure 4-3). 95 (90%) trialists responded to reveal 63 final publications for 43 trials. Unpublished trials (n=52) had often been completed (n=30), but several were ongoing (n=17) or terminated early (n=5). Reasons for non-publication are outlined in Table 4-2. Among reasons provided by 50 trials, the most common were ongoing manuscript preparation (36%) or an ongoing trial (34%). Literature searches yielded single publications for 5 of the 10 survey non-responders.

The final CIHR cohort consisted of 48 trials with 68 publications, representing a publication rate of 46%. 37 trials had produced a single publication, while 9 trials had presented results in two reports each. One of the remaining 2 trials had

yielded three publications, while the other had produced seven reports. 41 protocols were written in English, and 7 were in French. All publications appeared in English. 41 different researchers published trials in the CIHR cohort. 34 appeared as contact authors for one primary trial report, while 7 were responsible for two trials each.

4.3. Cohort characteristics

Overall trial characteristics are summarised for each cohort in Table 4-3. The same data stratified by study design are presented in Table 4-4. Data were mostly obtained from primary trial reports, supplemented by survey and protocol data. However, with regards to the quality of reporting, we did not use information from the protocol if it was missing from the publications, such as descriptions of power calculations or primary outcomes. The following descriptions of cohort characteristics will focus on parallel group and cross-over trials, as they constitute the vast majority of studies in the cohorts.

Table 4-3. Characteristics of randomised trials in PubMed, Ethics, and CIHR cohorts [Number (%) of studies in each trial cohort unless indicated otherwise]

	PubMed trials (n=519)	Ethics trials (n=102)	CIHR trials (n=48)
STUDY DESIGN			
Parallel group	383 (74%)	70 (69%)	39 (81%)
Cross-over	116 (22%)	30 (29%)	3 (6%)
Other	20 (4%)	2 (2%)	6 (13%)
SPECIALTY FIELD	Physiology 48 (9%)	Endocrinology 13 (13%)	Cardiology 10 (21%)
	Anaesthesiology 43 (8%)	Physiology 12 (12%)	Obstetrics/gynecology 8 (17%)
	Cardiology 41 (8%)	Oncology 7 (7%)	Surgery 7 (15%)
	Psychiatry 40 (8%)	Anaesthesiology 6 (6%)	Paediatrics 6 (13%)
	Paediatrics 37 (7%)	Paediatrics/Respirology 6 (6%) each	Critical care 4 (8%)
INTERVENTION			
Drug	393 (76%)	77 (75%)	27 (56%)
Surgery/Procedure	51 (10%)	11 (11%)	10 (21%)
Counselling/Lifestyle	55 (11%)	12 (12%)	8 (17%)
Equipment	20 (4%)	2 (2%)	3 (6%)
NUMBER OF STUDY GROUPS			
2	379 (73%)	68 (67%)	38 (79%)
3	85 (16%)	22 (22%)	3 (6%)
4	37 (7%)	8 (8%)	6 (13%)
> 4	18 (3%)	4 (4%)	1 (2%)
BLINDING			
Blinded	309 (60%)	67 (66%)	35 (73%)
None	166 (32%)	21 (21%)	12 (25%)
Unclear	44 (8%)	14 (14%)	1 (2%)
STUDY CENTRES			
Single	376 (72%)	53 (52%)	16 (33%)
Multiple	134 (26%)	49 (48%)	32 (67%)
Unclear	9 (2%)	0	0
FUNDING			
Full industry	167 (32%)	56 (55%)	0
Partial Industry	61 (12%)	17 (17%)	20 (42%)
Non-Industry	184 (35%)	22 (22%)	28 (58%)
None	54 (10%)	6 (6%)	0
Unknown	53 (10%)	1 (1%)	0

Table 4-4. Characteristics of randomised trials stratified by study design in PubMed, Ethics, and CIHR cohorts [Number (%) of studies in each trial cohort unless otherwise indicated]

	Parallel group trials			Cross-over trials			Other trial designs		
	PubMed cohort (n=383)	**Ethics cohort (n=70)**	**CIHR cohort (n=39)**	**PubMed cohort (n=116)**	**Ethics cohort (n=30)**	**CIHR cohort (n=3)**	**PubMed cohort (n=20)**	**Ethics cohort (n=2)**	**CIHR cohort (n=6)**
SPECIALTY FIELD	Anaesthesi-ology 40 (10%)	Endocrinology 11 (16%)	Obstetrics/ gynecology 8 (21%)	Physiology 38 (33%)	Physiology 11 (37%)	Haematology 1 (33%)	Dentistry 4 (20%)	Hematology 1 (50%)	Cardiology 5 (83%)
	Paediatrics 32 (8%)	Anaesthesiolo-gy 5 (7%)	Surgery 7 (18%)	Pharmacology 27 (23%)	Respirology 4 (13%)	Neurology 1 (33%)	Ophthalmology 4 (20%)	Oncology 1 (50%)	Paediatrics 1 (17%)
	Cardiology 31 (8%)	Cardiology 5 (7%)	Cardiology 5 (13%)	Cardiology 10 (9%)		Respirology 1 (33%)	Oncology 3 (15%)		
	Psychiatry 31 (8%)	Infectious diseases 5 (7%)	Paediatrics 5 (13%)	Endocrinology 7 (6%)			Paediatrics 2 (10%)		
	Gastroentero-logy 30 (8%)	Oncology 5 (7%)	Critical care 4 (10%)	Psychiatry 7 (6%)			Psychiatry 2 (10%)		
INTERVENTION									
Drug	278 (73%)	52 (74%)	19 (49%)	104 (90%)	23 (77%)	2 (67%)	11 (55%)	2 (100%)	6 (100%)
Surgery/Procedure	45 (12%)	9 (13%)	9 (23%)	4 (3%)	2 (7%)	1 (33%)	2 (10%)	0	0
Counselling/Lifestyle	45 (12%)	7 (10%)	8 (21%)	4 (3%)	5 (17%)	0	6 (30%)	0	0
Equipment	15 (4%)	2 (3%)	3 (8%)	4 (3%)	0	0	1 (5%)	0	0

	Parallel group trials			Cross-over trials			Other trial designs		
	PubMed cohort (n=383)	Ethics cohort (n=70)	CIHR cohort (n=39)	PubMed cohort (n=116)	Ethics cohort (n=30)	CIHR cohort (n=3)	PubMed cohort (n=20)	Ethics cohort (n=2)	CIHR cohort (n=6)
NUMBER OF STUDY GROUPS									
2	286 (75%)	47 (67%)	36 (92%)	81 (70%)	20 (67%)	2 (67%)	12 (60%)	1 (50%)	0
3	60 (16%)	15 (21%)	2 (5%)	21 (18%)	7 (23%)	1 (33%)	4 (20%)	0	0
4	25 (7%)	4 (6%)	1 (3%)	9 (8%)	3 (10%)	0	3 (15%)	1 (50%)	5 (83%)
> 4	12 (3%)	4 (6%)	0	5 (4%)	0	0	1 (5%)	0	1 (17%)
BLINDING									
Blinded	213 (56%)	49 (70%)	26 (67%)	86 (74%)	17 (57%)	3 (100%)	10 (50%)	1 (50%)	6 (100%)
None	132 (34%)	13 (19%)	12 (31%)	25 (22%)	7 (23%)	0	9 (45%)	1 (50%)	0
Unclear	38 (10%)	8 (11%)	1 (3%)	5 (4%)	6 (20%)	0	1 (5%)	0	0
STUDY CENTRES									
Single	254 (66%)	24 (34%)	13 (33%)	106 (91%)	28 (93%)	1 (33%)	16 (80%)	1 (50%)	2 (33%)
Multiple	122 (32%)	46 (66%)	26 (67%)	9 (8%)	2 (7%)	2 (67%)	3 (15%)	1 (50%)	4 (67%)
Unclear	7 (2%)	0	0	1 (1%)	0	0	1 (5%)	0	0
FUNDING									
Full industry	122 (32%)	45 (64%)	0	40 (34%)	10 (33%)	0	5 (25%)	1 (50%)	0
Partial Industry	39 (10%)	11 (16%)	13 (33%)	19 (16%)	6 (20%)	2 (67%)	3 (15%)	0	5 (83%)
Non-Industry	133 (35%)	10 (14%)	26 (67%)	41 (35%)	11 (37%)	1 (33%)	10 (50%)	1 (50%)	1 (17%)
None	45 (12%)	3 (4%)	0	8 (7%)	3 (10%)	0	1 (5%)	0	0
Unknown	44 (11%)	1 (1%)	0	8 (7%)	0	0	1 (5%)	0	0

4.3.1. Study topic

For parallel group trials, the most common specialty areas varied greatly between cohorts and study designs (Table 4-4). The only field that consistently ranked in the five most studied specialties was cardiology, followed by paediatrics. Cross-over trials were most frequently physiological and pharmacological studies.

4.3.2. Study design

In the PubMed cohort, 74% of studies were of parallel group design and 22% were of cross-over design (Table 4-3). The remainder were split-body (n=9), cluster (n=6), factorial (n=4), and 'n of 1' (n=1) trials. 69% and 29% of Ethics trials were of parallel group and cross-over designs respectively, as were 81% and 6% of CIHR trials. The other studies consisted of factorial (n=6 CIHR and 1 Ethics trial) and split-body (n=1 Ethics trial) designs.

20/519 (4%) PubMed trials (14 parallel group, 5 cross-over, 1 'n of 1' study) were randomised pilot or Phase 1 studies. One parallel group CIHR trial was a pilot study, as were two Ethics trials (1 parallel group and 1 cross-over trial).

In the PubMed and Ethics samples, more than 70% of parallel group trials investigated drugs as the primary intervention of interest, while 12-13% evaluated procedures or surgical interventions (Table 4-4). Counselling and lifestyle interventions were examined in approximately 10% of parallel group trials. The majority of cross-over studies evaluated drug interventions. CIHR parallel group trials, on the other hand, investigated drugs less frequently

(49%) and examined more surgical/procedural (23%) and counselling/lifestyle (21%) interventions.

While most trials randomised participants to 2 groups, a significant proportion (21%-33% in each cohort) examined 3 or more comparison groups (Table 4-3). Among parallel group studies, a larger percentage of CIHR trials (92%) randomised participants to 2 groups compared to PubMed (75%) and Ethics (67%) cohorts (Table 4-4). Proportions of two-arm cross-over trials were similar across cohorts (67-70%).

According to explicit text descriptions in protocols or publications, 56-70% of parallel group trials were characterised as blinded to some extent (Table 4-4). No masking was reportedly used for 1 in 3 parallel group trials in the PubMed and CIHR cohorts, compared to 1 in 5 for the Ethics cohort. Blinding among cross-over trials was used in 74% and 57% of PubMed and Ethics trials respectively, while all 3 CIHR cross-over trials were blinded.

Authors of the 34 trial reports that lacked information on the number of study sites were contacted for the PubMed cohort, and 25 (74%) responded with further information. Relatively large differences in the proportion of multicentre studies were observed between the 3 study cohorts (Table 4-3). Only a quarter of PubMed trials collected data from multiple study centres, compared to one half of Ethics trials and two thirds of CIHR trials.

Table 4-5. Sample sizes observed in three trial cohorts including and excluding physiological and pharmacokinetic trials

Trial design	INCLUDING PHYSIOLOGICAL AND PHARMACOKINETIC TRIALS			EXCLUDING PHYSIOLOGICAL AND PHARMACOKINETIC TRIALS		
	Number of trials	Median per trial [IPR$_{80}$][a]	Median per study group [IPR$_{80}$][a]	Number of trials	Median per trial [IPR$_{80}$][a]	Median per study group [IPR$_{80}$][a]
ALL TRIALS						
PubMed cohort	519	52 [12-310]	32 [12-159] n=393 trials[b]	438	64 [19-329]	34 [12-160] n=377 trials[b]
Ethics cohort	102	70 [12-624]	73 [15-299] n=71 trials[b]	89	92 [18-812]	73 [15-324] n=70 trials[b]
CIHR cohort	48	299 [61-2568]	137 [32-1072] n=45 trials[b]	N/A	N/A	N/A
PARALLEL GROUP						
PubMed cohort	383	80 [25-369]	32 [12-159]	368	85 [26-384]	34 [12-160]
Ethics cohort	70	151 [28-935]	66 [13-324]	69	155 [29-1000]	73 [15-350]
CIHR cohort	39	281 [48-2143]	141 [24-1072]	N/A	N/A	N/A
CROSS-OVER						
PubMed cohort	116	15 [8-38]	N/A	51	20 [10-44]	N/A
Ethics cohort	30	16 [7-43]	N/A	18	20 [7-75]	N/A
CIHR cohort	3	62 [61-129]	N/A	N/A	N/A	N/A
OTHER						
PubMed cohort	20	55 [12-556]	17 [5-202] n=10 trials[c]	19	52 [10-846]	20 [5-282] n=9 trials[c]
Ethics cohort	2	411 [10-812]	203 [N/A] n=1 trial[c]	2	411 [10-812]	203 [N/A] n=1 trial[c]
CIHR cohort	6	596 [275-9541]	119 [69-2385]	N/A	N/A	N/A

[a]Inner-80-percentile range (10^{th}-90^{th} percentile)
[b]Excluding cross-over, split-body, and 'n of 1' trials
[c]Excluding split-body and 'n of 1' trials

4.3.3. Sample size

The median sample sizes reported in trial publications varied greatly between the three study cohorts (Table 4-5). Among parallel group studies, PubMed trials randomised a median of 80 participants, with 32 per treatment group. On average, almost twice as many participants were recruited to parallel group Ethics trials (median 151 overall; 66 per group), and more than three times as many for parallel group CIHR trials (median 281 overall; 141 per

group). Cross-over trials randomised a median of 15-16 participants in the PubMed and Ethics cohorts. The average number of participants per group was not calculated for cross-over or split-body trials.

Excluding physiological and pharmacological studies produced slight increases in the sample sizes observed in the PubMed and Ethics cohorts (Table 4-5). No such studies existed in the CIHR cohort.

4.3.4. Funding sources

Publications for 177 PubMed trials lacked information about funding sources; trialists for 124 (70%) of these studies provided further information in survey responses. In the Ethics cohort, 16 primary trial reports were missing funding details, but protocols provided the necessary data for 15 of the studies. The remaining study did not have additional funding information from its protocol or questionnaire. Information about funding sources was thus missing for 53 PubMed trials, 1 Ethics trial, and no CIHR trials.

Over 30% of PubMed trials received only industry support with a further 12% reporting partial support, while 35% had only non-industry sources of funding (Table 4-3). In comparison, trials in the Ethics cohort were more frequently sponsored with full industry support (55%). Non-industry sponsorship was observed for 35% and 22% of PubMed and Ethics trials respectively. By the nature of the sample, the CIHR cohort contained no trials that were fully funded by commercial sources; 58% were supported by non-industry sources only.

4.3.5. Journal characteristics and quality of reporting

Primary trial reports in the PubMed cohort were published in 271 different journals, with only 13 journals publishing more than five trials in the single month studied. 89 and 20 different journals published trials from the Ethics and CIHR cohorts respectively. The vast majority (>90%) of PubMed and Ethics reports were published in specialty journals, compared to only 44% for the CIHR cohort (Table 4-6). Fewer than 4% of PubMed and Ethics studies were published as brief reports or letters; all CIHR trials were reported in full-length articles.

A power calculation was reported in the publications of 27% of PubMed and 37% of Ethics trials (Table 4-6). A much higher proportion (75%) was observed in CIHR trials. Primary outcomes were defined in publications of 45% and 62% of PubMed and Ethics trials respectively, compared to over 90% of CIHR trials.

The reporting of methods of random sequence generation and allocation concealment in the PubMed cohort was inadequate or unclear in approximately 80% of primary publications (Table 4-7). 66% of trials did not adequately detail losses to follow-up and exclusions. Reporting of methodological details was generally worse for cross-over studies than parallel group trials.

Table 4-6. Characteristics of publications in PubMed, Ethics, and CIHR cohorts (stratified by study design[a])

	All trials			Parallel group trials			Cross-over trials		
	PubMed cohort (n=519)	Ethics cohort (n=102)	CIHR cohort (n=48)	PubMed cohort (n=383)	Ethics cohort (n=70)	CIHR cohort (n=39)	PubMed cohort (n=116)	Ethics cohort (n=30)	CIHR cohort (n=3)
JOURNAL TYPE									
General	37 (7%)	3 (3%)	21 (44%)	31 (8%)	3 (4%)	17 (44%)	4 (3%)	0	1 (33%)
Specialty	482 (93%)	99 (97%)	27 (56%)	352 (92%)	67 (96%)	22 (56%)	112 (97%)	30 (100%)	2 (67%)
REPORT TYPE									
Short/Letter	17 (3%)	3 (3%)	0	11 (3%)	3 (4%)	0	10 (9%)	0	0
Full length	502 (97%)	99 (97%)	48 (100%)	372 (97%)	67 (96%)	39 (100%)	106 (91%)	30 (100%)	3 (100%)
POWER CALCULATION									
Reported	142 (27%)	38 (37%)	36 (75%)	122 (32%)	32 (46%)	31 (79%)	16 (14%)	5 (17%)	1 (33%)
Not reported	377 (73%)	64 (63%)	12 (25%)	261 (68%)	38 (54%)	8 (21%)	100 (86%)	25 (83%)	2 (67%)
PRIMARY OUTCOME									
Defined	232 (45%)	63 (62%)	45 (94%)	189 (49%)	47 (67%)	36 (92%)	36 (31%)	15 (50%)	3 (100%)
Not defined	287 (55%)	39 (38%)	3 (6%)	194 (51%)	23 (33%)	3 (8%)	80 (69%)	15 (50%)	0

[a]Study designs other than parallel group and cross-over trials are not shown due to small numbers

Table 4-7. Quality of reporting in primary publications of 519 randomised trials in the PubMed cohort

	Trial Design			
	All (n=519)	**Parallel group (n=383)**	**Cross-over (n=116)**	**Other (n=20)**
METHOD OF RANDOM SEQUENCE GENERATION				
Adequate[a]	109 (21%)	91 (24%)	11 (9%)	7 (35%)
Unclear	410 (79%)	292 (76%)	105 (91%)	13 (65%)
METHOD OF ALLOCATION CONCEALMENT				
Adequate[a]	97 (19%)	87 (23%)	7 (6%)	3 (15%)
Inadequate	2 (0.4%)	2 (1%)	0	0
Unclear	420 (81%)	294 (77%)	109 (94%)	17 (85%)
HANDLING OF ATTRITION				
Adequate[a]	174 (34%)	120 (31%)	45 (39%)	9 (45%)
Inadequate	150 (29%)	128 (33%)	17 (15%)	5 (25%)
Unclear	195 (38%)	135 (35%)	54 (47%)	6 (30%)

[a]Defined in Appendix 8

4.4. Discussion

The characteristics of three cohorts of randomised trials were examined: trials indexed on PubMed, published trials funded by a Canadian government research agency, and published trials approved by a Danish ethics committee. The descriptive findings in this chapter provide a cross-sectional epidemiological overview of trials from the populations represented by each cohort. Because the samples were comprehensive - including all trials within a given time period - they can be viewed as representative of the original source from which they were drawn. For the PubMed and Ethics cohorts, this would translate into relatively unrestricted samples, while the CIHR trials were selected through a rigorous peer review process. Clear differences in the

characteristics of study design and reporting were seen when comparing the individual trial cohorts. As expected, the largest contrast was between the CIHR trials and the other two cohorts.

4.4.1. Epidemiology of recent randomised trials indexed on PubMed

Of the three cohorts, the PubMed trials constituted the largest sample and the most representative of the current literature. Aside from the inherent limitations of the database itself, which will be discussed later in the chapter, the PubMed cohort did not exclude particular study designs, specialties, or journal types. The cohort's characteristics thus provide an overview of the epidemiology of recently published randomised trials.

Most previous reviews of research studies have examined samples restricted by journal or topic, with many considering only trials published in leading general medical journals. Two exceptions include a random sample of 113 MEDLINE-indexed controlled trials published in 1980;[96] and a 10% MEDLINE sample of non-cross-over randomised trials comparing two active treatments in 1976, 1981, 1986 and 1991.[30] These two studies thus provide a basis for comparison with the PubMed cohort.

Several observations about the PubMed cohort merit further discussion. 1% of the 842 full-text articles reviewed were non-randomised trials that claimed to be randomised. The mis-labelling was identified based on the text description of non-randomised participant allocation, but a subset of the 410 trials that did not present any details about the allocation procedure were likely to have been non-randomised as well. Examples of this type of mis-

labelling have been observed previously.[130] Their occurrence is problematic because the term 'randomised' conveys a false perception of the trial's quality.

Over a quarter of trials studied more than two randomised groups, while more than 10% had at least 4 treatment groups. These proportions are consistent with previous observations.[96] Given their high prevalence, the conduct and reporting of multi-arm studies warrant further examination.

One third of trials in each sample were unblinded, compared to only 15% in Meinert *et al.*'s sample from 1980. This difference may reflect an actual deterioration in methodological rigour, or an increase in the proportion of trials examining non-drug interventions that are less amenable to blinding. While masking is not always possible, a lack of masking has been shown to inflate treatment effect sizes.[131] Blinding of study investigators and outcome assessors is protective against wish bias,[132] where the clinician's preference for a new therapy may result in overestimation of its efficacy.

The assignment of blinding status to trials in our cohorts was based on descriptive terms used in the reports themselves. Because the terms 'single-', 'double-', and 'triple-blind' are each ambiguous with a variety of interpretations, the characteristic was dichotomised to include any blinding or no blinding.[133] In order to avoid ambiguity, it has been recommended that reports explicitly state which of the individuals involved in the study were blinded rather than using the traditional terminology.[133,134]

The small sample sizes observed in the PubMed cohort are concerning. Little improvement can be seen over the past two decades relative to the median sample size per study group of 19 observed in 1980,[96] or 23-39 observed from 1976-1991.[30] It can be postulated that the inclusion of physiological and pharmacological trials may have resulted in lower average sample sizes. However, when trials measuring primarily physiological or pharmacokinetic outcomes were excluded, median sample sizes remained low.

Using the median sample size of 32 participants per group in the PubMed cohort's parallel arm trials, a two-group comparison has only 11% power to detect a difference between event rates of 10% and 20% at the 0.05 significance level. Trials with inadequate power have a high false negative error rate, have an increased risk of spurious positive results, and are implicated as a source of publication bias.[6] These problems have prompted some individuals to label inadequately-powered trials as unethical.[135]

The apparent lack of progress in sample size should be interpreted with caution, as the samples being compared differ in other potentially confounding characteristics. MEDLINE has expanded its journal coverage by over 50% since 1980, either due to an increase in comprehensiveness of coverage or in the number of journals in print, or both. This may partly account for the lack of improvement in recent sample sizes, as smaller trials in lesser-known journals are now more likely to be identified. However, a higher percentage of multicentre trials was observed in the PubMed cohort compared to Meinert *et al.*'s MEDLINE sample from 1980 (26% versus 13%), as well as a lower proportion of cross-over trials (22% versus 66%).[96] These confounders

should act in the opposite direction to favour increased sample sizes in the PubMed cohort, suggesting that the deficiencies observed are actually worse than they appear.

With regards to characteristics of trial publications, it is apparent from the PubMed cohort that an enormous number of journals currently publish randomised trials, with no single journal accounting for more than 3% of trials published in the entire month. Only 7% of the trials were published in general medical journals. When compared to a sample of 166 clinical journals with the five highest impact factors in each of over 30 medical specialties (Douglas Altman, personal communication, 2003), 78% (212/271) of journals in the PubMed cohort were not included in the high impact sample. It is clear, therefore, that conclusions drawn from the many methodological reviews that are limited to high impact journals may well not be generalisable to the vast majority of published trials. Restricted samples, however, are customary in reviews of published trials - presumably due to practical constraints on the number of studies able to be reviewed. This may be acceptable if the aim is to focus on a single specialty.

Although the vast majority of PubMed trials were published as full-length reports, the reporting of important methodological characteristics remains poor. The revised CONSORT statement[134] for parallel group trials states that authors should report "clearly defined primary and secondary outcome measures," as well as details of power calculations, random sequence generation and allocation concealment. However, the majority of trials (55%) in the PubMed cohort failed to specify their primary outcomes, despite our use

of liberal criteria to assess whether such outcomes were defined. While disappointing, this figure is better than the prevalence of 27% observed in three leading general medical journals in 1985.[92] Only 6% of non-steroidal anti-inflammatory drug trials in rheumatoid arthritis defined primary outcomes in their reports.[97] Without explicitly-stated primary outcomes that are defined *a priori*, trialists have free rein to selectively report outcomes depending on results of significance tests rather than on pre-defined clinical importance.

The reporting of power calculations was even worse, with almost three-quarters of trials failing to mention one. This figure is consistent with a review of trials published in leading general medical journals in 1994, which found that 68% of negative trials failed to mention a power calculation.[31] More recent reviews have found higher rates of reporting power calculations in cohorts restricted to general medical[136] and specialty[137-140] journals, although the rates of reporting remain low in some specialties.[141] Reporting of an *a priori* power calculation can act as a surrogate indicator of adequate trial planning, can help to identify the primary outcome, and can indicate whether trials were terminated earlier than planned.[142]

Reporting of the methods of random sequence generation, methods of allocation concealment, and details of attrition were also poor. Inadequate reporting of allocation concealment has been shown to inflate pooled effect estimates by 17% overall.[131] However, a recent survey of 50 rheumatology trialists (response rate 80%) found that the assessment of quality markers based on publications often does not agree with actual trial conduct.[117] Although it was observed that the methods of random sequence generation

and allocation concealment were not described by 29 and 32 publications respectively, over 75% were found to have been performed adequately in practice based on survey responses. All methods identified as inadequate by survey were also deemed to be unclear from the trial reports. Assuming that the survey responses were accurate, the reliance on published reports to assess trial quality appears to be a sensitive but non-specific method, and may result in significant misclassifications.

Reporting of methodological details appears to be generally worse in cross-over studies compared to parallel group trials. Cross-over trials were also much smaller on average and more likely to be physiological studies. Certain aspects of trial methodology that are important to describe for parallel group trials may not be as essential to report for cross-over designs. For example, bias associated with inadequate methods of random sequence generation and allocation concealment may have less significance in cross-over trials because study participants will eventually be exposed to all intervention groups. However, adequate reporting of primary outcomes, power calculations, and attrition certainly remain important across study designs.

4.4.1.1. Summary

The PubMed cohort highlights several interesting features and deficiencies in the overall population of recently-published randomised trials. Sample sizes remain disappointingly low, facilitating the dissemination of false negative results and publication bias. Much improvement also remains to be made in the reporting of important methodological details. Of particular relevance to outcome reporting is the need for trialists to define their primary outcomes

explicitly in publications. The introduction of the revised CONSORT statement in 2001 and its adoption by journals is a positive step towards addressing the reporting deficiencies identified.[134]

4.4.2. Response rates to Publications Surveys

While a response rate of 90% was achieved in the CIHR cohort, a lower proportion (55%) of Ethics trialists responded. The discrepancy may be due to the use of frequent e-mail and telephone reminders in the CIHR cohort. The higher CIHR response rate may also reflect the greater perceived authority of a national funding agency compared to a regional ethics review committee.

The response rates are difficult to generalise to systematic reviews because the questionnaires were sent on behalf of funding or ethics bodies who had approved the trials. It is unclear whether similar response rates would be expected for requests sent by reviewers rather than institutions with authority. A randomised trial has suggested that a well-known signatory does not significantly increase response rates to postal questionnaires soliciting information about methodological details.[143] However, a previous attempt to identify unpublished trials retrospectively had relatively low yield.[144]

4.4.3. Generalisability of study cohorts

The three study cohorts were defined to enable an assessment of outcome reporting practices in broad, comprehensive samples of trials. It is therefore important to evaluate whether conclusions drawn from these samples will be generalisable to the overall population of trials in the medical literature.

4.4.3.1. PubMed cohort

The PubMed sample was unrestricted by study design, topic, or journal impact factor, and is representative of the current population of randomised trials published in journals covered by PubMed. A major advantage of the study cohort is its breadth of coverage, spanning specialty fields and journal types from over 70 countries.

However, despite encompassing MEDLINE's coverage of over 4600 journals, the PubMed database may not be representative of the whole literature.[145] It is likely that systematic differences exist between PubMed and non-PubMed indexed trial reports. The MEDLINE database is known to under-represent European journals.[145] Also, it was noted in 1995 that only 2% of journals indexed in major databases were from developing countries.[146] This may, however, reflect the actual proportion of research output from developing regions.

A review of 66 meta-analyses that included trials from both MEDLINE and non-MEDLINE journals demonstrated that non-MEDLINE indexed trials were more likely to be smaller, to be published earlier, to evaluate complementary medicine, and to report statistically significant results.[5] Overall, non-MEDLINE trials did not produce significantly different effect estimates than MEDLINE-indexed trials within the same meta-analyses.

Trials published in PubMed-indexed journals therefore tend to be larger, of higher methodological quality, and less likely to have significant results compared to non-PubMed indexed trials. This may result in an

underestimation of reporting deficiencies relative to the overall population of trials published in any journal. In addition, the exclusion of trials published in non-PubMed journals may not be important, as they generally have little impact on effect estimates obtained from meta-analyses. Finally, it should be noted that there are major challenges in studying non-MEDLINE journals systematically, particularly in identifying a comprehensive sample.

4.4.3.2. Ethics cohort

The publication rate of 37% in the Ethics cohort was similar to that observed in a sample of 154 trials approved by an ethics committee in Spain (31%).[147] The time period elapsed since approval was 2-3 years longer for the Danish cohort. Although the Ethics cohort was restricted primarily by study location, only 61% of the contact authors were based in Denmark, indicating that many trials were led by investigators in other countries. For single centre studies, there is little reason to believe that research activities conducted in Denmark would differ systematically from those conducted in other developed countries. However, within a particular country, there might be a systematic difference between trials conducted in major teaching hospitals, such as in Copenhagen, and in smaller centres.

Differences between the Ethics and PubMed cohorts may be attributed to PubMed trials that were carried out in developing countries. On average, trials and publications from the two cohorts were similar in most respects with the exception of funding sources, study sites, and sample size. The higher frequency of full industry sponsorship and multiple study sites among parallel group Ethics trials relative to the PubMed cohort suggests that Danish sites

participate more often as part of large, multicentre, industry-funded trials. Consistent with these differences was a larger observed median sample size compared to PubMed trials. Sampling an ethics committee would be expected to produce an over-representation of multicentre trials, as these must be considered by multiple ethics committees.

The Ethics cohort therefore represents a broad sample of published trials conducted in developed countries. It is likely to be slightly less representative of the overall published trial population compared to the PubMed cohort, but more generalisable than the CIHR cohort.

4.4.3.3. CIHR cohort

Published CIHR trials were generally more likely to have parallel study groups (including factorial trials), to investigate non-drug interventions, to recruit more participants from multiple study sites, and to have no industry involvement in funding. They were also much more likely to be published in a general medical journal, to define primary outcomes, and to report a power calculation. In terms of both methodology and reporting, the CIHR cohort clearly represents a higher standard of trials relative to the PubMed and Ethics samples, as well as to the population of published trials overall.

While parallel group trials appeared with similar frequency in the three cohorts, cross-over trials were less common in the CIHR sample. This discrepancy may reflect a tendency for large, multicentre studies to be of parallel group design.

CIHR studies examined more non-drug interventions than trials in the other two cohorts. One possible explanation is that drug trials are more likely to be funded by industry manufacturers, who have a clear interest in the investment. Non-drug interventions, on the other hand, must rely more heavily on non-industry sponsors as the interventions often do not present commercial opportunities. Consistent with this hypothesis is the observation that CIHR trials were more likely to be funded by non-commercial sources than those in the other study samples. On average, CIHR trials also had 4-6 times larger sample sizes than PubMed and Ethics trials, and were more likely to involve multiple study sites.

The CIHR cohort therefore represents trials of a high methodological standard, with superior reporting of important methodological details. It may be expected that deficiencies and bias in outcome reporting would be less prevalent in this population of trials.

4.4.4. Sampling biases relevant to outcome reporting

The PubMed, Ethics, and CIHR cohorts were clearly defined with specific inclusion and exclusion criteria. However, the final sample of trials may not be representative of the defined cohorts if the methods used to identify potential trials for inclusion were biased to systematically overlook particular types of studies. Of greatest concern would be an over-estimation of the true magnitude of outcome reporting bias using the three study samples. It is therefore important to examine sampling biases and their potential effect on the results of this thesis.

4.4.4.1. Factors applicable to all cohorts

Time period of sampling

For the PubMed cohort, it is unlikely that studies published in December differ systematically from those published in other months during the year 2000. Journals that are published less frequently than monthly would not have necessarily been systematically excluded because not all journals start their annual cycle in January. However, certain journals may have a 'fun' issue during the December holiday season, such as the *British Medical Journal*, which may have decreased the number of trials in the sample relative to other months.

For the Ethics and CIHR cohorts, the primary reasons for the choice of years were to maximise the sample size while enabling sufficient time for trials to be published. There is little reason to believe that the years chosen would be systematically different from other time periods of comparable duration in the 1990s.

Exclusions and missing files

Trials with primary reports published prior to the year 2000 were excluded from the PubMed cohort because the goal was to define a cohort of incident trials published for the first time in 2000. Such trials were identified from the text of reports, as well as Outcomes Survey responses that revealed earlier publications. However, some trials with publications earlier than 2000 may have been missed if the trial reports did not allude to prior publications, if the trialists did not respond to the survey, and if the pre-2000 reports were not identified by literature searches. This would have potentially inflated the

number of incident trials published in December 2000, but should not have biased the study results. Also, the occurrence of such ineligible trials would have presumably been uncommon.

Studies with primarily economic outcomes were excluded from all cohorts to focus the investigation on clinical and laboratory outcomes. Also, it would have been difficult to define the levels of outcome reporting with respect to meta-analysis of economic data. Evaluations of diagnostic test properties were excluded for similar reasons. Clearly the results of this thesis are not generalisable to these two types of trials. Finally, interim reports were excluded because the outcome data would be considered preliminary and incomplete.

Other sources of trial data for systematic reviews have also been excluded from the study cohorts, such as conference proceedings and abstracts. These would be expected to have greater outcome reporting deficiencies than full publications due to space limitations, and any systematic effect of their exclusion would therefore tend to act in a conservative manner. For these reasons, some would argue against the inclusion of these types of reports in systematic reviews.[148]

Publications reporting analyses for sub-groups of the original trial population, as well as associations or correlations of baseline risk factors were also excluded in order to focus on comparisons between the original randomised groups. It is, however, recognised that selective reporting of these results likely exists and warrants further study.

For the Ethics and CIHR trials, 30/1403 (2%) and 3/150 (2%) protocols were missing for each cohort respectively. It is likely safe to assume that they were missing at random. Two of the CIHR files had already been destroyed from archives after the standard storage period, while most of the missing Ethics files were in a single box that could not be located. Some of these studies may not have been eligible for inclusion in the study cohorts. Considering the low number of files involved and the low risk of an association between missing files and outcome reporting, their absence should not bias the results produced from the trial cohorts.

4.4.4.2. Factors specific to the PubMed cohort

Despite avoiding the commonly-used restrictions to particular journal types or specialty areas, the PubMed cohort was limited by the search strategy used, the timing of PubMed indexing, the initial exclusion of some citations based on abstract alone, and the language in which publications were written.

Search strategy

It was important to identify all randomised trials published in December 2000 and indexed on PubMed. The search strategy used was derived from a highly sensitive strategy proposed for the Cochrane Collaboration in 1994 and revised in 2002.[114] Additional terms were added to increase the search sensitivity: 'cross-over studies[mh]' and 'multicenter study[pt]'. Therefore, the modified search strategy should have retrieved the vast majority of trials indexed on PubMed.

Timing of trial indexing on PubMed

The time of indexing of trial reports may constitute an additional sampling bias specific to PubMed. It was noted that studies were continually being added to the database, even at the final search 18 months post-publication (Table 4-1). Studies indexed after this date would have therefore been excluded from the cohort. It is probable that studies published in journals with a lower impact factor and smaller readership would have been reviewed and indexed later than those published in prominent journals. The PubMed cohort may thus be representative of a higher standard of trials. Foreign language journals may also have taken longer to evaluate due to time spent on translation, although such trials would have been excluded from the sample regardless based on language.

A PubMed search conducted in May 2003 retrieved 1590 citations, an increase of 48 from the final search used to define the PubMed cohort in July 2002. It is unclear how many of these would be eligible trials, but assuming the same yield as the pre-July 2002 citations, approximately a third may have been eligible for inclusion. It is unlikely that these few trials would have significant biased our results.

Exclusion of studies based on abstract

The first stage of identifying eligible trials from citations retrieved from PubMed involved a review of available abstracts. Studies were excluded if it was obvious that they were not eligible for inclusion. Trials were not excluded if any doubt existed as to their study design, even when PubMed categorised them as non-trials, such as 'letters' or 'reviews'. It is possible that some

citations were wrongly excluded if the abstract was not representative of the actual study design. However, this is unlikely to have been a common occurrence.

Language

The restriction of the PubMed cohort to English and French reports was imposed for practical reasons, namely insufficient resources for translation. Twenty-four potential trials identified in the PubMed search were not in these two languages, and their full-text papers could therefore not be reviewed. It may be postulated that important, high quality trials tend to be published in English language journals for the widest dissemination of results, such that exclusion of up to 25 foreign language trials will not have had a significant impact on results.

A comparison of 133 trials published in English-language journals to 96 trials published in French, German, Italian, and Spanish medical journals revealed that non-English trial reports were less likely to define primary outcomes or to justify their sample size.[149] In a retrospective empirical analysis of language bias in meta-analyses, Jüni *et al.* (2002)[150] found that non-English language trials were of lower methodological quality than trials published in English, and concluded that they generally had little effect on pooled estimates of treatment effect. Egger *et al.* (1997)[151] observed that among 62 pairs of English and German trial publications with the same authors, studies were more likely to be published in local German language journals as opposed to international English journals if their results were negative. Study quality was comparable between the trial reports in each pair. Conversely, in a separate review of 50

meta-analyses, non-English trials overall had a 16% (95% CI 3% to 26%) larger pooled effect estimate compared to trials published in English.[5] The effects of including non-English trial reports within individual meta-analyses varied widely, ranging from an increase of 42% to a decrease of 23% in pooled effect estimates. The overall evidence suggests that trials published in non-English journals are of similar or inferior quality, with the potential to slightly inflate pooled effect estimates in meta-analyses.

By excluding non-English publications, the PubMed cohort may therefore represent trials of higher quality with fewer statistically significant results. This bias may have an impact on outcome reporting if higher quality trials have better reporting practices, although fewer significant results may reduce the number of potential outcomes to selectively report based on statistical significance. The systematic effect overall would have likely acted in a conservative direction, although any effect would have been minor due to the small number of non-English publications retrieved by our PubMed search.

Summary

Considering the direction of operation for all sampling biases together, the overall effect - if any - would likely be to underestimate the true extent of biased trial reporting in the PubMed cohort. Certainly at the very minimum, more important trials would not have been missed. The cohort can therefore be considered to be broadly representative of all PubMed-indexed randomised trials published in the year 2000 with clinical and laboratory outcomes.

4.4.4.3. Factors specific to CIHR and Ethics cohorts

Identification of publications

The CIHR and Ethics cohorts relied on the Publications Survey and literature searches to identify published reports for the corresponding protocols. Response and recall bias in the survey may have prevented the identification of publications for particular types of trials. Trialists may have been less likely to reply if their study was unpublished, or may not have remembered the study in question, particularly if their participation was limited as one of many sites in a multicentre study. In one instance, a trialist in the Ethics cohort stated that his trials had never been completed, despite publications being identified by literature search and his name being listed as one of the participating investigators. Because literature searches were conducted for all protocols in the Ethics cohort, such inaccuracies in survey responses were largely avoided. For the CIHR cohort, literature searches were conducted only for non-responders, as it was felt that such large trials would not be mistakenly described as unpublished by trialists who would have an interest in reporting all research output to a funding agency.

Publications for non-responders may not have been identified in either cohort if the journal was not indexed on PubMed, EMBASE, or the Cochrane Controlled Trials Register. However, such trials would presumably have been rare. Of 151 survey responders in the Ethics cohort, 7 (5%) revealed publications that were not retrieved by literature searching. The literature searching was thus comprehensive and would not have missed a significant number of publications.

As discussed previously, trials published in journals that are not indexed on any of the major databases searched may be of lesser quality. If they were to bias the assessment of outcome reporting, the impact of their exclusion would likely be in a conservative direction.

4.5. Chapter summary

The three study cohorts represent distinct populations of published randomised trials. The PubMed cohort is the most comprehensive in its coverage, and is restricted primarily by the limitations inherent to the database. The overall characteristics of PubMed trials represent a cross-sectional view of the epidemiology of recent randomised trials, and several deficiencies in trial design and reporting were noted. The Ethics cohort may be viewed as representative of the population of trials conducted in developed countries, while the CIHR cohort represents a high standard of trials whose protocols were subjected to rigorous methodological and clinical review. Potential sampling biases exist for all three cohorts, but their overall effect - if any - on the evaluation of outcome reporting would likely be small and act in a conservative manner to under-estimate the extent of reporting deficiencies.

Chapter 5 - Completeness of outcome reporting in randomised trials

5.1. Introduction

The previous chapter provided an overview of the three study cohorts with regards to trial characteristics and reporting of important methodological details. A discussion of the generalisability of conclusions drawn from the cohorts was also presented. The following three chapters will focus on the reporting of trial outcomes. This chapter will begin with an outline of Outcomes Survey responses, followed by a descriptive overview of the population of outcomes identified for each cohort. Results relating to the prevalence of incomplete outcome reporting will then be presented and discussed.

Outcomes data for the three study cohorts were extracted from several sources. The PubMed cohort relied on publications and the Outcomes Survey; the Ethics sample obtained outcomes data from publications and protocols; and the CIHR cohort used publications, protocols, and the Outcomes Survey. All survey and publications data were included up to May 20, 2003.

5.2. Results

5.2.1. Responses to the Outcomes Survey of trialists

5.2.1.1. Completeness of responses

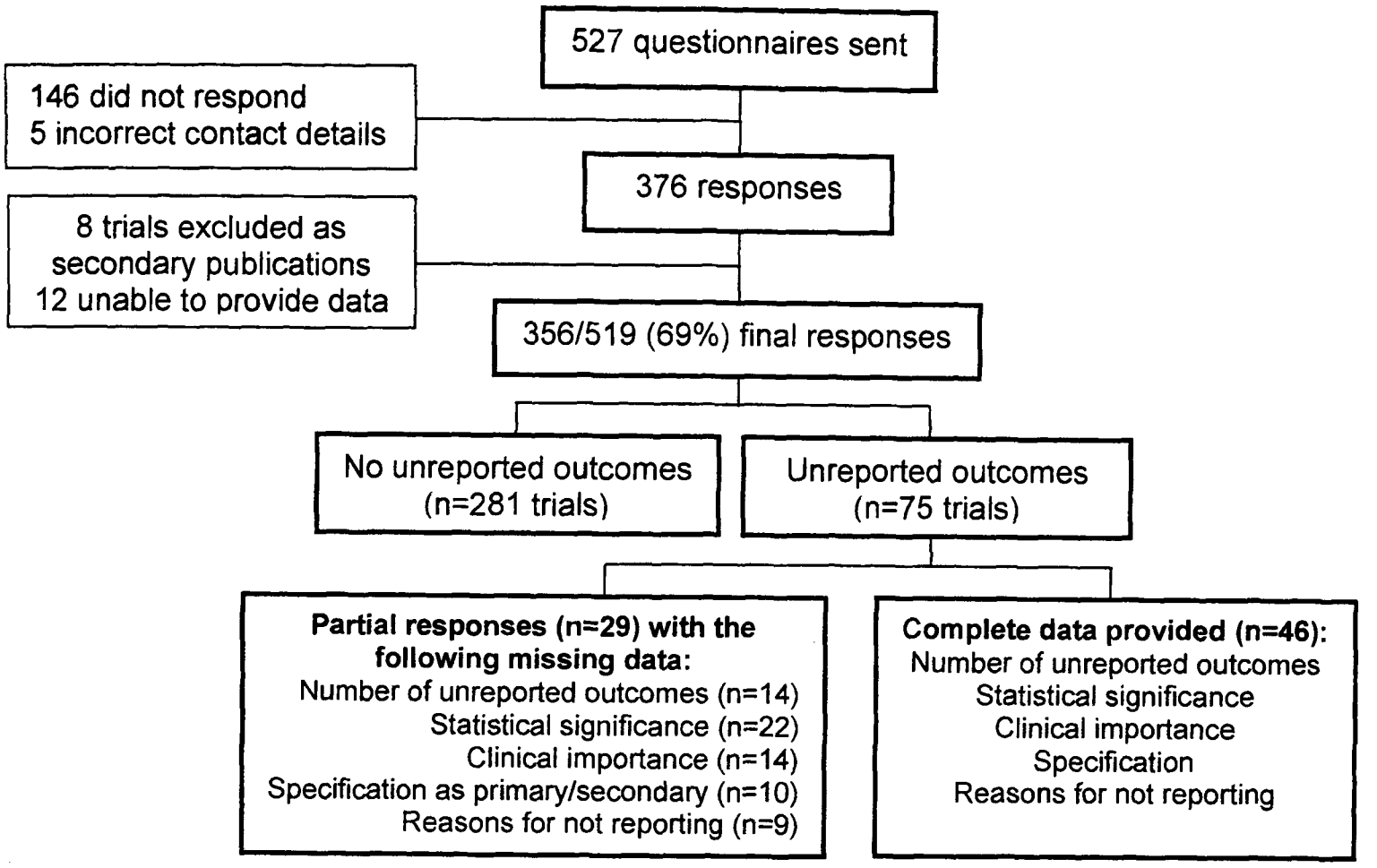
PubMed cohort

For the PubMed cohort, the Outcomes Survey questionnaires and two reminders were sent to trialists from February 2002 to August 2002. 69% (356/519) of trialists returned partially or fully completed questionnaires (Figure 5-1a). Among 75 responders who provided a list of unreported outcomes, 46 (61%) gave complete data for each of these outcomes, while 29 (39%) provided only partial details for them. Among 163 non-responders, 5 trialists did not receive the questionnaire due to inaccurate contact details. Twelve authors stated that they were unable to complete the survey due to a lack of time (n=7), industry control of data (n=1), multiple study centres (n=2), and other unspecified reasons (n=2). Although responses were received from these 12 trialists, they were classified as non-responders for the purposes of analysis because they did not provide any data.

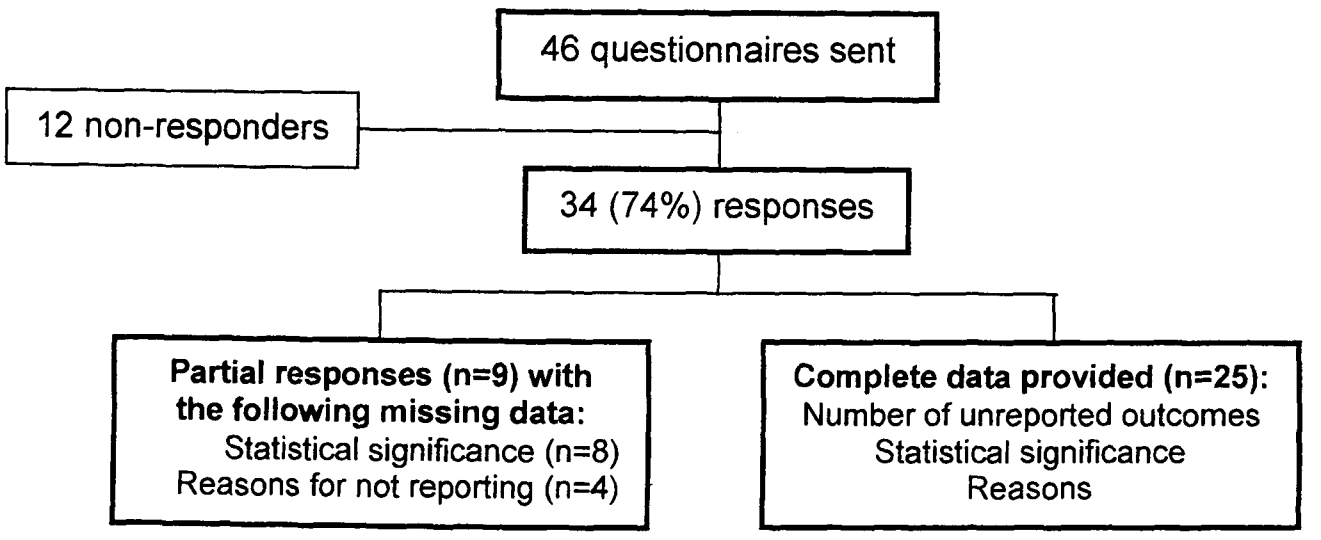
CIHR cohort

Outcomes Survey questionnaires for the CIHR cohort were sent in March 2003, followed by bi-weekly e-mail and telephone reminders until May 2003. A list of unreported outcomes identified from the comparison of protocols to publications was provided to each trialist. A response rate of 74% (34/46) was obtained for the Outcomes Survey (Figure 5-1b). 25/34 (74%) responses provided complete data, while 9/34 (26%) were missing information about either the statistical significance of outcomes or the reasons for not reporting

them. Two trialists were not surveyed because their studies contained no unreported outcomes based on protocols, publications, and responses to the relevant question in the Publications Survey.



a) PubMed cohort



b) CIHR cohort

Figure 5-1. Number and completeness of Outcomes Survey responses from PubMed and CIHR trialists

No trialist in either PubMed or CIHR cohorts supplemented the list of unreported outcomes in the Outcomes Survey with additional omitted outcomes.

5.2.1.2. Comparability of responders and non-responders

In the PubMed and CIHR cohorts, Outcomes Survey response rates were similar across study designs, journal types, and number of study sites (Table 5-1). However, differences in response rates were observed between the sources of funding for PubMed trials, as well as sample sizes for CIHR trials. Trialists for PubMed studies with no funding or non-industry funding were more likely to respond than those with full industry sponsorship (response rate 75-80% versus 65%). Accordingly, when comparing the trial characteristics of responders and non-responders in the PubMed cohort (Table 5-2), the latter were more frequently supported solely by commercial sources compared to the former (53% versus 30%). In the CIHR cohort, a higher response rate was observed for trials with sample sizes above the median of 299 (Table 5-1). In comparing the median sample size of responders to non-responders, the former recruited many more participants (median 506 versus 109) (Table 5-2). Differences between non-responders and responders were not formally tested because specific hypotheses were not formulated.

5.2.1.3. Reliability of survey responses

Trialists were asked whether there were any unreported outcomes in their studies. PubMed trialists were asked this question in the Outcomes Survey, while Ethics and CIHR trialists were asked in the initial Publications Survey. The concordance between these survey responses and the review of

publications and/or protocols was evaluated to provide an indication of the reliability of survey responses in each trial cohort.

Table 5-1. Outcomes Survey response rates stratified by trial characteristics and cohort

Trial characteristic	Response rate (%)	
	PubMed trials (n=519)	**CIHR trials (n=46)[a]**
STUDY DESIGN		
Parallel group	258/383 (67%)	27/38 (71%)
Cross-over	81/116 (70%)	3/3 (100%)
Other	17/20 (85%)	4/5 (80%)
JOURNAL TYPE		
General	25/37 (68%)	15/20 (75%)
Specialty	331/482 (69%)	19/26 (73%)
STUDY SITES		
Single centre	268/376 (71%)	10/15 (67%)
Multicentre	88/134 (66%)	24/31 (77%)
SAMPLE SIZE		
≥ cohort median[b]	173/261 (66%)	20/23 (87%)
< cohort median	183/258 (71%)	14/23 (61%)
FUNDING		
Industry only	108/167 (65%)[c]	N/A
Partial industry	46/61 (75%)[c]	15/20 (75%)
Non-industry	147/184 (80%)[c]	19/26 (73%)
None	54/54 (100%)[c]	N/A

[a] Two trials contained no unreported outcomes and were not surveyed
[b] Median sample sizes were 52 and 299 for PubMed and CIHR cohorts respectively
[c] Denominators correspond to trials with known sources of funding (n=466 trials overall)

Table 5-2. Trial characteristics of Outcomes Survey responders and non-responders in PubMed and CIHR cohorts

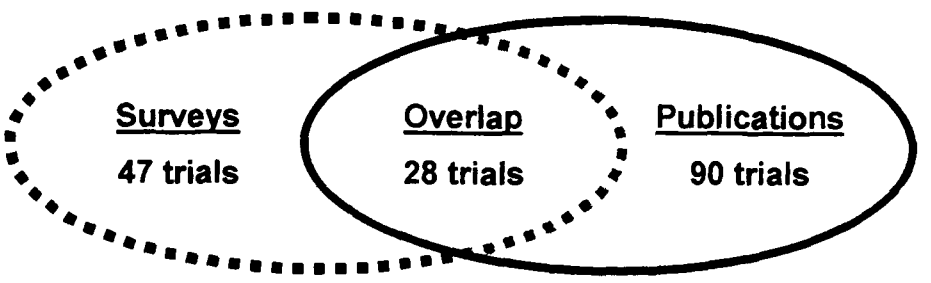
Cohort	Design[a]		General Journal (%)	Full industry funding (%)	Multicentre (%)	Median sample size [IPR_{80}]	Median number of reported outcomes [IPR_{80}]
	Parallel group (%)	Cross-over (%)					
PUBMED							
Non-responders (n=163)	125 (77%)	35 (21%)	12 (7%)	59/111[b] (53%)	46/154[b] (30%)	56 [15-353]	15 [4-45]
Responders (n=356)	258 (72%)	81 (23%)	25 (7%)	108/355[b] (30%)	88 (25%)	50 [12-269]	13 [4-40]
CIHR[c]							
Non-responders (n=12)	11 (92%)	0	5 (42%)	0	7 (58%)	109 [30-1127]	16 [6-29]
Responders (n=34)	27 (79%)	3 (9%)	15 (44%)	0	24 (71%)	506 [66-3577]	22 [5-45]

[a] Other designs not included in table
[b] Denominators correspond to trials with known sources of funding or number of study sites
[c] Two trials contained no unreported outcomes and were not surveyed

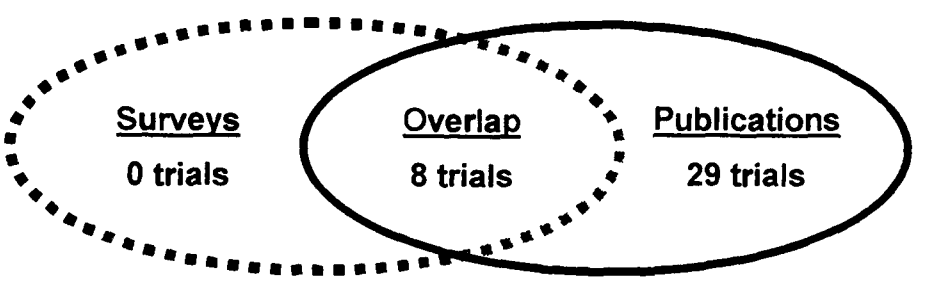
With regards to identifying unreported outcomes, discrepancies between trialist responses and evidence from publications or protocols were observed in 25%-85% of trials. Among 356 survey responders in the PubMed cohort, 21% (75/356) of trialists stated that there were unreported outcomes in their study. From the trial publications of the same 356 studies, 33% (118/356) contained unreported outcomes that were mentioned in the Methods but not the Results section. However, the overlap was only 8% (28/356) (Figure 5-2a). In other words, 25% (90/356) of survey responses stated that there were no unreported outcomes despite the publications indicating otherwise.

Similarly, before providing them with a list of unreported outcomes in the Outcomes Survey, CIHR trialists were asked in the Publications Survey whether there were any unreported outcomes in their published study (Figure 5-2b). Among 37/48 (77%) trials with responses to this particular question, all 37 had evidence of unreported outcomes in the protocols and publications. However, 78% (29/37) of these trialists stated in the questionnaire that no outcomes had been omitted from publications.

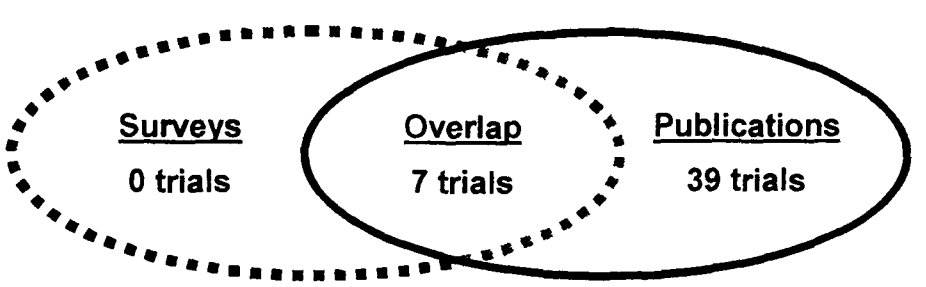
Ethics trialists were also asked about the existence of unreported outcomes in the Publications Survey. Among the 46/102 (45%) trials with responses to this question, all 46 had unreported outcomes identified from protocols and publications (Figure 5-2c). However, 85% (39/46) of these trialists stated that unreported outcomes did not exist.



a) Outcomes Survey responders with unreported outcomes based on responses (n=75) and publications (n=118) in PubMed cohort



b) Publications Survey responders with unreported outcomes based on responses (n=8) and publications (n=37) in CIHR cohort



c) Publications Survey responders with unreported outcomes based on responses (n=7) and publications (n=46) in Ethics cohort

Figure 5-2. Consistency between survey responses and publications in confirming the existence of one or more unreported outcomes in three trial cohorts

Further discrepancies in the CIHR cohort were noted between the study protocols and the reasons given by trialists for omitting outcomes. 26% (12/46) of trialists explained that at least one unreported outcome listed in the questionnaire was not actually intended for between-group analyses. However, 42% (5/12) of these trialists had explicitly specified the outcomes under consideration as secondary outcomes in the protocol. An additional trialist stated that a particular outcome had been omitted due to a lack of clinical importance, even though the outcome had been specified as primary in the protocol.

5.2.2. Overview of outcome characteristics in randomised trials

5.2.2.1. Review of outcome data collection

Data on study outcomes in PubMed trials were obtained from trial reports and Outcomes Survey responses. Reported outcomes were ascertained from trial publications, while unreported outcomes were recorded from both publications and questionnaires. As defined in Chapter 2, reported outcomes consisted of those that were fully, partially, and qualitatively reported in the publications. By definition, the statistical significance of reported outcomes was known from the trial reports, and the levels of reporting were distinguished only by the amount of data presented. In contrast, the statistical significance of unreported outcomes was available through the Outcomes Survey alone.

For the Ethics cohort, outcomes data were extracted from trial protocols and publications. Reported outcomes were recorded from publications, while unreported outcomes were noted if they were specified in protocols but not reported in publications, or if they were mentioned in the Methods but not the Results sections of publications. The statistical significance of unreported outcomes was unknown because the Outcomes Survey could not be conducted due to practical considerations, as outlined in Chapter 2.

Data on study outcomes from CIHR trials were collected from trial protocols, publications, and Outcomes Surveys. Reported outcomes were defined by publications, while unreported outcomes were defined in three ways - those that were specified in protocols but not reported in publications; those that were listed in the Methods but not the Results sections of publications; and those that were

provided in Outcomes Surveys. The statistical significance of unreported outcomes was provided by Outcomes Survey responses.

Table 5-3. Number of efficacy and safety outcomes measured in PubMed, Ethics, and CIHR cohorts

Cohort	Number of outcomes (% of total) [a]					
	n trials	Total outcomes	n trials	Efficacy outcomes	n trials	Safety outcomes
PubMed	519	10,557	505	8325 (79%)	308	2232 (21%)
Ethics	102	3613	99	2694 (75%)	69	919 (25%)
CIHR	48	1390	48	1233 (89%)	26	157 (11%)

[a]Defined in publications, surveys (PubMed and CIHR cohorts), or protocols (Ethics and CIHR cohorts)

5.2.2.2. Overall number of trial outcomes

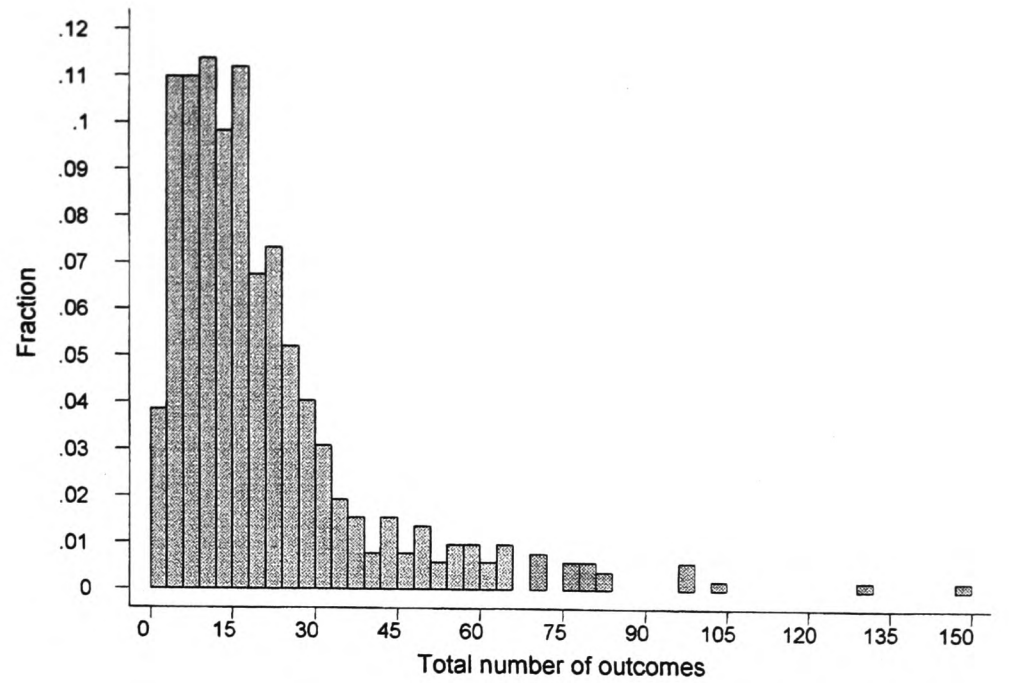
Overall numbers of outcomes stratified by efficacy or safety data are displayed for each cohort in Table 5-3. These numbers incorporate data on both reported and unreported outcomes measured in a trial. A total of 10,557 outcomes (79% efficacy and 21% safety outcomes) were measured in the 519 PubMed trials. 505 trials measured at least one efficacy outcome, while one or more safety outcomes was recorded in 308 studies. In the Ethics cohort, 102 trials measured 3613 outcomes. 75% of these were efficacy outcomes from 99 trials, while 25% were safety outcomes from 69 trials. In the CIHR cohort, a total of 1390 outcomes (89% efficacy and 11% safety outcomes) were measured in 48 trials, with 48 and 26 trials measuring at least one efficacy and safety outcome respectively. The number of trials with at least one efficacy outcome or at least one safety outcome represents the denominator for subsequent analyses stratified by efficacy and safety outcomes.

The frequency distributions for the number of outcomes measured per trial exhibited positive skewness in the three cohorts (Figure 5-3). The median numbers of measured outcomes are shown in Table 5-4. Including all study designs, a median of 15 [IPR_{80}* 4-43] outcomes per trial (n=519) were observed. Stratification by efficacy and safety outcomes yielded a median of 11 [IPR_{80} 3-36] and 4 [IPR_{80} 1-17] per trial respectively. These numbers were similar across study designs.

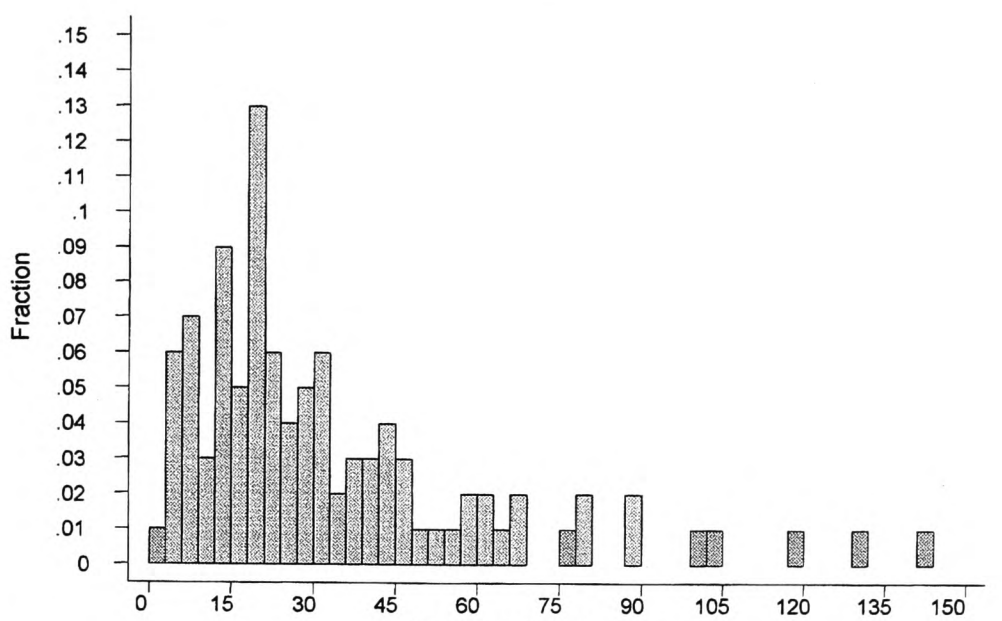
Ethics and CIHR trials measured more outcomes per study compared to the PubMed cohort (Table 5-4). A median of 24 [IPR_{80} 7-78] outcomes were recorded per Ethics trial (n=102), with a median of 19 [IPR_{80} 5-63] efficacy and 6 [IPR_{80} 1-37] safety outcomes per trial. Similarly, in the CIHR cohort, the median number of outcomes per trial (n=48) was 26 [IPR_{80} 10-57]. Stratified by efficacy and safety data, 20 [IPR_{80} 6-54] and 5 [IPR_{80} 1-11] outcomes respectively were measured per trial. Within both Ethics and CIHR cohorts, similar numbers of outcomes per trial were observed across study designs (Table 5-4).

---

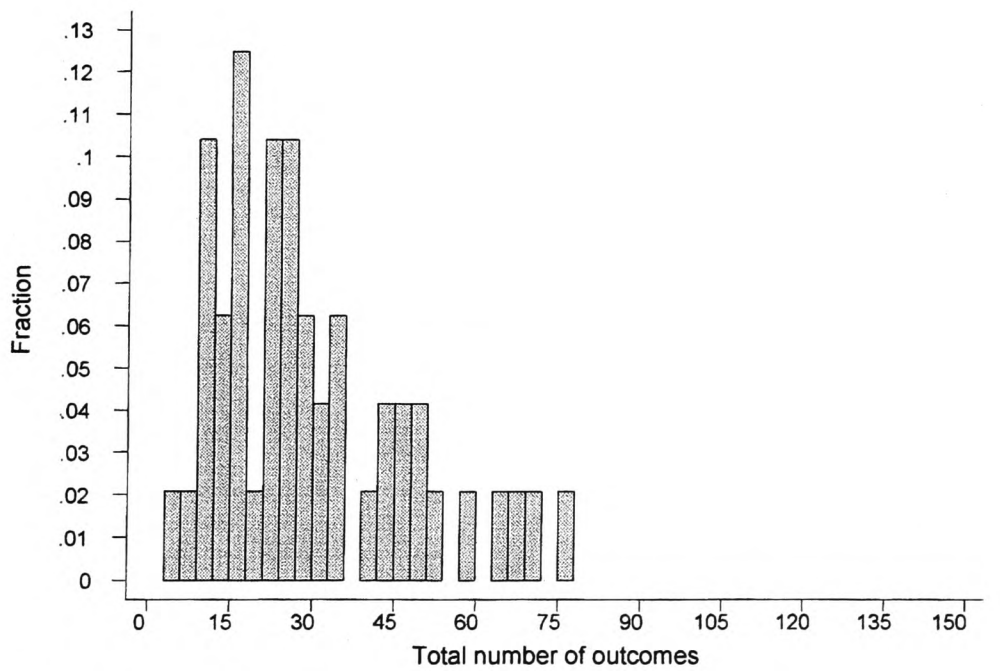
* Inner-80-percentile range (IPR_{80}) is bound by the 10^{th} and 90^{th} percentiles[216]



a) PubMed cohort (median 15 outcomes, IPR_{80} 4-43)



b) Ethics cohort (median 24 outcomes, IPR_{80} 7-78)



c) CIHR cohort (median 26 outcomes, IPR_{80} 10-57, n=48 trials)

Figure 5-3. Relative frequency histogram of the total number of measured outcomes per trial in the three cohorts (medians and inner-80-percentile ranges [IPR_{80}] are provided)

Table 5-4. Median number of outcomes per trial by study design and cohort

Trial design	Median number of outcomes per trial [IPR_{80}]					
	n trials	All outcomes	n trials	Efficacy outcomes	n trials	Safety outcomes
ALL						
PubMed cohort	519	15 [4-43]	505	11 [3-36]	308	4 [1-17]
Ethics cohort	102	24 [7-78]	99	19 [5-63]	69	6 [1-37]
CIHR cohort	48	26 [10-57]	48	20 [6-54]	26	5 [1-11]
PARALLEL GROUP						
PubMed cohort	383	16 [4-43]	375	11 [3-37]	237	4 [1-18]
Ethics cohort	70	30 [13-89]	68	19 [6-69]	56	9 [1-40]
CIHR cohort	39	24 [9-63]	39	19 [6-63]	18	7 [1-15]
CROSS-OVER						
PubMed cohort	116	14 [5-51]	110	10 [4-50]	62	4 [1-14]
Ethics cohort	30	21 [5-53]	29	20 [4-57]	12	2 [1-6]
CIHR cohort	3	26 [11-45]	3	24 [6-38]	3	5 [2-7]
OTHER						
PubMed cohort	20	13 [5-27]	20	12 [3-22]	9	4 [1-15]
Ethics cohort	2	19 [8-29]	2	15 [8-22]	1	7 [N/A]
CIHR cohort	6	34 [12-52]	6	32 [10-49]	5	3 [1-6]

To assess whether the higher number of total outcomes per Ethics and CIHR trial was due to the availability of protocol data, the median numbers were re-calculated using data from publications alone. This produced medians [IPR_{80}] of 15 [4-43], 21 [6-67], and 21 [7-45] for the total number of outcomes per trial in the PubMed, Ethics, and CIHR cohorts respectively.

A sensitivity analysis was conducted for the Ethics and CIHR cohorts to assess the impact of including vague unreported outcomes. These were variables that were vaguely described in protocols without an analysis plan but not reported in publications, and it was thus unclear whether they were intended for inter-group comparisons. As explained in Chapter 3, these potential outcomes had been excluded from the primary analyses in order to obtain a conservative estimate of the number of outcomes measured per trial. Including 947 vague variables from 69 trials resulted in a slightly higher median of 28 [IPR_{80} 8-78] outcomes per Ethics trial (n=102 trials). The estimate for the CIHR cohort was unchanged with the inclusion of 167 vague variables from 22 trials.

5.2.2.3. Characteristics of trial outcomes

Type of outcome data

Trials measured a variety of outcome types consisting predominantly of binary and continuous data. The type of outcome data was available for both reported and unreported outcomes in the Ethics and CIHR cohorts, but this information was not known for unreported outcomes in the PubMed cohort because protocols were not available.

Of 505 PubMed trials with efficacy outcomes, 445 (88%) trials reported continuous efficacy outcomes, while 256 (51%) and 52 (10%) trials reported binary and ordinal outcomes respectively. Twenty-five (5%) studies reported time-to-event outcomes. For 308 trials with safety outcomes, 153 (50%) reported continuous safety outcomes, while 284 (92%) reported binary outcomes. Six (2%) and 1 (0.3%) trial reported ordinal and time-to-event safety outcomes respectively. These proportions are not mutually exclusive.

Similar prevalences were observed in the Ethics cohort, although ordinal and survival outcomes were used more frequently. Of 99 studies with efficacy data, continuous efficacy outcomes were measured in 85 (86%) trials, while binary efficacy outcomes were recorded in 49 (49%) trials. Ordinal and time-to-event efficacy outcomes were present in 25 (25%) and 17 (17%) trials respectively. With respect to safety outcomes, 36/69 (52%) trials measured continuous outcomes, while 55 (80%) recorded binary outcomes. Five (7%) and 3 (4%) recorded ordinal and time-to-event safety outcomes respectively.

While the prevalence of continuous efficacy outcomes (41/48, 85%) in the CIHR cohort was similar to that in the PubMed and Ethics cohorts, a higher proportion of CIHR trials measured binary (44/48, 92%) and time-to-event (20/48, 42%) efficacy outcomes. Eight (17%) trials reported ordinal efficacy outcomes. For 26 trials with safety outcomes, 7 (27%) reported continuous outcomes, while 24 (92%) reported binary outcomes. Three (12%) and 1 (4%) trial reported ordinal and time-to-event safety outcomes respectively.

Statistical significance of reported outcomes

The prevalence of statistically significant efficacy outcomes reported in publications was comparable across the three trial cohorts, while more variation was observed with safety outcomes. Among efficacy outcomes with known significance, the overall proportion of significant outcomes was 33% (2456/7470), 35% (729/2107), and 32% (298/945) in PubMed, Ethics, and CIHR cohorts respectively. Among trials with efficacy outcomes, 79% (399/505) produced at least one statistically significant efficacy outcome in the PubMed cohort, compared to 84% (83/99) and 88% (42/48) of Ethics and CIHR trials respectively.

In terms of safety outcomes with known significance, 19% (370/1916), 20% (117/594), and 25% (26/105) of all outcomes were significant in PubMed, Ethics, and CIHR cohorts respectively. Only 35% (109/308) of PubMed trials reported a significant safety outcome, compared to 49% (34/69) and 50% (13/26) of trials in the Ethics and CIHR cohorts respectively.

Number and nature of publication-defined primary outcomes

The proportion of trials defining primary outcomes in their publications varied across cohorts, ranging from 45% (232/519) in the PubMed cohort to 62% (63/102) and 94% (45/48) in the Ethics and CIHR cohorts respectively. In the PubMed cohort, primary outcomes were identified through explicit description in 67% (155/232) of trials; power calculation in 19% (45/232); and study objectives in 14% (32/232). Similar proportions were observed in Ethics trials, with 75% (47/63) defining their primary outcomes explicitly in the reports, 16% (10/63) in the power calculation, and 10% (6/63) in the study objectives. All publication-defined primary outcomes in the CIHR cohort were specified explicitly as primary.

The vast majority of publication-defined primary outcomes measured efficacy data. 82% (756/925) of primary outcomes in the PubMed cohort were efficacy outcomes, as were 99% (175/177) and 91% (69/76) of primary outcomes in the Ethics and CIHR cohorts respectively. Among trials that defined primary outcomes in their reports, only 3%-10% in each cohort specified at least one primary safety outcome, compared to 93%-98% for efficacy outcomes.

The number of primary outcomes specified in publications for each trial is detailed in Table 5-5. PubMed trials generally defined a larger number of primary outcomes in their reports than the other two cohorts. 38% (89/232) of PubMed trials specified more than two primary outcomes, compared to 27% (17/63) and 9% (4/45) for Ethics and CIHR trials respectively.

Table 5-5. Number of publication-defined primary outcomes per trial

Number of primary outcomes per trial	Number of trials (%)		
	PubMed (n=232 trials)	Ethics (n=63 trials)	CIHR (n=45 trials)
1	102 (44%)	39 (62%)	35 (78%)
2	41 (18%)	7 (11%)	6 (13%)
3-5	46 (20%)	7 (11%)	2 (4%)
6-10	22 (9%)	6 (10%)	1 (2%)
11-15	11 (5%)	4 (6%)	0
>15	10 (4%)	0	1 (2%)

The proportion of trials with statistically significant primary outcomes was comparable across cohorts. Among trials that defined primary outcomes, 58% (134/232) had significant primary outcomes in the PubMed cohort, compared to 60% (38/63) of Ethics and 49% (22/45) of CIHR trials.

5.2.3. Completeness of outcome reporting in randomised trials

The first objective of this thesis was to assess the prevalence of incomplete outcome reporting in randomised trials. Incompletely reported outcomes were those for which insufficient data were provided for inclusion in a meta-analysis. This definition encompassed partially, qualitatively, and unreported outcomes (see Figure 2-1). Partially reported variables included some of the data necessary for meta-analysis, while qualitative reporting consisted solely of a *p*-value or statement about statistical significance. Unreported outcomes were essentially omitted from the Results sections of trial publications with no useful data provided. It should be noted that the type of data required for full reporting differs depending on whether the outcome consists of unpaired or paired data.

Unreported outcomes in a random sample of 19 Ethics trials were verified by two individuals, with disagreements resolved by consensus. Corrections were made to 9/269 (3%) unreported outcomes in 2/19 (11%) trials.

5.2.3.1. Overview of reporting among all outcomes

A general overview of the distribution of outcomes across the four levels of reporting is shown in Table 5-6. Ignoring the clustering by trial, approximately one third to one half of all outcomes in each cohort were incompletely reported for meta-analysis. Efficacy outcomes in the CIHR cohort were more often fully reported compared to PubMed and Ethics cohorts (63% versus 41%-45%). Unreported outcomes were identified less frequently in the PubMed cohort than in the other two cohorts (8% versus 21%-23% for efficacy; 11% versus 32%-36% for safety). CIHR trials generally appeared to either fully report or omit their outcomes, as the prevalence of partially or qualitatively reported outcomes was

relatively low. Compared to efficacy outcomes, full reporting of safety outcomes was less common within the Ethics and CIHR cohorts.

Table 5-6. Overall number of efficacy and safety outcomes stratified by level of reporting

Trial design	Number of trials	Total number of outcomes	Number of outcomes at each level of reporting (% of total)			
			Fully reported	Partially reported	Qualitatively reported	Unreported
PUBMED COHORT						
Efficacy outcomes	505	8325	3785 (45%)	3041 (37%)	832 (10%)	667 (8%)
Safety outcomes	308	2232	1095 (49%)	376 (17%)	513 (23%)	248 (11%)
ETHICS COHORT						
Efficacy outcomes	99	2694	1107 (41%)	733 (27%)	300 (11%)	554 (21%)
Safety outcomes	69	919	275 (30%)	104 (11%)	213 (23%)	327 (36%)
CIHR COHORT						
Efficacy outcomes	48	1233	777 (63%)	93 (8%)	77 (6%)	286 (23%)
Safety outcomes	26	157	73 (46%)	8 (5%)	26 (17%)	50 (32%)

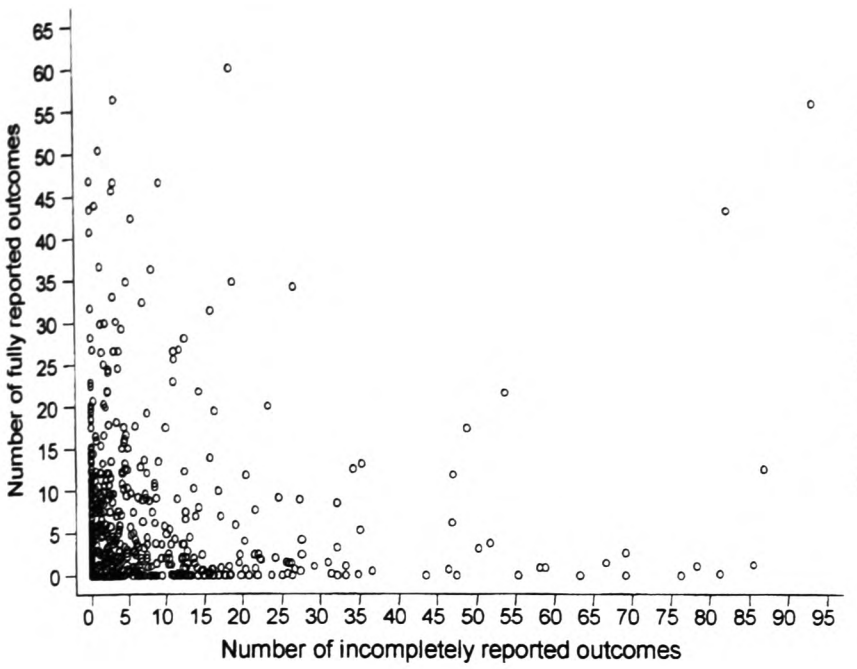
Binary outcomes were more likely to be fully reported than continuous outcomes. Complete reporting was observed in 85%-86% of binary outcomes across cohorts (319/375 in PubMed; 412/477 in Ethics; 395/457 in CIHR cohort), as opposed to 61% (289/471) of continuous outcomes in the PubMed cohort, 37% (851/2278) in Ethics trials, and 70% (347/496) in the CIHR cohort.

5.2.3.2. Proportion of incompletely reported outcomes per trial

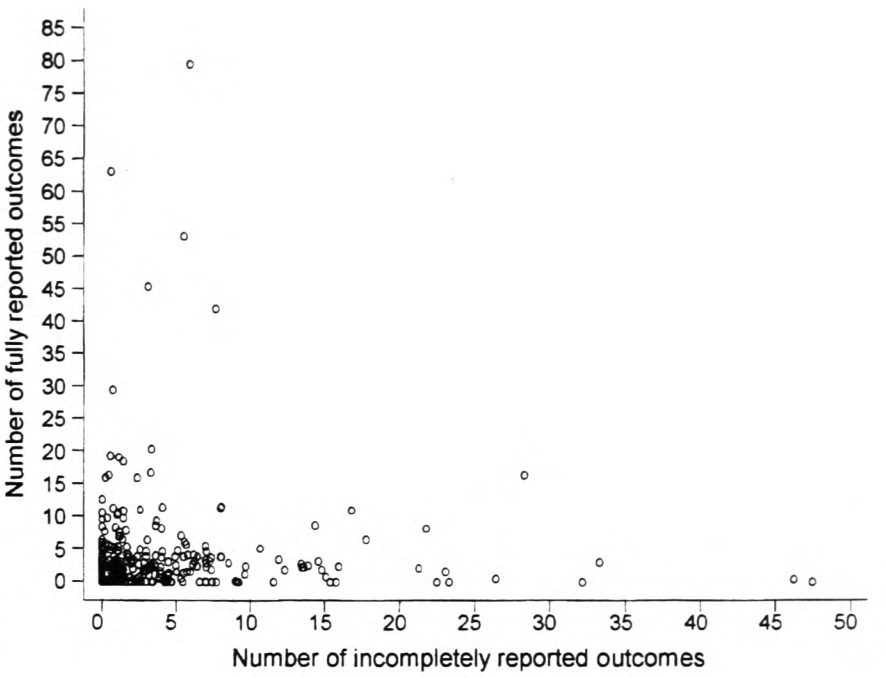
Because the number of outcomes per trial varied from one to over 100, it is difficult to summarise the proportion of outcomes that were incompletely reported. For this reason, the database has been examined descriptively in several ways. The number of incompletely reported outcomes per trial varied greatly relative to the number of fully reported outcomes (Figure 5-4) or the total number of outcomes (Figure 5-5). The proportions of incompletely reported outcomes in

each trial assumed distributions shown in Figure 5-6. The figures suggest that when higher numbers of outcomes were measured in a trial, the majority tended to be either fully or incompletely reported. This was most clearly observed in the PubMed cohort.

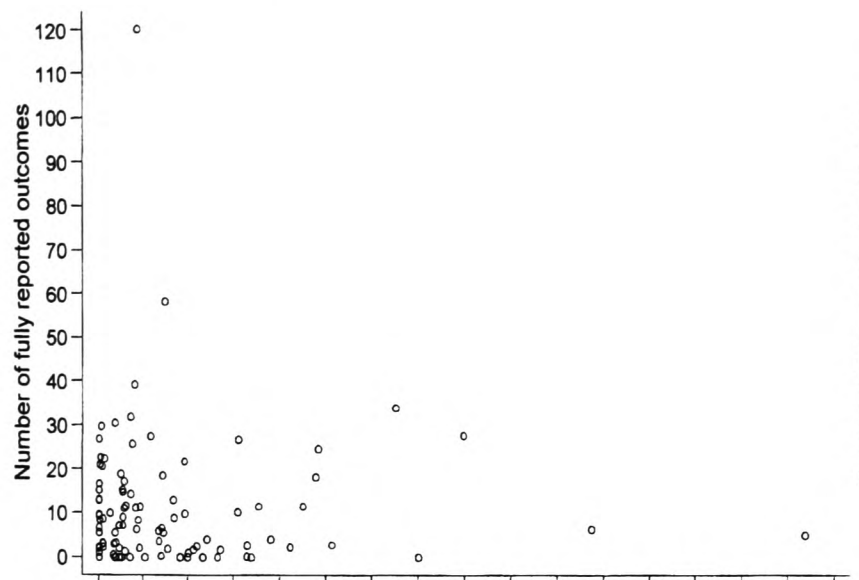
For incompletely reported efficacy outcomes, the median proportion was lower overall in CIHR trials (31%) than in PubMed (42%) and Ethics (48%) cohorts (Table 5-7). Median proportions of incompletely reported safety outcomes were similar across the three cohorts, ranging from 50%-60%. Stratified by study design, parallel group trials contained lower percentages of incompletely reported efficacy and safety outcomes compared to cross-over trials in the three cohorts (22%-35% versus 58%-100% respectively for efficacy; 24%-50% versus 71%-100% for safety).



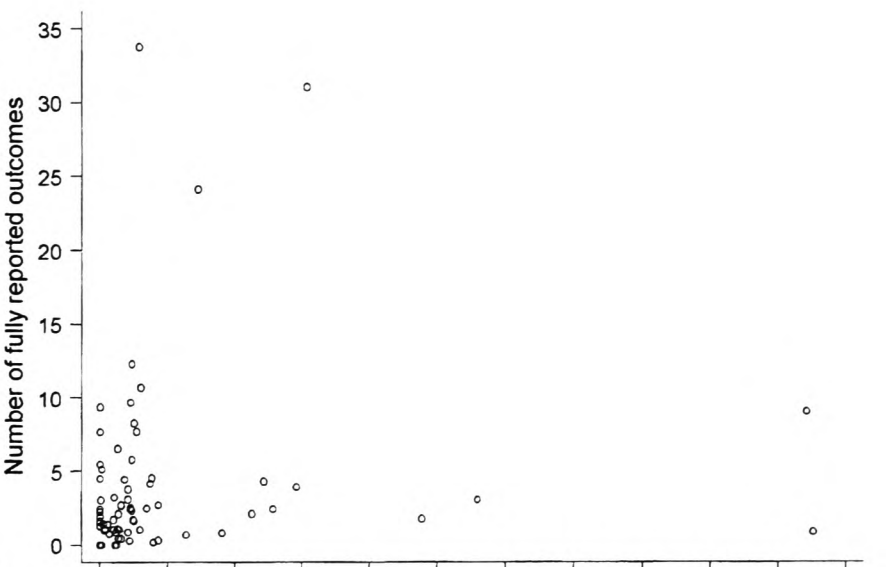
a) PubMed efficacy outcomes (n=505 trials)



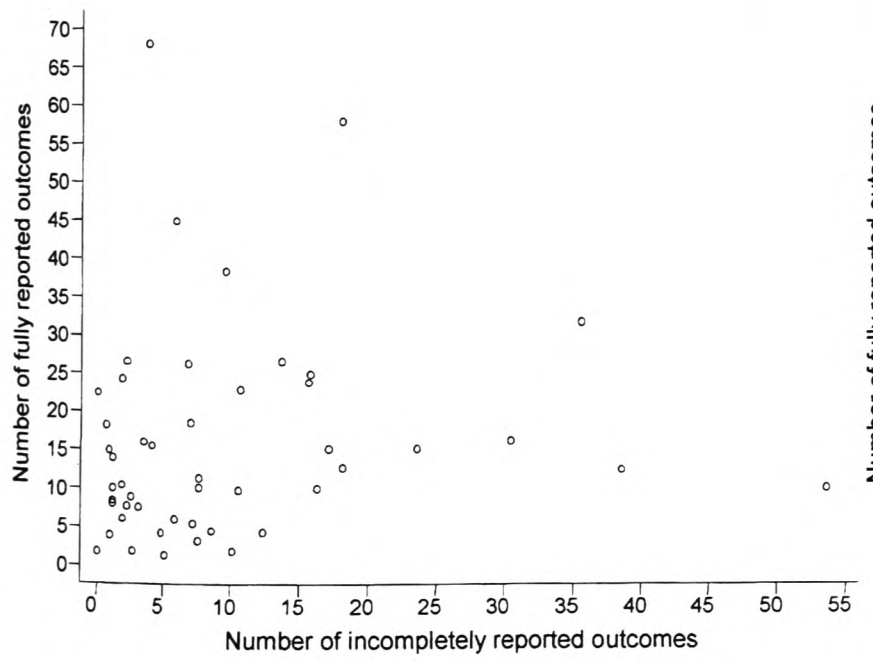
b) PubMed safety outcomes (n=308 trials)



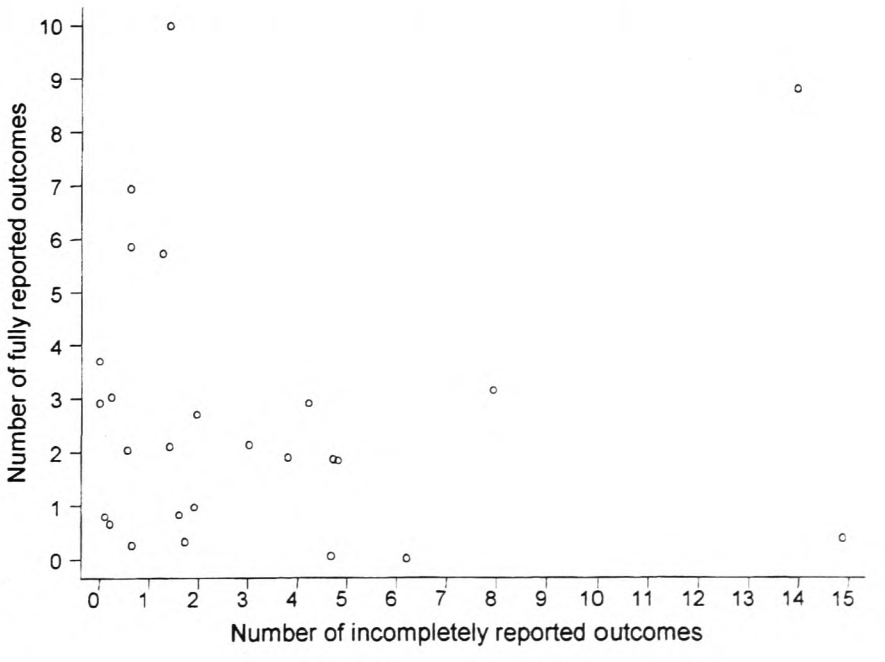
c) Ethics cohort efficacy outcomes (n=99 trials)



d) Ethics safety outcomes (n=69 trials)

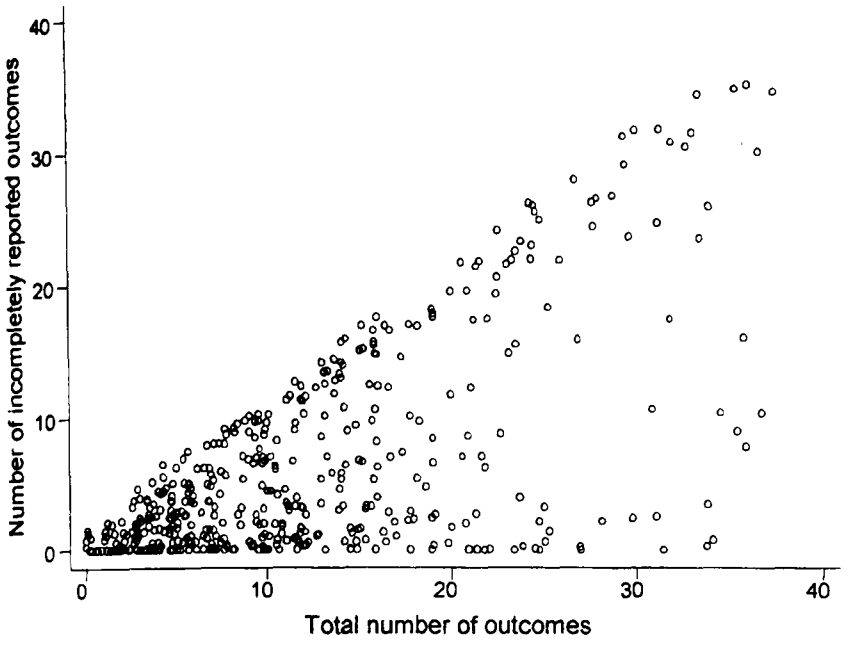


e) CIHR cohort efficacy outcomes (n=48 trials)

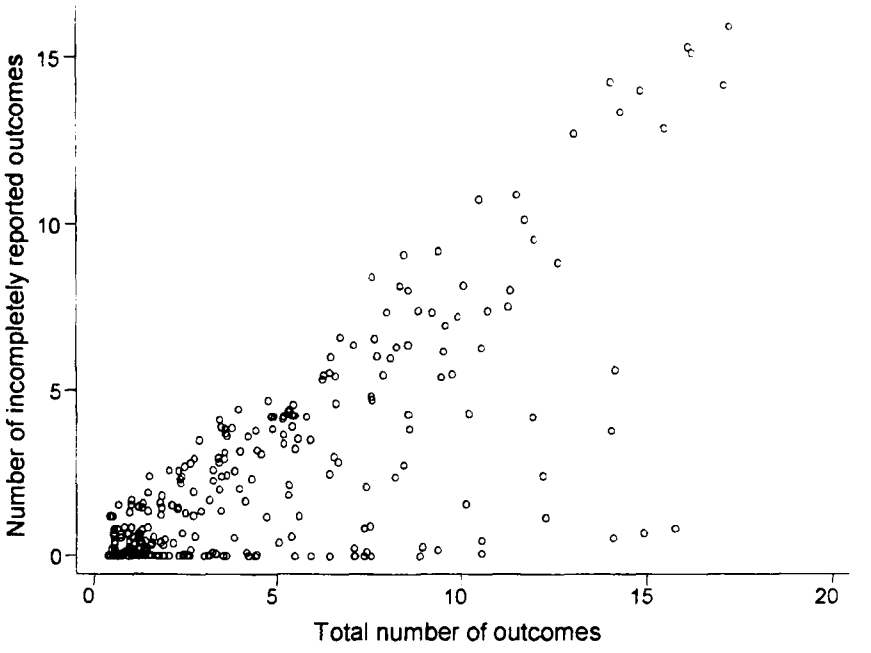


f) CIHR safety outcomes (n=26 trials)

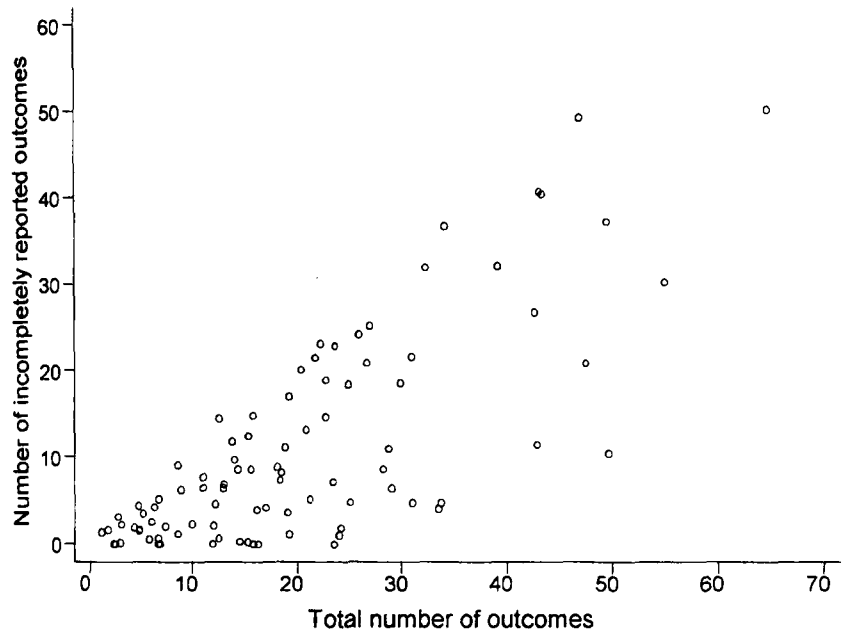
Figure 5-4. Scatterplot of the number of fully reported outcomes versus incompletely reported outcomes per trial in the three cohorts, stratified by efficacy and safety outcomes



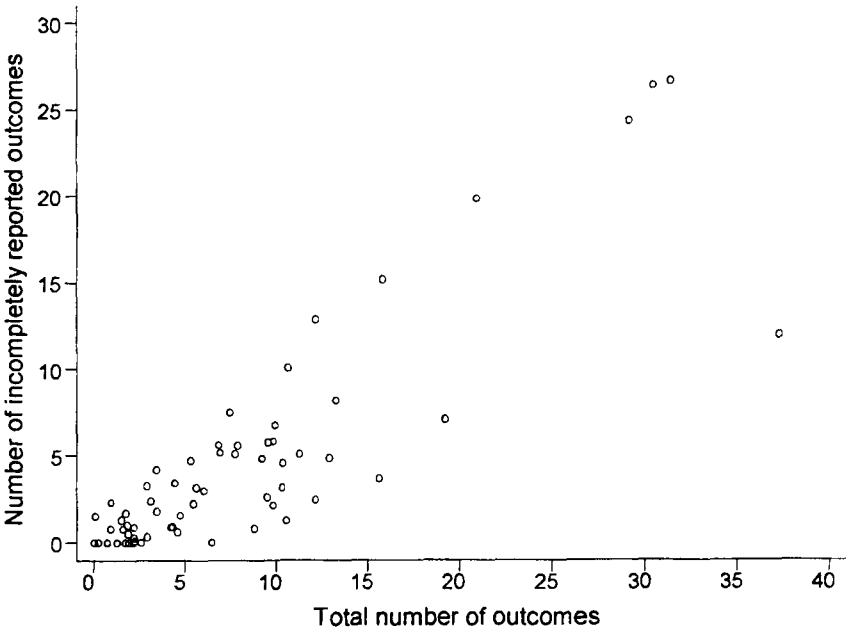
a) PubMed efficacy outcomes (n=455 trials)



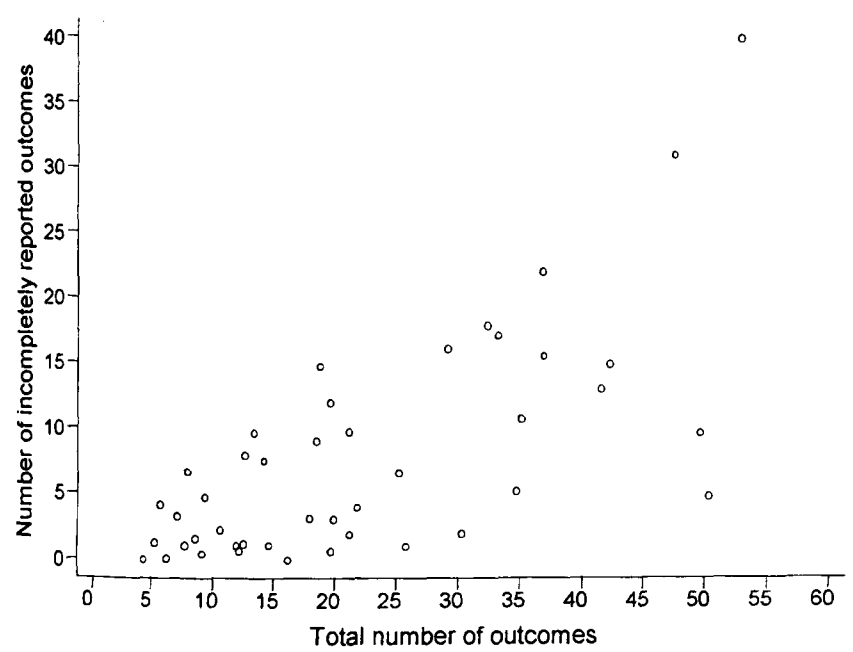
b) PubMed safety outcomes (n=278 trials)



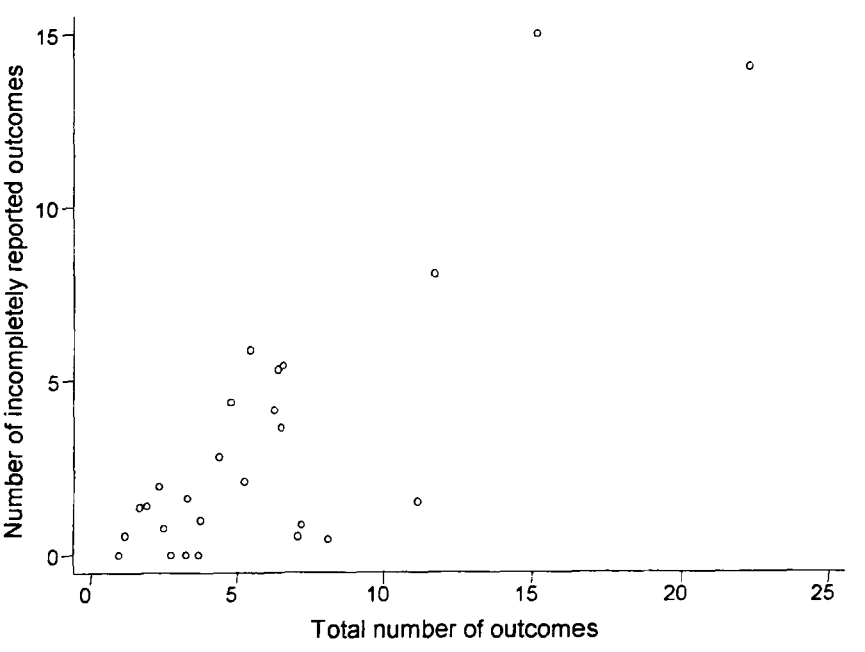
c) Ethics efficacy outcomes (n=90 trials)



d) Ethics safety outcomes (n=63 trials)

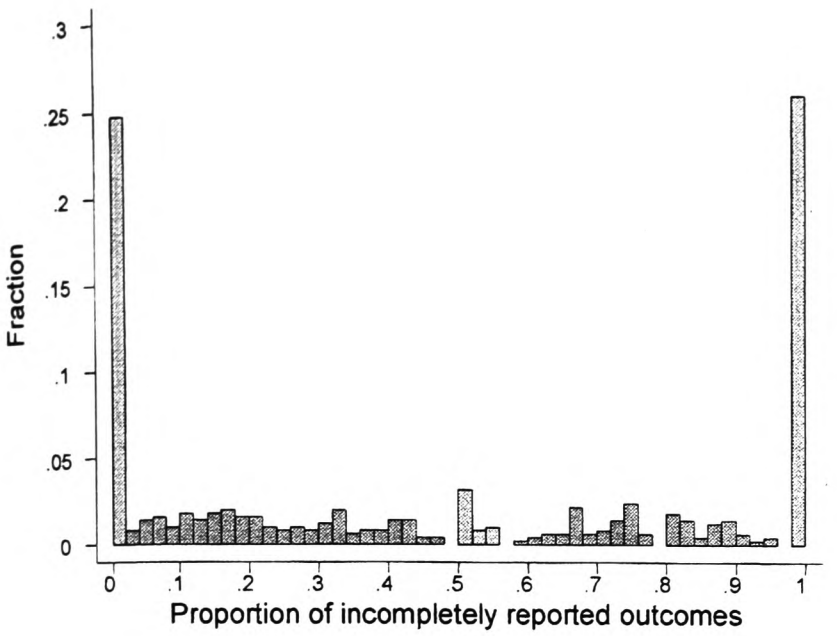


e) CIHR efficacy outcomes (n=44 trials)

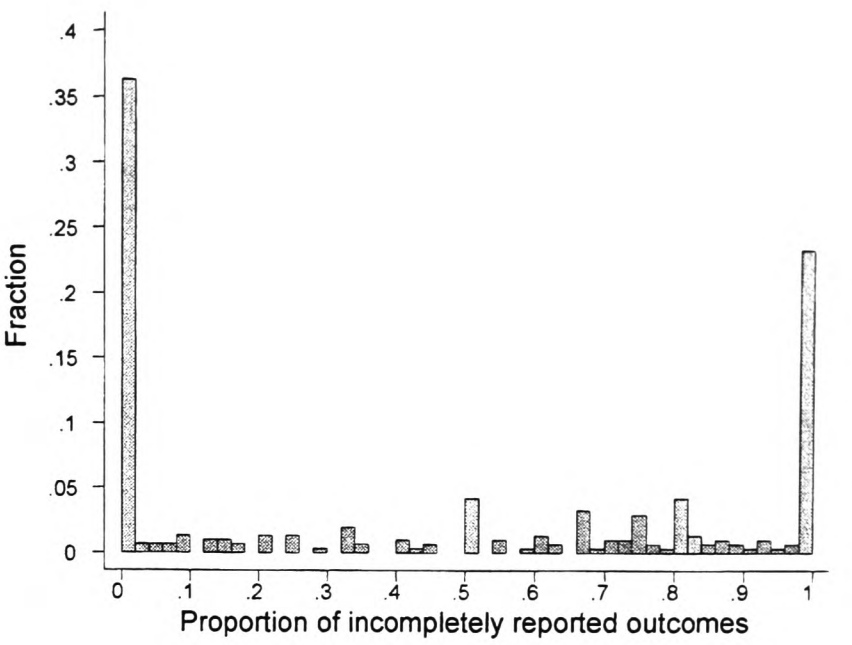


f) CIHR safety outcomes (n=24 trials)

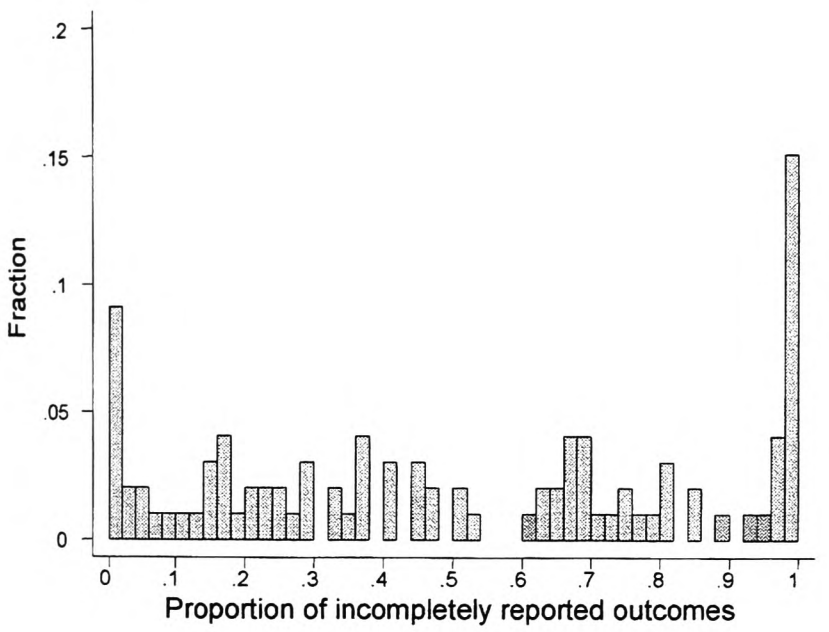
Figure 5-5. Scatterplot of the number of incompletely reported outcomes versus the total number of outcomes per trial (Trials with number of outcomes>90th percentile were excluded; data points are jiggled to avoid overlap)



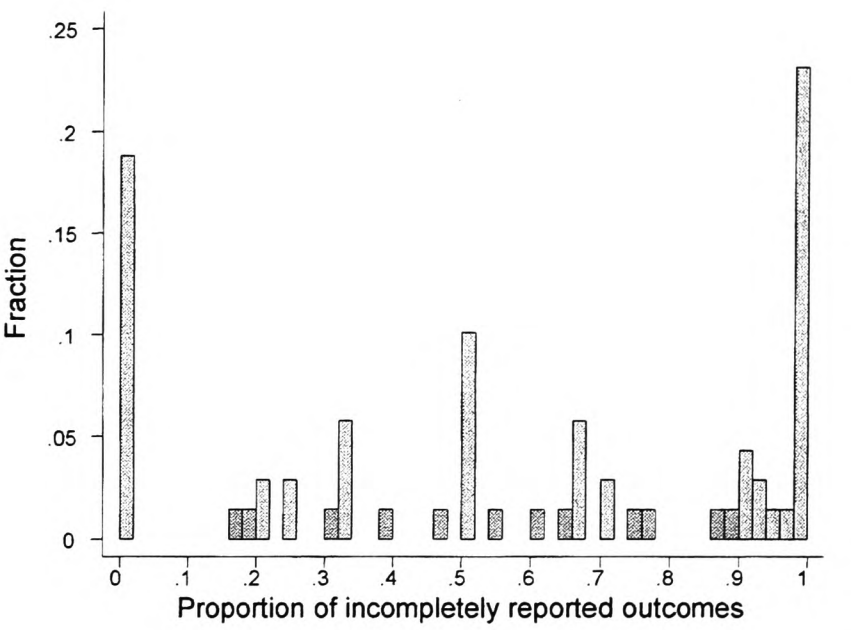
a) PubMed efficacy outcomes (n=505 trials)



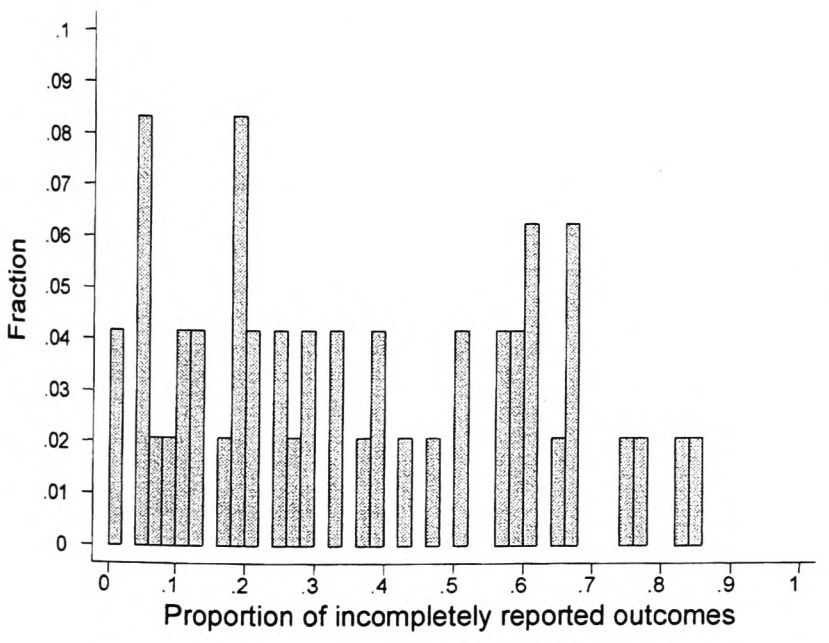
b) PubMed safety outcomes (n=308 trials)



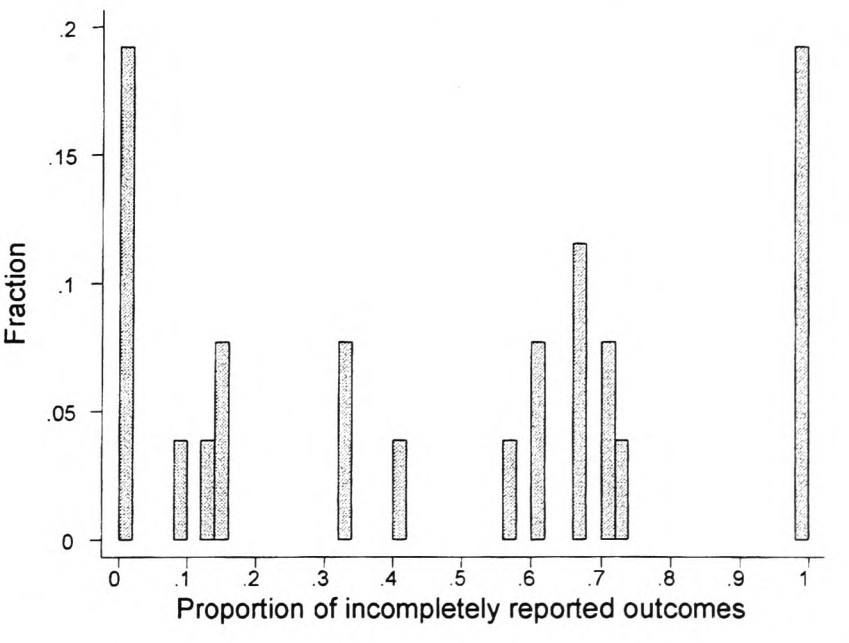
c) Ethics efficacy outcomes (n=99 trials)



d) Ethics safety outcomes (n=60 trials)



e) CIHR efficacy outcomes (n=48 trials)



f) CIHR safety outcomes (n=26 trials)

Figure 5-6. Relative frequency histogram of the proportion of incompletely reported efficacy and safety outcomes per trial in the three cohorts

Table 5-7. Median proportion of incompletely reported efficacy and safety outcomes per trial by study design

Trial design	Median % per trial [IPR_{80}]			
	n trials	Efficacy outcomes	n trials	Safety outcomes
ALL				
PubMed cohort	505	42% [0-100%]	308	50% [0-100%]
Ethics cohort	99	48% [4-100%]	69	60% [0-100%]
CIHR cohort	48	31% [5-67%]	26	59% [0-100%]
PARALLEL GROUP				
PubMed cohort	375	22% [0-100%]	237	25% [0-100%]
Ethics cohort	68	35% [4-78%]	56	50% [0-100%]
CIHR cohort	39	33% [5-74%]	18	24% [0-100%]
CROSS-OVER				
PubMed cohort	110	100% [48-100%]	62	82% [0-100%]
Ethics cohort	29	98% [0-100%]	12	100% [0-100%]
CIHR cohort	3	58% [4-67%]	3	71% [40-100%]
OTHER				
PubMed cohort	20	59% [0-100%]	9	80% [0-100%]
Ethics cohort	2	91% [82-100%]	1	71% [N/A]
CIHR cohort	6	18% [7-39%]	5	67% [60-100%]

Table 5-8. Median proportion of incompletely reported efficacy and safety outcomes per trial by journal type and funding source[a]

Trial characteristic	Median % per trial [IPR_{80}]			
	n trials	Efficacy outcomes	n trials	Safety outcomes
PUBMED COHORT				
General medical journal	37	40% [0-100%]	26	58% [0-100%]
Specialty journal	468	43% [0-100%]	282	47% [0-100%]
Full industry funding	163	46% [0-100%]	133	56% [0-100%]
Partial or non-industry funding	290	42% [0-100%]	142	27% [0-100%]
ETHICS COHORT				
General medical journal	3	15% [4-21%]	3	56% [25-100%]
Specialty journal	96	50% [4-100%]	66	62% [0-100%]
Full industry funding	54	50% [11-97%]	50	55% [0-100%]
Partial or non-industry funding	44	46% [0-100%]	18	61% [0-100%]
CIHR COHORT[b]				
General medical journal	21	21% [4-67%]	13	14% [0-71%]
Specialty journal	27	33% [10-67%]	13	67% [14-100%]

[a]Trials with unknown funding sources were excluded from the analysis of funding
[b]Stratification by funding source was not done because no trials had full industry sponsorship

The median proportions of incompletely reported outcomes per trial were also stratified by journal type and funding source (Table 5-8). There was evidence of greater completeness of reporting among CIHR trials published in general medical journals compared to specialty journals (33% versus 21% for efficacy and 67% versus 14% for safety outcomes). Also, fewer incompletely reported safety outcomes were observed in PubMed trials that were not fully funded by industry (27% versus 56%).

Sensitivity analyses

A sensitivity analysis was conducted for the Ethics and CIHR cohorts to assess the impact of including unreported outcomes with unclear outcome status. As noted earlier, these were unreported variables that were vaguely described in protocols but not reported in publications, and it was thus unclear whether they were intended for inter-group comparisons. Including these potential outcomes (947 variables in 69 Ethics trials; 167 variables in 22 CIHR trials) would over-estimate the true number of total and unreported outcomes. The median proportion of incompletely reported outcomes in all Ethics trials was increased to 53% [IPR_{80} 6%-100%] for efficacy outcomes (n=99 trials) and 67% [IPR_{80} 0%-100%] for safety outcomes (n=73 trials). Corresponding figures for the CIHR cohort were virtually unchanged to 31% [IPR_{80} 6%-67%] for efficacy outcomes (n=48 trials) and 60% [IPR_{80} 0%-100%] for safety outcomes (n=28 trials). For parallel group Ethics trials, an increased median of 40% [IPR_{80} 6%-78%] of efficacy outcomes (n=68 trials) and 62% [IPR_{80} 0%-100%] of safety outcomes (n=56 trials) were incompletely reported when vague unreported outcomes were included. In CIHR parallel group trials, the inclusion of these outcomes increased the median proportion of incompletely

reported efficacy and safety outcomes to 38% [IPR_{80} 6%-74%] in 39 trials and 33% [IPR_{80} 0%-100%] in 20 trials respectively.

A second sensitivity analysis was conducted in which trials with fewer than 5 efficacy outcomes were excluded, as proportions with small numbers of outcomes are unreliable. The resulting median proportion of incompletely reported efficacy outcomes was similar for all eligible PubMed trials (43% [IPR_{80} 0%-100%], n=409 trials) and slightly increased for parallel group trials (27% [IPR_{80} 0%-100%], n=301 trials). The median proportions were virtually unchanged for all Ethics trials (49% [IPR_{80} 5%-100%], n=90 trials) and parallel group trials (36% [IPR_{80} 4%-78%], n=63 trials). Median proportions for the CIHR cohort were also virtually unchanged in all trials (33% [IPR_{80} 6%-67%], n=47 trials) and parallel group trials (35% [IPR_{80} 6%-74%], n=38 trials). This analysis could not be conducted with safety outcomes, as the median of 4 safety outcomes observed per trial was already low.

5.2.3.3. Proportion of trials with incompletely reported outcomes

The proportion of trials with at least one efficacy or safety outcome at the four levels of reporting is shown in Table 5-9. The majority of trials contained incompletely reported efficacy and safety outcomes. Three-quarters of PubMed trials did not fully report all of their efficacy outcomes, compared to 91% of Ethics trials and 96% of CIHR trials. Although all Ethics and CIHR parallel group trials presented at least one fully reported efficacy outcome, only 86% of parallel group PubMed trials did the same - meaning that 14% did not fully report even one efficacy outcome (Table 5-9a). A much lower

proportion of cross-over trials presented any fully reported efficacy outcomes in PubMed (31%) and Ethics (52%) cohorts.

Similarly, 64%-81% of trials in the three cohorts did not fully report all of their safety outcomes. 81%-89% of parallel group trials presented at least one fully reported safety outcome (Table 5-9b). Lower proportions were observed in cross-over trials (60% of PubMed; 33% of Ethics; 67% of CIHR trials).

Table 5-9. Proportion of trials with at least one outcome at specific levels of reporting

a) Efficacy outcomes

Trial design by study cohort	Total number of trials	Number of trials (%) with ≥1 efficacy outcome at specific levels of reporting				
		Full	Incomplete			
			Any	Partial	Qualitative	Unreported
ALL						
PubMed cohort	505	375 (74%)	380 (75%)	270 (53%)	138 (27%)	169 (33%)
Ethics cohort	99	84 (85%)	90 (91%)	54 (55%)	40 (40%)	64 (65%)
CIHR cohort	48	48 (100%)	46 (96%)	22 (46%)	12 (25%)	42 (88%)
PARALLEL GROUP						
PubMed cohort	383	328 (86%)	260 (68%)	163 (43%)	103 (27%)	125 (33%)
Ethics cohort	68	68 (100%)	62 (91%)	29 (43%)	27 (40%)	48 (71%)
CIHR cohort	39	39 (100%)	37 (95%)	19 (49%)	9 (23%)	34 (87%)
CROSS-OVER						
PubMed cohort	116	36 (31%)	103 (89%)	95 (82%)	30 (26%)	36 (31%)
Ethics cohort	29	15 (52%)	26 (90%)	24 (83%)	12 (41%)	15 (52%)
CIHR cohort	3	3 (100%)	3 (100%)	1 (33%)	2 (67%)	2 (67%)
OTHER						
PubMed cohort	20	11 (55%)	17 (85%)	12 (60%)	5 (25%)	8 (40%)
Ethics cohort	2	1 (50%)	2 (100%)	1 (50%)	1 (50%)	1 (50%)
CIHR cohort	6	6 (100%)	6 (100%)	2 (33%)	1 (17%)	6 (100%)

b) Safety outcomes

Trial design by study cohort	Total number of trials	Number of trials (%) with ≥1 safety outcome at specific levels of reporting				
		Full	Incomplete			
			Any	Partial	Qualitative	Unreported
ALL						
PubMed cohort	308	236 (77%)	196 (64%)	98 (32%)	102 (33%)	85 (28%)
Ethics cohort	69	53 (77%)	56 (81%)	19 (28%)	26 (38%)	39 (57%)
CIHR cohort	26	21 (81%)	21 (81%)	4 (15%)	8 (31%)	16 (62%)
PARALLEL GROUP						
PubMed cohort	237	193 (81%)	141 (59%)	59 (25%)	70 (30%)	65 (27%)
Ethics cohort	56	48 (86%)	46 (82%)	14 (25%)	21 (38%)	36 (64%)
CIHR cohort	18	16 (89%)	13 (72%)	3 (17%)	5 (28%)	10 (56%)
CROSS-OVER						
PubMed cohort	62	37 (60%)	48 (77%)	35 (56%)	29 (47%)	17 (27%)
Ethics cohort	12	4 (33%)	9 (75%)	4 (33%)	4 (33%)	3 (25%)
CIHR cohort	3	2 (67%)	3 (100%)	1 (33%)	1 (33%)	2 (67%)
OTHER						
PubMed cohort	9	6 (67%)	7 (78%)	4 (44%)	3 (33%)	3 (33%)
Ethics cohort	1	1 (100%)	1 (100%)	1 (100%)	1 (100%)	0
CIHR cohort	5	3 (60%)	5 (100%)	0	2 (40%)	4 (80%)

5.2.3.4. Prevalence of unreported outcomes

The average number of unreported efficacy outcomes identified per trial was lowest in the PubMed cohort. A median of 2 [IPR_{80} 1-7] efficacy and 2 [IPR_{80} 1-6] safety outcomes were omitted in each PubMed trial. Corresponding figures for Ethics trials were 4 [1-24] and 3 [1-23] unreported outcomes. CIHR trials omitted a median of 5 [1-16] efficacy and 2 [1-7] safety outcomes each.

The proportion of trials with at least one unreported efficacy outcome was also large, ranging from 33% in the PubMed cohort to 65% and 88% in the Ethics and CIHR cohorts respectively (Table 5-9a). 28% of PubMed trials, 57% of

Ethics trials, and 62% of CIHR trials omitted at least one safety outcome each (Table 5-9b).

In order to assess whether the larger proportions observed in the Ethics and CIHR cohorts were due to the availability of protocol data, the proportions were re-calculated using data from publications alone. Similar proportions of trials across cohorts had evidence of unreported outcomes in their publications (34% of PubMed trials; 29% of Ethics trials; 27% of CIHR trials).

5.2.4. Clinical importance of incompletely reported outcomes

To assess whether poor outcome reporting affects clinically important outcomes, the reporting of primary outcomes was evaluated. Only 74% of the PubMed trials presented even one fully reported, publication-defined primary outcome (Table 5-10). Thus all primary outcomes were incompletely reported in 26% of trials. In the Ethics and CIHR cohorts, 21% and 13% of trials respectively presented only incompletely reported primary outcomes. 36% of PubMed trials presented at least one incompletely reported primary outcome, compared to 27% and 16% of Ethics and CIHR trials respectively. Based on publications alone, 3% of PubMed and Ethics trials defined a primary outcome in the Methods section and then failed to report it in the Results section.

PubMed trialists were also requested to rate the clinical importance and provide the pre-specification of unreported outcomes. Survey responses for 64 PubMed trials provided data on the clinical importance of 276 (238 efficacy and 38 safety) unreported outcomes. Fourteen of these trials omitted 86 efficacy outcomes that were categorised by trialists as being highly important.

In addition, 66 Outcomes Survey responders in the PubMed cohort indicated the pre-specification of their unreported outcomes. 24% (13/54) of these trials omitted 36 primary efficacy outcomes from their publications, while 17% (3/18) omitted 9 primary safety outcomes. These data will be described in more detail in Chapter 6.

Table 5-10. Proportion of trials with at least one publication-defined primary outcome at specific levels of reporting

Trial cohort	Number of trials with primary outcomes	Number of trials (%) with ≥1 primary outcome at specific levels of reporting				
			Incomplete			
		Full	Any	Partial	Qualitative	Unreported
PubMed	232	171 (74%)	83 (36%)	78 (34%)	11 (5%)	6 (3%)
Ethics	63	50 (79%)	17 (27%)	15 (24%)	2 (3%)	2 (3%)
CIHR	45	39 (87%)	7 (16%)	7 (16%)	0	0

Finally, protocol-defined primary outcomes were omitted from publications for 26% (20/76) of Ethics trials and 13% (6/48) of CIHR trials. These discrepancies between protocols and publications will be discussed in more detail in Chapter 7.

5.3. Discussion

This chapter presented the results from a comprehensive assessment of the completeness of outcome reporting in three trial cohorts. Publications and protocols were reviewed, and trialists were surveyed in order to ascertain the prevalence of incomplete outcome reporting. The results highlight important deficiencies in the reporting of trial outcomes. The generalisability of these

findings to the overall trial population was discussed along with the cohort characteristics in Chapter 4. The remainder of this chapter will discuss the study methodology and findings within the context of the evidence currently available in the medical literature. A critical assessment of the internal validity of these results will be presented in the following chapter.

5.3.1. Previous applications of the study methodology

This thesis constitutes the first systematic assessment of outcome reporting in randomised trials. The degree of reporting for each outcome was defined objectively at four levels based on the data requirements for meta-analysis. To our knowledge, this methodology has not been previously applied to a cohort of studies.

Two approaches were applied to three cohorts of randomised trials. The PubMed approach adopted the perspective of the systematic reviewer by using information available from publications and surveys alone. The reliance on self-reported data from trialists may not have been ideal in terms of objectivity, but it is often the only means available for systematic reviewers to obtain unpublished information. The PubMed approach not only enabled an assessment of outcome reporting bias, but also provided insight into the usefulness of contacting trialists for further information about funding and unreported outcomes.

The second approach used in the thesis provided a more reliable assessment of outcome reporting deficiencies by also using study protocols. Protocols constitute a more objective source of information because they contain details

specified *a priori*. Five reviews of protocols have previously been published. A review of 75 biomedical study protocols (including 10 randomised trials) approved by a human research committee, along with 33 subsequent publications, observed that study hypotheses and details of statistical analysis were often inadequately described.[123] Consent was not required from each protocol author or sponsor. The review differed from our objectives in that it focused primarily on identifying deficiencies in the description of appropriate hypotheses and analyses, rather than on reporting of results. Protocols and publications were only compared to each other for consistency in study design. The authors observed that 13 of 33 publications reported a different design to that described in the protocol. This included one randomised trial that was published as an uncontrolled trial.

A second study compared 47 Cochrane systematic reviews to their published protocols, and surveyed authors regarding the reasons for any changes.[152] Major discrepancies in at least one protocol section were found in over 90% of reviews, with almost half making changes to pre-specified outcomes. All 65% of survey responders explained that the changes were made for pragmatic reasons.

Another study also compared Cochrane systematic reviews to their published protocols.[153] The authors found that 10/28 reviews performed subgroup analyses that were not pre-specified in the protocol.

Fourthly, a pilot study compared outcomes specified in 15 protocols to those reported in publications of clinical studies approved by a research ethics

committee.[98] The authors found a large number of discrepancies in terms of both new and omitted outcomes. A source of potential bias in the study design was the requirement for consent from investigators to examine their protocols.

Finally, a recent study examined 42 industry-funded trial protocols and reports submitted to a Swedish drug regulatory authority as part of the approval process for five selective serotonin reuptake inhibitors.[19] Despite most submitted reports describing both intention-to-treat and per-protocol analyses, corresponding journal publications selectively reported the latter analyses over the less favourable intention-to-treat analyses.

Two positive features of these previous studies were incorporated into the methods applied to the CIHR and Ethics cohorts. The first feature was the lack of requirement for consent from investigators to review their protocols, thus minimising selection bias. The second feature was the use of surveys to confirm unreported outcomes and to elicit reasons for omitting them, thus enabling the identification of potential bias in the selection of outcomes to report.

5.3.2. Contacting trialists for information on outcomes - is it useful?

Four separate surveys of trialists were conducted in this thesis - one for each of the PubMed and Ethics cohorts, and two for the CIHR cohort. Publications Surveys for the Ethics and CIHR cohorts asked trialists for a list of publications. Researchers were also asked in the same questionnaire whether any unreported outcomes existed. The Outcomes Survey for the

PubMed cohort enquired about any unreported outcomes, including their statistical significance, specification, clinical importance, and reasons for not reporting. Similar Outcomes Survey questionnaires were sent to CIHR trialists, but a few modifications were made. A list of unreported outcomes obtained from the comparison of protocols to publications was included. Trialists were asked to supplement the list and also to provide the statistical significance and reasons for omitting each outcome.

5.3.2.1. Response rates

Response rates for the single Ethics survey (55%) and the 2 CIHR surveys (90% and 74%) of trialists were relatively high, particularly for the latter cohort. Researchers in the two cohorts may have felt an obligation to respond in order to maintain good standing with an agency that could affect their future research. The authority of funding agencies and the use of frequent telephone reminders may have helped to increase response rates for CIHR trialists. Because the questionnaires were sent on behalf of ethics and funding bodies, it is difficult to interpret these response rates in the context of soliciting additional information from trialists for systematic reviews. Whether similar response rates would be obtained in surveys conducted by reviewers is unclear, although evidence suggests that response rates are similar regardless of whether an authority figure is involved. A randomised study found that response rates to requests for unpublished trial details were not significantly improved when signed by the editor of the *British Medical Journal* compared to an 'unknown' researcher.[143] The observed response rates in the trial were 72% and 26% for acquiring information on unreported methodological details and baseline data respectively, which suggests that

obtaining actual data may be more difficult than obtaining information about study characteristics.

The survey of PubMed trialists is the most relevant with regards to systematic reviews, as the sample's coverage was broad and the questionnaires were sent from an 'unknown' researcher who did not have access to trial protocols. The response rate achieved (69%) was comparable to that obtained for the Ethics cohort (55%), and may have been improved by our affiliation with a university institution.[119] Overall, over a fifth of responders provided a list of unreported outcomes in the PubMed cohort. Among these responses, almost a quarter of PubMed trialists revealed unreported primary outcomes that were not evident based on the publication alone.

Unreported data are an obvious concern for systematic reviewers, as their absence may bias the results of a meta-analysis. The results of the PubMed Outcomes Survey indicate that an adequate response rate to requests for data regarding funding, study sites, and the existence of unreported outcomes is achievable. However, it is important to note that we were writing to trialists whose studies were published recently, which is often not the case for systematic reviews. It is also unclear whether survey responders would have been willing to provide actual data for the unreported outcomes. Previous experiences suggest that obtaining aggregate data or information about unpublished studies from trialists can be very difficult, and similar drawbacks would be expected when requesting unreported data in published trials.[143,144,154,155]

5.3.2.2. Response bias

Although relatively high response rates were achieved in the Publications and Outcomes Surveys in all cohorts, there was strong potential for response bias due to the potentially sensitive nature of the information solicited from trialists. Based on trial characteristics, non-responders differed from responders primarily in sources of funding and study size. Authors of PubMed trials funded solely by industry sources were less likely to respond compared to those whose trials had partial or no industry funding. This finding is consistent with previous observations that industry-funded researchers may be less willing or unable to offer data from their studies.[154,156-158] Potential reasons for non-response from industry-funded trials include the protection of commercial interests, the inability to access industry-held data, and the reluctance to reveal biased practices. A few trialists in our cohorts indicated that they were unable to provide answers because the data were held by sponsors.

CIHR responders were more likely to have conducted larger, multicentre trials compared to non-responders. A possible explanation is that smaller sample sizes may be a marker for poor trial design and conduct, which could be associated with poorer quality of reporting.

In addition to these measurable factors, non-responders likely differed from responders in several other respects.[154] Trialists who did not reply to the survey may have been more busy than responders, or may have had fewer resources available for returning their response. One trialist commented by e-mail that her department did not provide funds for international faxes or postage. Non-responders may also have been more reluctant to retrieve

archived data for older trials, may have understood less English, or may have been less willing to admit to poor research practices.

The last reason is probably the most relevant to our assessment of outcome reporting bias. Trialists who selectively report outcomes may not wish to admit to this practice and may therefore be less likely to respond - or even worse, they may respond with inaccurate details of their study conduct. Trialists may also be less inclined to indicate that a lack of statistical significance was the primary reason for not reporting an outcome if they were aware that this was poor research practice.

Several observations suggest that survey responses were often unreliable in terms of under-reporting the existence of omitted outcomes. Firstly, when evidence in the publications or protocols revealed that some outcomes had been omitted, and before we indicated to Ethics or CIHR trialists that their protocols had been reviewed, 25%-85% of responders in the three trial cohorts incorrectly stated that all outcomes had been reported. It is possible that some trialists simply could not recall any unreported outcomes if the trial had been conducted several years earlier. Higher rates of discrepancies were observed in Ethics and CIHR cohorts because protocols were available to identify unreported outcomes more reliably than in the PubMed cohort. Secondly, among 12 CIHR trials with unreported outcomes that, according to trialists, were omitted because they were "not intended for inter-group comparisons", one half had explicitly specified these variables as secondary outcomes in the protocol. Finally, it is notable that none of the CIHR trialists supplemented the questionnaire list with additional unreported outcomes that

we had not already identified from their protocols. Whether this was due to none existing, poor recall, or a reluctance to reveal further omissions is unclear.

One PubMed trialist from a school of pharmacy stated in his Outcomes Survey response that "your goals are difficult, as authors who do leave data out of their publications are most commonly found in the pharmaceutical industry. Therefore to receive impartial comment from them is likely to be difficult." This obstacle to soliciting potentially self-incriminating information is not only applicable to industry and their commercial interests, but it is also a potential concern for non-industry researchers with their academic and career interests.[158,159]

It is evident that despite high response rates, the reliability of self-reported data with regards to identifying unreported outcomes was inadequate for a significant proportion of responders, thereby limiting their usefulness in systematic reviews. However, several unreported primary outcomes were identified through survey responses, which justifies continuing efforts to solicit this type of information from trialists.

5.3.3. Characteristics of trial outcomes

A large number of outcomes were measured and analysed in the three cohorts of randomised trials, which raises issues about multiplicity and spurious findings. The median numbers of total outcomes per trial ranged from 15 in the PubMed cohort, to 24 and 26 in the Ethics and CIHR cohorts respectively. The differences between PubMed trials and the other two

cohorts may partly reflect the availability of study protocols to identify additional unreported outcomes for Ethics and CIHR trials. However, even when restricted to outcomes reported in publications, the median number of outcomes per trial was higher in Ethics and CIHR cohorts (21 outcomes per trial for both cohorts) compared to the PubMed cohort (15 outcomes per trial), suggesting that other factors exist. Potential factors include the tendency for Ethics and CIHR trials to have larger sample sizes and multiple study centres, which may be associated with more outcome measurements.

The number of outcomes observed per trial across the three cohorts is higher than in previous reviews restricted to specialty fields[97] and high impact journals,[92] which found a median of 8 total outcomes and 6 efficacy outcomes per trial respectively. These reviews relied on publications only, and differences in the number of outcomes may reflect the inclusion of additional outcomes identified by protocols and Outcomes Surveys in the thesis cohorts. However, even if we restrict our cohorts to outcomes listed in the publications alone, the median number of outcomes remains high (15-21 outcomes per trial). Secondly, additional outcomes were identified in our cohorts through literature searches for multiple trial publications. A third explanation could involve differences in defining individual outcomes. For example, Gøtzsche[97] did not count variables measured at multiple time points as separate outcomes. Finally, the two previous reviews were conducted over a decade ago, and the number of outcomes measured in trials may have increased over time.

Despite being higher than in previous reviews of restricted samples, the number of outcomes observed in each of the three trial cohorts is likely an underestimate of the true number of measured outcomes. Firstly, some unreported outcomes would not have been identified due to non-response to the Outcomes Survey. Even when protocols were available, unreported outcomes may have been missed if they had not been specified in the protocol. Secondly, survey responders likely under-reported the number of omitted outcomes. A large proportion of investigators (25%-85%) denied the existence of unreported outcomes despite clear evidence from the publications or protocols indicating otherwise. Finally, some unreported outcomes that were vaguely described in the protocols of Ethics and CIHR trials were not counted as outcomes. As a result, our estimates represent the lower limit of the true number of trial outcomes. A sensitivity analysis including the vague outcomes produced a slight increase in the median number of outcomes per trial in the Ethics cohort, but did not affect the CIHR trials.

Trials appear to focus primarily on measuring efficacy rather than safety data, as suggested by the larger proportion of trials recording efficacy outcomes and the much higher median number of efficacy outcomes in each trial. In addition, safety outcomes were rarely specified as primary outcomes. The discrepancy in importance given to efficacy and safety outcomes is consistent with previous studies demonstrating that the amount of space allocated to safety outcomes in journal reports is inadequate.[120,160]

The proportion of trials with at least one statistically significant efficacy outcome was similar across cohorts (79%-88%), as was the proportion of trials with at least one significant primary outcome (49%-60%). These figures are compatible with previous reviews of medical journals.[28,31-33] The consistency observed between the three thesis cohorts was unexpected given the differences in statistical power resulting from much larger sample sizes in Ethics and CIHR trials compared to PubMed trials. It is possible that the comparable prevalence of significant efficacy outcomes in the PubMed cohort is due to a higher degree of study publication bias or a greater amount of data dredging to produce a significant result in the smaller trials. Another possible but unlikely explanation is that PubMed trialists were more likely to have investigated interventions that produced large effect sizes, which compensated for the lack of power.

For safety outcomes, PubMed trials were less likely to report a significant outcome compared to the Ethics and CIHR trials (35% versus 49%-50%). This is consistent with the lack of power in PubMed trials relative to the other two cohorts.

Finally, the number of publication-defined primary outcomes per trial was often greater than the 1 or 2 main outcomes generally recommended by guidelines such as the revised CONSORT statement[161] and the International Conference on Harmonisation guidelines.[129] As further confirmation of their high methodological quality, less than 10% of the CIHR trials defined more than 2 primary outcomes in their reports, compared to 27% and 38% of Ethics

and PubMed trials. The use of multiple primary outcomes can present difficulties in interpretation if only some are statistically significant.

5.3.4. Prevalence of incomplete outcome reporting

The first objective of this thesis was to assess the prevalence of incomplete outcome reporting across broad trial cohorts. By combining a review of trial publications and protocols with a survey of investigators, the number of reported and unreported outcomes per study was ascertained. Four primary levels of outcome reporting were defined with respect to the amount of data required for meta-analysis. These levels consisted of fully, partially, qualitatively, and unreported outcomes.

Two additional composite categories of reporting were defined. Reported outcomes consisted of those for which some data were presented in the trial reports. These encompassed fully, partially, and qualitatively reported outcomes. Outcomes with insufficient data presented for meta-analysis were classified as incompletely reported. These included partially, qualitatively, and unreported outcomes. It is important to distinguish incompletely reported outcomes, which encompassed the three lowest primary reporting levels, from the lowest level of unreported outcomes, for which no data were presented in trial reports.

5.3.4.1. Incomplete reporting of outcomes across all trials

Incomplete reporting of outcomes can have a substantial effect on meta-analyses by prohibiting the inclusion of particular outcomes. Our review of three trial cohorts demonstrated that incomplete reporting was common, even

in the cohort of higher quality CIHR trials. A median of 31%-48% of efficacy outcomes per trial were incompletely reported for meta-analysis, compared to 50%-60% of safety outcomes. The majority of trials did not fully report all of their efficacy (75%-96%) and safety (64%-81%) outcomes. Continuous outcomes were more often reported incompletely compared to binary outcomes, which would be expected due to the relative simplicity of providing event rates by group for binary outcomes. The results were robust to sensitivity analyses that excluded trials with fewer than 5 efficacy outcomes.

As discussed in Chapter 3, excluding variables that were vaguely-described in the protocol and unreported in the publications would under-estimate the degree of incomplete reporting, whereas incorporating them would provide an over-estimate. Our main analysis excluded these outcomes in order to be conservative. Including them in a sensitivity analysis increased the proportions of incompletely reported outcomes in the Ethics cohort overall from 48% to 53% for efficacy and from 60% to 67% for safety outcomes. For parallel group CIHR trials, the proportions were increased from 33% to 38% for efficacy and from 24% to 33% for safety outcomes.

5.3.4.2. Incomplete reporting of outcomes by study design, journal type, and funding source

Among the various study designs, parallel group trials had the lowest median proportion of incompletely reported outcomes per trial, ranging from 22%-35% for efficacy outcomes and 24%-50% for safety outcomes across the three study cohorts. The adoption of evidence-based reporting guidelines such as the revised CONSORT statement[134] for parallel group trials should help to

reduce poor outcome reporting in journal publications. The guidelines assert that "for each primary and secondary outcome, a summary of results for each group, and the estimated effect size and its precision" should be reported.[134]

The trials reviewed in this thesis were mostly published after the original CONSORT statement was introduced in 1996.[162] Previous studies have observed an improvement in the reporting of methodological details for trials published primarily in high-impact journals after the adoption of the original CONSORT statement.[163-165] However, these studies also showed that the overall quality of reporting remained inadequate. Only one of the reviews assessed the reporting of results, and found that a mean of approximately 7 checklist items were reported out of a possible 10.[165] The deficiencies would have presumably been worse in lower impact journals. Although the quality of trial reporting may have improved further since the introduction of the revised CONSORT guidelines in 2001, the potential impact of these standardised guidelines will be limited if journal editors do not strictly enforce adherence to them. Our results indicate that adherence to the CONSORT guidelines in terms of fully reporting the study outcomes was poor among trialists in the three cohorts.

Although parallel group trials exhibited a relatively high prevalence of incompletely reported outcomes, the situation was even worse for cross-over trials. A median of 58%-100% of efficacy and 71%-100% of safety outcomes were incompletely reported in these trials. Although an extension to the CONSORT guidelines is currently being prepared for cross-over trials [Douglas Altman, personal communication, 2003], standardised reporting

guidelines are not yet readily available for this study design. As a result, outcome reporting deficiencies may stem from a lack of awareness with regards to the type of information that should be presented for analysis of paired data. Authors may feel that, as with parallel group trials, providing overall group numbers, means, and standard deviations is sufficient for continuous outcomes. Similarly, for binary outcomes, independent proportions for each study group may be perceived as sufficient. However, this type of data is inadequate for both meta-analysis and the interpretation of individual cross-over trial results. Either raw paired data for each study participant, an exact *p*-value from a paired analysis, or the precision of the mean difference is required.[166]

The proportion of incompletely reported efficacy outcomes per trial varied across journal types and sources of funding. PubMed trials with full industry funding had a much higher proportion of incompletely reported safety outcomes, while CIHR trials that were published in specialty journals had higher proportions of incompletely reported efficacy and safety outcomes. Trials published in general medical journals are likely to be subjected to more rigorous peer review, resulting in fewer outcome reporting deficiencies. However, the reporting of safety outcomes in prominent general medical journals has previously been shown to provide inadequate data for meta-analysis in 38% of trials.[160]

With regards to sources of funding, authors of trials sponsored fully by industry sources may be more willing to provide a generic statement claiming no difference in safety profiles between interventions, rather than providing full

details that could be interpreted unfavourably for the sponsored drug. Among industry submissions to drug licensing agencies in Finland and Sweden, reports of unpublished trials were found to contain significantly more safety data than those of published trials, suggesting that safety data may often by suppressed in publications.[74]

5.3.4.3. Prevalence of unreported outcomes

The observed proportions of incompletely reported outcomes are likely an underestimate due to under-reporting of omitted outcomes by trialists in the Outcomes Surveys, as well as the existence of vague unreported variables that were not recorded as outcomes. Despite this under-reporting, complete omission of outcomes from trial reports was still commonly observed in the Ethics and CIHR cohorts, occurring in three-quarters of trials for efficacy outcomes and over one half of trials for safety outcomes. Unreported outcomes accounted for over 20% and 30% of the total number of efficacy and safety outcomes respectively. A median of 4-5 efficacy outcomes and 2-3 safety outcomes per trial were omitted from publications. Unreported outcomes were identified less frequently in the PubMed cohort, which can be viewed as the least accurate estimate due to the reliance on self-reported data rather than protocols to identify omitted outcomes. The Ethics and CIHR cohorts thus provide a more valid estimate of the prevalence of unreported outcomes. As stated accurately in a survey response from a trialist at a prominent institution who published his trial in a general medical journal, "there is always info in a study that doesn't get into the final manuscript, in my mind."

A previous review of 196 non-steroidal anti-inflammatory drug (NSAID) trials revealed that 30% of reports mentioned outcomes in the Methods but not Results sections.[97] To our knowledge, no other review has commented on this type of discrepancy within studies. The review of NSAID trials is consistent with the thesis results if survey data are excluded, with 27%-34% of trial reports containing evidence of unreported outcomes in the three cohorts. This consistency across thesis cohorts supports the assertion that the lower prevalence of unreported outcomes identified in the PubMed cohort is due to the absence of protocol data rather than superior trial reporting.

5.3.5. Clinical importance of incompletely reported outcomes

Although our results have shown that outcome reporting deficiencies are common, it is important to evaluate their potential clinical impact. Critics may argue that incompletely reported outcomes are likely to be those that are not clinically important, and their absence may therefore be inconsequential in a systematic review. However, among trials with primary outcomes defined in their publications, 16%-36% across the three cohorts contained incompletely reported primary outcomes. In addition, at least one primary outcome was omitted by a quarter of the Outcomes Survey responders who revealed unreported outcomes in the PubMed cohort. Furthermore, 26% of Ethics trials and 13% of CIHR trials omitted a protocol-defined primary outcome from their publications. It is therefore evident that incomplete reporting affects clinically important outcomes.

5.4. Chapter summary

The reliability of a systematic review depends to a large extent on the identification and availability of all existing outcome data. We used protocols, publications and trialist surveys to assess the completeness of outcome reporting in three broad cohorts of randomised trials. Incomplete reporting of outcomes prevents their inclusion in meta-analyses, and introduces the potential for bias.

Based on our results, it is evident that outcome reporting in trial publications is often inadequate for meta-analysis, with a median of 22%-35% of efficacy and 24%-50% of safety outcomes being incompletely reported per parallel group trial. The deficiencies were worse for cross-over trials. When protocol data were available, complete omission of efficacy and safety outcomes was identified in 65%-88% and 57%-62% of trials respectively. The high prevalences of incompletely reported and unreported outcomes observed across trial cohorts represent conservative estimates of outcome reporting deficiencies due to the potential for under-reporting by trialists. Incompletely reported outcomes were frequently considered to be clinically important by the investigators themselves, which highlights potentially serious implications for systematic reviews.

Chapter 6 - Bias in the reporting of trial outcomes

6.1. Introduction

The previous chapter highlighted deficiencies in the completeness of outcome reporting in publications of randomised trials. Data provided for both efficacy and safety outcomes were frequently insufficient for meta-analysis. Because of potential difficulties in obtaining data beyond what is provided in trial publications, it is necessary to ascertain whether incomplete outcome reporting occurs independently of trial results. Data missing at random may reduce the global amount of knowledge available in the medical literature - hence reducing precision - but would not bias the results of a meta-analysis in any particular direction. Data that are selectively suppressed *post hoc* based on their analysis, however, may result in a preponderance of statistically significant results in the literature.

This chapter will evaluate whether the selection of outcomes to fully report in trial publications is associated with potential bias in relation to statistical significance. Reasons for omitting outcomes will be assessed, and a quantitative analysis of outcome reporting bias will be presented. The chapter will conclude with a discussion of the limitations of data collection and analysis for our assessment of outcome reporting.

6.2. Characteristics of unreported outcomes

6.2.1. Statistical significance

As part of the Outcomes Survey, trialists in the PubMed and CIHR cohorts were asked to provide the statistical significance of unreported outcomes. Among outcomes for which this information was provided, 66% (167/253) of unreported efficacy outcomes and 96% (48/50) of unreported safety outcomes were non-significant in the PubMed cohort. 99% (99/100) of omitted outcomes (84 efficacy and 16 safety) in the CIHR cohort were non-significant.

Table 6-1. Number of trials with at least one unreported outcome at various ratings of clinical importance according to 64 Outcomes Survey responses in the PubMed cohort

Clinical importance	Proportion of trials among responders (%)[a]	
	Efficacy outcomes n=53 trials	Safety outcomes n=16 trials
High	14 (26%)	0 (0%)
Moderate	26 (49%)	5 (31%)
Low	29 (55%)	13 (81%)

[a]Denominator corresponds to trials with unreported outcomes that provided survey data on clinical importance

6.2.2. Clinical importance

PubMed trialists were also asked to rate the clinical importance of unreported outcomes as high/moderate/low, and to indicate their pre-specification as primary or secondary outcomes. As mentioned briefly in Chapter 5, survey responses for 64 PubMed trials provided data on the clinical importance of 276 (238 efficacy and 38 safety) unreported outcomes (Table 6-1). One quarter of these 64 trials omitted 86 efficacy outcomes that were categorised as having high clinical importance. According to non-mutually exclusive

reasons provided by trialists, these important outcomes were omitted due to space constraints (n=6 trials), planned submission or ongoing analysis (n=4 trials), lack of statistical significance (n=3 trials), and a lack of clinical importance (n=2 trials). Among unreported efficacy outcomes with known statistical significance, those rated as having low importance were less likely to be significant compared to those classified as highly important (18% [14/79 outcomes] versus 43% [37/86 outcomes]). All 38 unreported safety outcomes were classified as having low or moderate clinical importance; 13 trialists provided the statistical significance of 30 omitted safety outcomes, all of which were non-significant.

Table 6-2. Number of trials with at least one unreported outcome specified as primary, secondary, or neither according to 66 Outcomes Survey responses for the PubMed cohort

Specification	Proportion of trials among responders (%)[a]	
	Efficacy outcomes n=54 trials	Safety outcomes n=18 trials
Primary	13 (24%)	3 (17%)
Secondary	28 (52%)	8 (44%)
Unspecified	22 (41%)	8 (44%)

[a]Denominator corresponds to trials with unreported outcomes that provided survey data on outcome specification

Another indication of clinical importance was the pre-specification of unreported outcomes. As described briefly in Chapter 5, 66 Outcomes Survey responders in the PubMed cohort indicated the pre-specification of their unreported outcomes (Table 6-2). These 66 responders included the 64 trials with clinical importance ratings in Table 6-1. 24% (13/54) of these trials omitted 36 primary efficacy outcomes from their publications, while 17% (3/18) omitted 9 primary safety outcomes. Among all primary outcomes with known

statistical significance, unreported primary outcomes were less likely to be statistically significant compared to those that were reported in publications (26% [11/43] versus 47% [421/901]). The non-mutually exclusive reasons given by trialists for omitting these outcomes included planned submission or ongoing analysis (n=7 trials), space constraints (n=6 trials), lack of statistical significance (n=4 trials), and lack of clinical importance (n=3 trials).

6.2.3. Reasons for omitting outcomes from trial publications

Among Outcomes Survey responses with unreported outcomes, 69/75 (92%) and 34/46 (74%) provided reasons for omitting either efficacy or safety outcomes in the PubMed and CIHR cohorts respectively (Table 6-3). A lack of clinical importance was commonly cited as a reason for omitting efficacy outcomes in both PubMed (37%) and CIHR (62%) cohorts. 24% (14/59) of PubMed trials and 45% (13/29) of CIHR trials omitted efficacy outcomes due to a lack of statistical significance. In 29% (4/14) and 46% (6/13) of these trials respectively, a lack of clinical importance was also cited as a concurrent reason for omitting efficacy outcomes. While the most common reason in PubMed trials was journal space restrictions (47%) imposed by authors (36%) and journals (19%), this justification was not as prominent in CIHR trials (17%). For over 20% of trials in both cohorts, authors were planning to report these efficacy outcomes in future manuscripts, and 7%-17% had outcomes that were still undergoing analysis.

With regards to safety outcomes, the most common reasons for omitting them in PubMed and CIHR trials were a lack of clinical importance (75% and 45% of trials respectively) and a lack of statistical significance (50% and 45% of

trials respectively) (Table 6-3). A lack of both statistical and clinical significance were given concurrently as reasons for omitting an outcome in 50% (4/8) of PubMed trials and 40% (2/5) of CIHR trials. A quarter of trials in both cohorts omitted safety outcomes because of space constraints imposed by authors or journals.

In reference to the list of unreported outcomes that we identified from CIHR protocols, trialists stated that 46 efficacy outcomes were unreported for 11/29 (38%) trials because they were not actually intended for between-group analyses (Table 6-3). However, 24 (52%) of these outcomes were specified explicitly as secondary outcomes in the protocols. One safety outcome was not intended for inter-group comparisons.

Table 6-3. Reasons for not reporting one or more outcomes per trial based on survey responses in PubMed and CIHR cohorts

a) PubMed cohort (n=69 responses with unreported efficacy or safety outcomes)

	Number of trials (%)	
Reason[a]	**Efficacy outcomes (n = 59 trials)**	**Safety outcomes (n = 16 trials)**
Space constraints	28 (47%)	4 (25%)
Journal-imposed	11 (19%)	1 (6%)
Author -imposed	21 (36%)	3 (19%)
Not clinically important	22 (37%)	12 (75%)
p>0.05	14 (24%)	8 (50%)
Not yet submitted	13 (22%)	1 (6%)
Not yet analysed	10 (17%)	1 (6%)

[a]Reasons are not mutually exclusive (>1 may have been given per trial)

b) CIHR cohort (n=34 responses with unreported efficacy or safety outcomes)

	Number of trials (%)	
Reason[a]	**Efficacy outcomes (n=29 trials)**	**Safety outcomes (n=11 trials)**
Space constraints	5 (17%)	3 (27%)
Journal-imposed	3 (10%)	2 (18%)
Author -imposed	2 (7%)	1 (9%)
Not clinically important	18 (62%)	5 (45%)
p>0.05	13 (45%)	5 (45%)
Not yet submitted	6 (21%)	1 (9%)
Not yet analysed	2 (7%)	0 (0%)
Non-outcome[b]	11 (38%)	1 (9%)
Data not collected	6 (21%)	1 (9%)
Low event rates	2 (7%)	2 (18%)

[a]Reasons are not mutually exclusive (>1 may have been given per trial)
[b]Not intended for between-group comparisons

6.3. Quantitative analysis of outcome reporting bias

6.3.1. Review of methods

In order to calculate the odds ratio for outcome reporting bias (defined as the ratio of odds for a fully reported outcome being statistically significant relative to the odds for an incompletely reported outcome being significant), the level of reporting was dichotomised as fully reported versus incompletely reported. Statistical significance was dichotomised at the traditional $p<0.05$ level. Odds ratios relating completeness of reporting to statistical significance in each trial were pooled using a random effects meta-analysis (DerSimonian and Laird method).[121]

Trials were excluded from the calculation of odds ratios if they had no outcomes in adjacent cells in the 2x2 table, or if they had no outcomes with known statistical significance. Adjacent 'zero' cells would occur if all trial

outcomes were either fully or incompletely reported, or if all outcomes were either statistically significant or non-significant.

Level of reporting	$p<0.05$	$p \geq 0.05$
Full	12	15
Partial	2	0
Qualitative	0	3
Unreported	0	2

Collapse

Level of reporting	$p<0.05$	$p \geq 0.05$
Full	12	15
Incomplete	2	5

Odds ratio=2.0

Figure 6-1. Example of dichotomising the level of outcome reporting to calculate the odds ratio for outcome reporting bias

An anonymised example from the trial cohorts will help to illustrate the described analysis. A large factorial trial co-sponsored by industry and government sources measured 34 efficacy outcomes and no safety outcomes. Final results for 32 outcomes were presented in two publications, one of them being in a well-known journal. Figure 6-1 shows the outcomes stratified by level of reporting and statistical significance. Twenty-four survival outcomes were fully reported as hazard ratios with 95% confidence intervals, half of which were statistically significant. Another three non-significant continuous outcomes were fully reported with mean changes from baseline and exact *p*-values. Two significant continuous outcomes were recorded as partially reported because they were reported in a graph without error bars or exact *p*-values. Three continuous qualitatively reported outcomes were described as not having differed significantly between groups. Finally, two non-significant outcomes were omitted from the trial reports. One of the unreported outcomes had been defined as a primary outcome in the protocol. The trialist explained that both outcomes were to be reported in future publications. The

four levels of reporting were collapsed to form two levels by combining partially, qualitatively, and unreported outcomes. The final 2x2 table consisted of 27 fully reported outcomes, 12 of which were significant, and 7 incompletely reported outcomes, 2 of which were significant. The resulting odds ratio for outcome reporting bias is 2.0.

6.3.2. Association between level of reporting and statistical significance

Across the three trial cohorts, the pooled odds ratio for outcome reporting bias ranged from 2.0-2.7 for efficacy outcomes and 1.9-7.7 for safety outcomes respectively (Table 6-4). The odds ratios were significantly greater than 1 in all instances except for safety outcomes in 4 CIHR trials. Stratifying by study design yielded similar results for parallel group trials (Table 6-4). The confidence intervals for bias in cross-over trials were wide because there were few such studies.

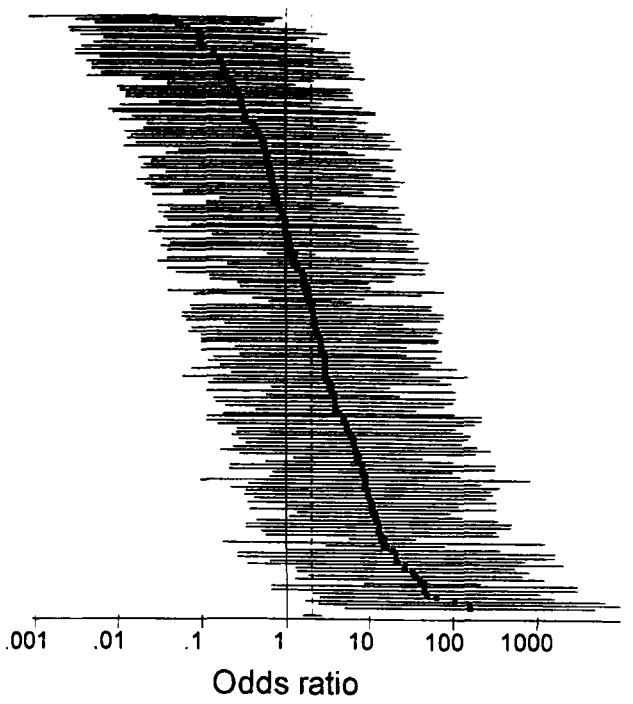
Table 6-4. Pooled odds ratio for outcome reporting bias (fully versus incompletely reported outcomes) by study design and cohort

Trial design	Efficacy outcomes		Safety outcomes	
	n trials[a]	OR[b] (95% CI)	n trials[a]	OR[b] (95% CI)
ALL				
PubMed cohort	161	2.0 (1.6 - 2.7)	43	1.9 (1.1 - 3.5)
Ethics cohort	45	2.4 (1.4-4.1)	17	4.4 (1.7-11.7)
CIHR cohort	30	2.7 (1.5-5.0)	4	7.7 (0.53-111)
PARALLEL GROUP				
PubMed cohort	135	1.9 (1.4 - 2.6)	33	3.1 (1.6 - 5.8)
Ethics cohort	33	3.0 (1.6-5.8)	16	5.0 (1.8-13.9)
CIHR cohort	24	2.4 (1.1-5.1)	2	43.4 (2.9-662)
CROSS-OVER				
PubMed cohort	22	2.7 (1.4 - 5.3)	9	0.32 (0.10 - 1.1)
Ethics cohort	11	1.4 (0.60-3.1)	0	N/A
CIHR cohort	2	1.6 (0.21-11.7)	1	0.25 (0.01-8.6)

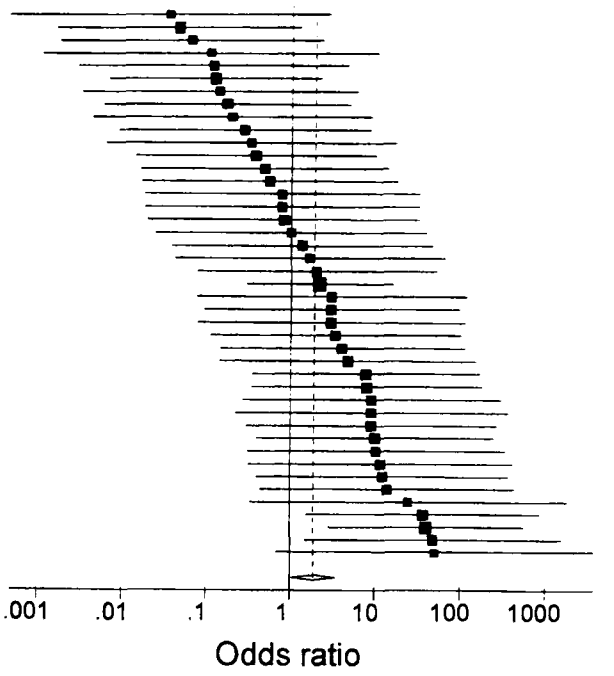
[a]Trials were excluded if odds ratio could not be calculated due to adjacent 'zero' cells in the 2x2 table
[b]Odds ratio >1 signifies that fully reported outcomes have a higher odds of being statistically significant ($p<0.05$) compared to incompletely reported outcomes

Individual odds ratios varied greatly within each cohort (Figure 6-2). Because of small numbers of outcomes in some trials, it may have been necessary to add 0.5 to all cells (done by software) in order to accommodate empty cells in the 2x2 table; it was thus possible to obtain extreme odds ratios. The inner-80-percentile ranges for efficacy outcomes were 0.2-19 in 161 PubMed trials, 0.33-27 in 45 Ethics trials, and 0.13-18 in 30 CIHR trials. For safety outcomes, the inner-80-percentile ranges were 0.12-25 in 43 PubMed trials, 0.33-93 in 17 Ethics trials, and 0.25-164 in 4 CIHR trials. Using the chi-square test, considerable heterogeneity between trial estimates of outcome reporting bias was observed for efficacy outcomes in the PubMed (p=0.001) and Ethics (p=0.002) cohorts, but not for safety outcomes (p=0.099 and p=0.093 respectively). Significant heterogeneity between CIHR trials was not observed for efficacy (p=0.154) or safety (p=0.074) outcomes.

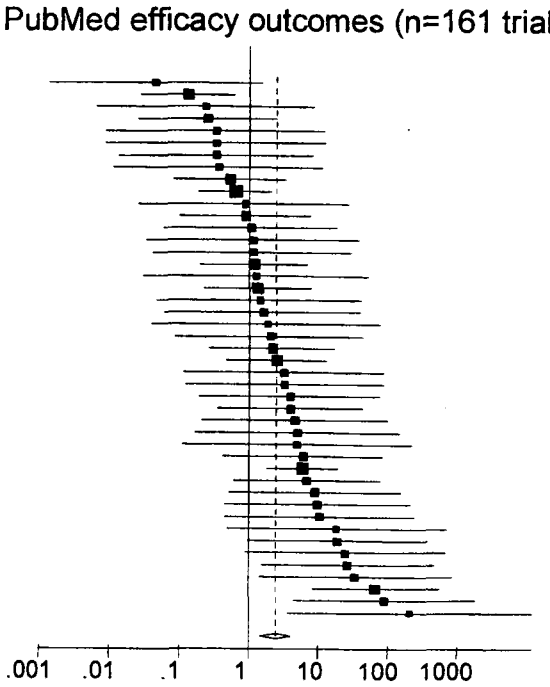
Outcomes Survey data were not available for the Ethics cohort, and unreported outcomes could therefore not be included in the odds ratio calculation because their statistical significance was unknown. In order to assess the potential impact of this missing data, odds ratios were re-calculated for the PubMed and CIHR cohorts with the unreported outcomes excluded. The odds ratios were virtually unchanged in the PubMed cohort, and reduced in the CIHR cohort. For efficacy outcomes across all study designs, the odds ratios were 2.1 (95% CI 1.6-2.9) in 143 PubMed trials, and 2.2 (95% CI 1.0-4.6) in 24 CIHR trials. For safety outcomes, the odds ratios were 1.9 (95%CI 1.0-3.5) in 41 PubMed trials, and 5.8 (95% CI 0.53-64) in 4 CIHR trials.



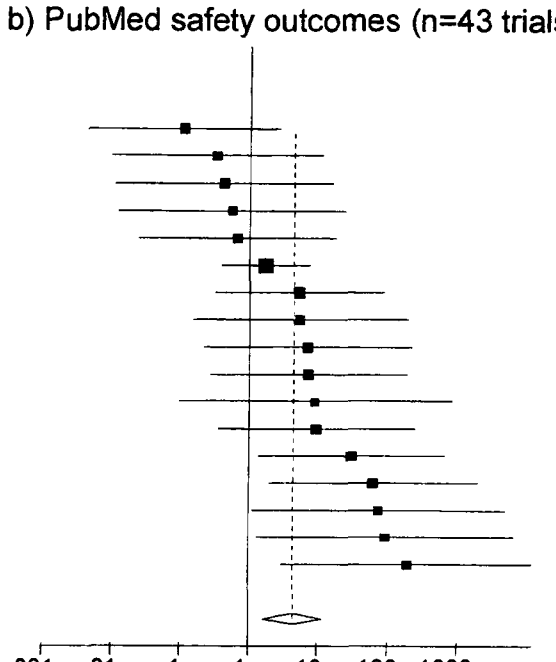
a) PubMed efficacy outcomes (n=161 trials)



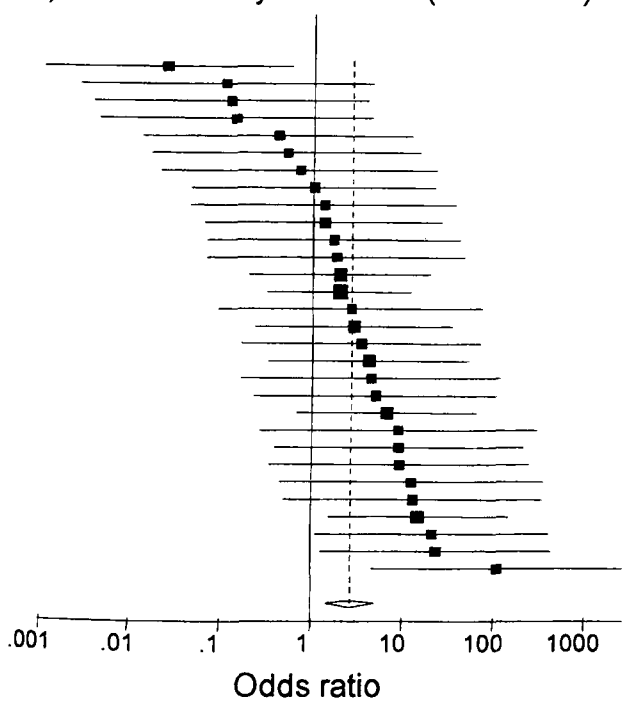
b) PubMed safety outcomes (n=43 trials)



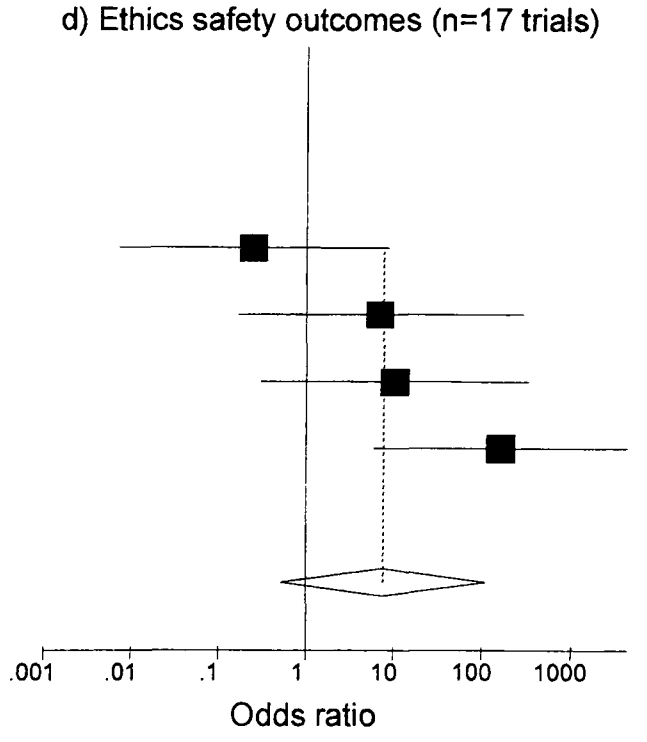
c) Ethics efficacy outcomes (n=45 trials)



d) Ethics safety outcomes (n=17 trials)



e) CIHR efficacy outcomes (n=30 trials)



f) CIHR safety outcomes (n=4 trials)

Figure 6-2. Forest plots for the odds ratio of outcome reporting bias (fully versus incompletely reported outcomes) by efficacy and safety in three trial cohorts

The specific reasons for excluding trials from each cohort are detailed in Table 6-5. Trials were more commonly excluded due to all outcomes being fully reported as opposed to all being incompletely reported. Half of the excluded PubMed trials had only fully reported efficacy outcomes (rows 1,5,6), compared to 37% with only incompletely reported outcomes (rows 2,7,8) (Table 6-5a). More than twice as many Ethics trials were excluded with all fully reported efficacy outcomes (59%) relative to trials excluded with all incompletely reported efficacy outcomes (26%). Almost three-quarters of excluded CIHR trials had efficacy outcomes that were all fully reported, while no CIHR trials were excluded for having all incompletely reported efficacy outcomes. Trials in the three cohorts were also much more likely to be excluded with all statistically non-significant outcomes (rows 4,6,8) rather than all significant ones (rows 3,5,7) (Table 6-5a).

Similar trends were observed for exclusions from the analysis of safety outcomes (Table 6-5b). Trials were more commonly excluded because they fully reported all safety outcomes with known statistical significance (rows 1,5,6), as opposed to having no fully reported outcomes (rows 2,7,8). Trials were also more frequently excluded for having all non-significant outcomes (rows 4,6,8) compared to only significant outcomes (rows 3,5,7).

The characteristics of excluded trials were generally similar to those of included trials in the efficacy and safety analysis of bias (Table 6-6). The main differences were a higher number of eligible outcomes and a lower proportion of cross-over trials among included studies.

Table 6-5. Reasons for exclusion of trials from the calculation of odds ratios for reporting bias

a) Eligible efficacy outcomes with known statistical significance

Row	REASON	Number (%) of excluded trials		
		PubMed Cohort (n=344)	Ethics Cohort (n=54)	CIHR Cohort (n=18)
1	All outcomes were fully reported	97 (28%)	21 (39%)	9 (50%)
2	All outcomes were incompletely reported	87 (25%)	11 (20%)	0
3	All outcomes were significant ($p<0.05$)	5 (1%)	1 (2%)	0
4	All outcomes were non-significant ($p\geq0.05$)	36 (10%)	6 (11%)	5 (28%)
5	All outcomes were fully reported and significant	33 (10%)	4 (7%)	3 (17%)
6	All outcomes were fully reported and non-significant	42 (12%)	7 (13%)	1 (6%)
7	All outcomes were incompletely reported and significant	16 (5%)	1 (2%)	0
8	All outcomes were incompletely reported and non-significant	25 (7%)	2 (4%)	0
9	Unknown statistical significance of all efficacy outcomes	3 (1%)	1 (2%)	0

b) Eligible safety outcomes with known statistical significance

Row	REASON	Number (%) of excluded trials		
		PubMed Cohort (n=344)	Ethics Cohort (n=54)	CIHR Cohort (n=18)
1	All outcomes were fully reported	27 (10%)	9 (17%)	5 (23%)
2	All outcomes were incompletely reported	11 (4%)	3 (6%)	0
3	All outcomes were significant ($p<0.05$)	0	0	0
4	All outcomes were non-significant ($p\geq0.05$)	52 (20%)	6 (12%)	2 (9%)
5	All outcomes were fully reported and significant	22 (8%)	4 (8%)	4 (18%)
6	All outcomes were fully reported and non-significant	91 (34%)	17 (33%)	6 (27%)
7	All outcomes were incompletely reported and significant	6 (2%)	0	0
8	All outcomes were incompletely reported and non-significant	37 (14%)	6 (12%)	5 (23%)
9	Unknown statistical significance of all safety outcomes	19 (7%)	7 (13%)	0

Table 6-6. Characteristics of trials included and excluded from the meta-analysis of odds ratios for outcome reporting bias

a) Efficacy outcomes

COHORT	Design[a]		General journal (%)	Non-commercial or no funding (%)	Multicentre (%)	Median sample size of parallel group trials [IPR_{80}][b]	Median number of eligible outcomes[c] [IPR_{80}][b]
	Parallel (%)	Cross-over (%)					
PUBMED							
161 trials included	84%	14%	7%	50%	32%	100 [27-414] n=135 trials	15 [5-44]
344 trials excluded	70%	26%	8%	52%	24%	74 [25-341] n=240 trials	8 [2-25]
ETHICS							
45 trials included	73%	24%	0	27%	49%	147 [29-618] n=33 trials	23 [6-69]
54 trials excluded	65%	33%	6%	28%	48%	182 [31-1479] n=35 trials	11 [3-28]
CIHR							
30 trials included	80%	7%	43%	63%	67%	278 [64-2088] n=24 trials	25 [9-49]
18 trials excluded	83%	6%	44%	50%	67%	393 [48-2143] n=15 trials	10 [3-21]

[a]Other designs not included in table
[b]Inner-80-percentile range (10^{th}-90^{th} percentiles)
[c]Outcomes for which statistical significance was known

b) Safety outcomes

COHORT	Design[a]		General journal (%)	Non-commercial or no funding (%)	Multicentre (%)	Median sample size of parallel group trials [IPR_{80}][b]	Median number of eligible outcomes[c] [IPR_{80}][b]
	Parallel (%)	Cross-over (%)					
PUBMED							
43 trials included	77%	21%	16%	20%	57%	132 [40-653] n=33 trials	9 [5-27]
265 trials excluded	77%	20%	7%	41%	29%	95 [25-420] n=203 trials	2 [1-11]
ETHICS							
17 trials included	94%	0	0	12%	59%	224 [29-1000] n=16 trials	12 [4-49]
52 trials excluded	77%	23%	6%	10%	65%	225 [31-1270] n=40 trials	2 [0-11]
CIHR							
4 trials included	50%	25%	50%	50%	25%	2255 [135-4374] n=2 trials	9 [5-23]
22 trials excluded	73%	9%	50%	45%	86%	371 [39-1994] n=16 trials	3 [1-6]

[a]Other designs not included in table
[b]Inner-80-percentile range (10^{th}-90^{th} percentiles)
[c]Outcomes for which statistical significance was known

6.3.2.1. Sensitivity analyses

Three sensitivity analyses were conducted to assess the robustness of the odds ratios for outcome reporting bias (Table 6-7). For the PubMed and CIHR cohorts, sensitivity analyses restricted to those trials with Outcomes Survey responses had little impact on the odds ratios. A second sensitivity analysis excluded physiological and pharmacokinetic trials. The odds ratios were not greatly affected in the PubMed and Ethics cohorts. The CIHR cohort did not contain any physiological or pharmacokinetic trials.

A final sensitivity analysis was conducted to assess the impact of dichotomising the level of reporting differently (Table 6-7). The statistical significance of fully and partially reported outcomes was compared to that of qualitatively reported and unreported outcomes. Dichotomising the level of reporting in this manner produced higher pooled odds ratios in each trial cohort for both efficacy and safety outcomes.

Table 6-7. Sensitivity analyses to assess the robustness of the odds ratio for outcome reporting bias

a) Efficacy outcomes

Trial design	All trials		Excluding Outcomes Survey non-responders		Excluding physiological and pharmacokinetic trials		Fully/partially reported versus qualitatively/unreported	
	Number of trials[a]	OR[b] (95% CI)	Number of trials[a]	OR[b] (95% CI)	Number of trials[a]	OR[b] (95% CI)	Number of trials[a]	OR[c] (95% CI)
ALL								
PubMed cohort	161	2.0 (1.6 - 2.7)	114	2.1 (1.6-2.9)	148	1.9 (1.4-2.5)	134	3.2 (2.4-4.2)
Ethics cohort	45	2.4 (1.4-4.1)	N/A	N/A	40	2.8 (1.6-4.9)	30	3.3 (1.8-6.1)
CIHR cohort	30	2.7 (1.5-5.0)	25	2.9 (1.5-5.8)	N/A	N/A	20	5.1 (2.5-10)
PARALLEL GROUP								
PubMed cohort	135	1.9 (1.4 - 2.6)	92	1.8 (1.3-2.6)	130	1.9 (1.4-2.5)	98	2.9 (2.0-4.0)
Ethics cohort	33	3.0 (1.6-5.8)	N/A	N/A	32	3.3 (1.7-6.5)	22	3.7 (1.7-8.1)
CIHR cohort	24	2.4 (1.1-5.1)	19	2.6 (1.1-6.2)	N/A	N/A	8	8.6 (2.9-25)

[a]Trials were excluded if odds ratio could not be calculated due to adjacent 'zero' cells in 2x2 table
[b]Odds ratio > 1 signifies that fully reported outcomes have a higher odds of being statistically significant ($p<0.05$) compared to incompletely reported outcomes
[c]Odds ratio > 1 signifies that fully/partially reported outcomes have a higher odds of being statistically significant ($p<0.05$) compared to unreported/qualitatively reported outcomes

a) Safety outcomes

Trial design	All trials		Excluding Outcomes Survey non-responders		Excluding physiological and pharmacokinetic trials		Fully/partially reported versus qualitatively/unreported	
	Number of trials[a]	OR[b] (95% CI)	Number of trials[a]	OR[b] (95% CI)	Number of trials[a]	OR[b] (95% CI)	Number of trials[a]	OR[c] (95% CI)
ALL								
PubMed cohort	43	1.9 (1.1 - 3.5)	31	1.4 (0.66-3.0)	38	2.4 (1.3-4.4)	38	4.2 (2.5-7.0)
Ethics cohort	17	4.4 (1.7-11.7)	N/A	N/A	17	4.4 (1.7-11.7)	14	7.4 (2.2-25.4)
CIHR cohort	4	7.7 (0.53-111)	4	7.7 (0.53-111)	N/A	N/A	4	12.3 (1.5-99)
PARALLEL GROUP								
PubMed cohort	33	3.1 (1.6 - 5.8)	21	2.6 (1.1-6.3)	33	3.1 (1.6-5.8)	27	5.3 (2.7-10.3)
Ethics cohort	16	5.0 (1.8-13.9)	N/A	N/A	16	5.0 (1.8-13.9)	13	9.7 (3.0-31.7)
CIHR cohort	2	43.4 (2.9-662)	2	43.4 (2.9-662)	N/A	N/A	2	25.3 (2.1-308)

[a]Trials were excluded if odds ratio could not be calculated due to adjacent 'zero' cells in 2x2 table
[b]Odds ratio > 1 signifies that fully reported outcomes have a higher odds of being statistically significant ($p<0.05$) compared to incompletely reported outcomes
[c]Odds ratio > 1 signifies that fully/partially reported outcomes have a higher odds of being statistically significant ($p<0.05$) compared to unreported/qualitatively reported outcomes

6.3.2.2. Trial characteristics and outcome reporting bias

Exploratory meta-regression was limited to efficacy outcomes in the PubMed cohort, as there were too few trials in the other data sets to permit reliable assessments of covariates. Thirteen pre-specified trial characteristics were included in univariate meta-regression analyses, followed by exploratory backward stepwise meta-regression.

The coefficients from exploratory univariate regression analyses are shown in Table 6-8. Multicentre trials were associated with a significantly reduced degree of outcome reporting bias compared to single centre trials (p=0.03). None of the other covariates was statistically significant.

Correlations were examined between explanatory variables, as the regression analysis might have been misleading if high collinearity was present. Three pairs of variables demonstrated correlation coefficients greater than 0.5. Sample size was correlated with the multicentre status (r=0.52) and trial design (r=0.54). The reporting of a power calculation was correlated with the specification of primary outcomes (r=0.57). Adequate blinding was also correlated with drug interventions (r=0.43).

Thirteen trial characteristics were initially examined in an exploratory analysis using backward stepwise meta-regression. No significant covariates were identified. Two variables from the univariate analyses, sample size and reporting of a power calculation, were then excluded based on collinearity with other variables. Backward stepwise meta-regression was repeated starting with the remaining 11 covariates. The final exploratory model included the

specification of primary outcomes, which was associated with increased bias (p=0.035), and multicentre status, which was associated with less bias (p=0.007) (Table 6-9).

Table 6-8. Univariate meta-regression analyses for the association between the odds ratio for efficacy outcome reporting bias and thirteen trial characteristics

Trial characteristic	Coding (number of trials)	Regression coefficient	Standard error	*p*-value
Trial design	0 = parallel (135) 1 = other (26)	0.41	0.39	0.30
Type of intervention	0 = drug (120) 1 = other (41)	0.16	0.32	0.61
Journal type	0 = general (11) 1 = specialty (150)	0.70	0.51	0.17
Type of journal report	0 = short/letter (2) 1 = full length (159)	-1.1	1.4	0.41
Log(sample size)	Continuous (161)	-0.17	0.11	0.11
Type of funding	0 = full industry (60) 1 = other (89)	0.34	0.29	0.23
Number of sites	0 = single centre (108) 1 = multicentre (51)	-0.65	0.30	0.03
Specification of primary outcome(s)	0 = not specified (74) 1 = specified (87)	0.36	0.27	0.19
Power calculation	0 = described (50) 1 = not mentioned (111)	-0.03	0.30	0.91
Reporting of random sequence generation	0 = adequate (40) 1 = inadequate (121)	0.01	0.32	0.97
Reporting of allocation concealment	0 = adequate (44) 1 = inadequate (117)	0.12	0.31	0.69
Use of blinding	0 = adequate (105) 1 = inadequate (56)	-0.17	0.29	0.56
Reporting of attrition	0 = adequate (54) 1 = inadequate (107)	0.33	0.29	0.26

Table 6-9. Final exploratory meta-regression model for the association between trial characteristics and efficacy outcome reporting bias

Trial characteristic	Coding (number of trials)	Regression coefficient	Standard error	*p*-value
Specification of primary outcome(s)	0 = not specified (74) 1 = specified (87)	0.60	0.29	0.035
Number of sites	0 = single centre (108) 1 = multicentre (51)	-0.83	0.31	0.007

The pooled odds ratios for outcome reporting bias, stratified by specification of primary outcomes and multicentre status, are shown in Table 6-10. The largest degree of bias was found in 49 single centre trials that specified their primary outcome(s) in the publication (OR 3.4, 95% CI 2.0-5.7), while the smallest was found in 13 multicentre trials that did not specify such outcomes (OR 0.74, 95% CI 0.24-2.3).

Table 6-10. Pooled odds ratios for outcome reporting bias stratified by specification of primary outcomes and multicentre status

Specification of primary outcomes	Single centre trial		Multicentre trial	
	n trials	OR[a] (95% CI)	n trials	OR[a] (95% CI)
Specified	49	3.4 (2.0-5.7)	38	1.6 (0.93-2.7)
Not specified	59	1.9 (1.4-2.8)	13	0.74 (0.24-2.3)

[a]Odds ratio > 1 signifies that fully reported outcomes have a higher odds of being statistically significant ($p<0.05$) compared to incompletely reported outcomes

6.4. Discussion

Systematic reviewers must often rely on published data to derive overall estimates of treatment effect. An important question, therefore, is whether the completeness of outcome reporting is associated with specific outcome characteristics, particularly the statistical significance of inter-group comparisons.

6.4.1. Reasons for omitting outcomes

Based on PubMed and CIHR Outcomes Survey responses, the most common reasons given by trialists for omitting particular trial outcomes included clinical

irrelevance and a lack of statistical significance. Space constraints and planned submission of results at a later date were also commonly cited.

The failure to report all measured outcomes within a trial publication appears to arise from investigators' decisions based on a combination of the results, the importance of the outcome, and the external factor of journal space restrictions. This is consistent with previous observations that the failure to publish entire studies stems primarily from investigators' rather than editors' actions, although the authors' failure to submit a manuscript may occur in anticipation of journal rejection.[14,42,43,47,49,55]

6.4.1.1. Clinical importance

As part of the Outcomes Survey, trialists in the PubMed cohort were asked to indicate the relevance of unreported outcomes by rating their clinical importance and providing their pre-specification. A quarter of studies with unreported efficacy outcomes ranked at least one of them as being highly important or primary outcomes. Almost a fifth of trials with unreported safety outcomes ranked at least one of them as a primary outcome, although none were rated as highly important. Finally, 63% of PubMed trials and 38% of CIHR trials omitted efficacy outcomes for reasons other than a lack of clinical importance. It is therefore evident that, in many cases, important outcomes are omitted from trial publications.

Although many PubMed and CIHR trialists explained that they omitted outcomes due to a lack of clinical importance, it is unclear whether this determination of importance was made *a priori* or whether it was data-driven.

If an outcome was deemed to be unimportant in the planning stages of a trial, then it does not seem logical that resources and participants' time should have been wasted on recording irrelevant measures. An alternative explanation is that the importance of outcomes was defined *post hoc* based on results. At its worst, a trial may begin with an initial 'shotgun' approach of collecting as many outcomes as possible in the hopes of yielding a significant result. This approach to assigning clinical importance has the potential to introduce significant bias if it results in particular types of outcomes being systematically classified *post hoc* as 'unimportant' based on data analysis, with the consequence of being incompletely reported or omitted entirely from publications. Even non-significant exploratory outcomes, if deemed to have potential importance *a priori*, should be reported to enable confirmation in future studies and incorporation in future meta-analyses.

As an alternative explanation for the failure to report secondary outcomes, some may argue that in the context of a trial with non-significant primary outcomes, the non-primary outcomes become irrelevant.[167] One trialist in the CIHR cohort explained that certain secondary outcomes were unreported because "reporting on these 'negative' secondary outcomes in the context of a 'negative' trial would be of secondary importance." This might be the case if the secondary outcomes were intended as confirmatory or supportive evidence for primary outcomes, and if the trial was adequately powered for the primary outcome.

While this view may sometimes be reasonable in the interpretation of a single trial, it remains controversial.[168,169] Such omissions have the potential to bias

the results of subsequent systematic reviews by preventing the inclusion of particular outcomes. Trials on a specific intervention may not all agree on their choice of primary and secondary outcomes, and the choice of primary outcomes in a meta-analysis may not match the primary outcomes defined in each trial. It is thus important that all primary and secondary outcomes be reported for potential inclusion as main outcomes in meta-analyses.

Furthermore, it is problematic to base the decision to report particular outcomes on data analyses. The dependence of secondary outcome reporting on the statistical significance of primary outcomes has the potential to introduce bias. Particularly for trials that are under-powered for the primary outcome, this may mean that secondary outcomes would rarely be reported because the primary outcomes are unlikely to be significant. Even though trials may often be under-powered with respect to secondary outcomes, non-significant secondary outcomes have the potential to become statistically significant when pooled in a meta-analysis. One can also doubt whether statistically significant secondary outcomes in a trial would be omitted just as readily as non-significant ones when the primary outcome is not significant.

6.4.1.2. Journal space restrictions

A lack of journal space was the most frequently cited reason for omitting efficacy outcomes among PubMed trials, most of which were published in specialty journals. Space restrictions were cited less often for CIHR trials, many of which were published in general medical journals. A possible explanation for the difference between the two cohorts is that greater space restrictions were imposed by specialty journals. This is consistent with the

finding that incompletely reported outcomes were more prevalent among CIHR trials published in specialty compared to general medical journals.

For journal space limitations to be an unbiased justification for omitting outcomes from trial reports, the selection of outcomes to report should be independent of data analysis. However, our analysis of odds ratios for outcome reporting bias indicates that fully reported outcomes are more likely to be statistically significant compared to incompletely reported outcomes.

According to survey responses, the omission of outcomes due to space limitations was imposed by both journals and trialists. However, the responsibility must ultimately lie with the journals, as author-imposed omissions may occur in anticipation of journal editing. With the advent of electronic journals and the opportunity to publish additional data online, space limitations may become less of an issue for trial reporting.[170] However, this privilege is certainly not the norm at present among the 271 different print journals that published trial reports in the PubMed cohort.

6.4.1.3. Statistical significance

A preference for statistically significant outcomes was evident from the survey responses. It is remarkable that even among high quality CIHR trials, almost half of the trialists who omitted at least one efficacy or safety outcome stated that they did so based on a lack of statistical significance. The same reason was provided by a quarter of PubMed trialists for unreported efficacy outcomes and one half of trialists for safety outcomes. This constitutes direct evidence that many outcomes remain unreported based on statistical significance.

Similar findings have been observed for non-publication of entire trials, where surveys of trialists have shown that 15-28% of trials remain unpublished due to null or 'uninteresting' results.[14,47]

Additional supporting evidence for bias against non-significant outcomes comes from our observation that the vast majority of unreported outcomes in PubMed (71%) and CIHR (99%) trials were non-significant. These proportions are much higher than among reported outcomes in PubMed (30%) and CIHR (31%) cohorts. Furthermore, reported primary outcomes from PubMed trials were almost twice as likely to be significant compared to unreported primary outcomes. Finally, statements in trial publications also supported the existence of biased reporting practices. One report explained that "since results ... did not reveal significant changes, data are not given in the tables."

The true prevalence of outcomes that were omitted for reasons of statistical non-significance is likely to be higher than what we found in our survey of trialists. Firstly, responders might have been less inclined to admit that a lack of statistical significance was the primary reason for omitting an outcome, particularly if they were aware that this practice is biased. Secondly, the omission of outcomes due to other reasons given by trialists, particularly space restrictions and a lack of clinical importance, may well be directly associated with a lack of statistical significance. Outcomes could be deemed *post hoc* to have little clinical relevance if they fail to show significant findings, while the omission of 'uninteresting' non-significant outcomes may be more common when accommodating space limitations. This view is supported by the observation that non-significance and a lack of clinical importance were

frequently cited together as reasons for omitting an outcome. In addition, unreported outcomes in the PubMed cohort that were rated as having low clinical importance were more likely to be non-significant compared to those with high importance (18% versus 43%).

In addition to evidence from our trialist surveys, the association between statistical significance and both clinical importance and space restrictions have been supported by certain journal instructions to authors. Journal reviewers and editors have previously encouraged the selective reporting of the most 'interesting' results, such as those demonstrating statistically significant differences.[94] For example, the *American Journal of Psychiatry* asks for "all significant and *important* nonsignificant results" [my emphasis] to be reported,[70] which implies different reporting standards based on statistical significance. Another unidentified journal stated that it was "impossible for [them] to use space to publish negative trials."[171] The editor of a major environmental/toxicological journal issued a rejection letter to a researcher, explaining that although "the manuscript [was] very well written and the study was well documented..., the negative results translate into a minimal contribution to the field."[28] In yet another example, the journal *Diabetologia* stated in its instructions to authors in 1984 that "mere confirmation of known facts will be accepted only in exceptional cases; the same applies to reports of experiments and observations having no positive outcome."[67] A lack of journal space and clinical importance can thus be seen as being intrinsically linked to a lack of statistical significance.

It is therefore evident that the justifications provided by PubMed and CIHR trialists for omitting outcomes are largely inadequate. The reasons tend to be associated with a preference for statistically significant outcomes, and thus act either directly or indirectly to perpetuate outcome reporting bias within trial publications. The tendency to exclude statistically non-significant outcomes would create a preponderance of significant outcomes in the medical literature, resulting in potentially biased meta-analyses. This bias would act in addition to, and in the same direction as, the suppression of whole trials from the literature (study publication bias).

6.4.2. Association between level of reporting and statistical significance

This is the first study to assess the magnitude and direction of outcome reporting bias across a representative population of studies. A direct measure of bias is the association between the level of reporting and statistical significance. The pooled odds ratio for outcome reporting bias was 2.0-2.7 for efficacy outcomes across the three trial cohorts, and 1.9-7.7 for safety outcomes. Parallel and cross-over trials had similar magnitudes of bias. The magnitude of outcome reporting bias is similar to that of study publication bias, which was found to be an odds ratio of 2.54 in a meta-analysis of five cohort studies.[44]

The pooled odds ratio of outcome reporting bias was robust in a variety of sensitivity analyses, which supports the reliability of the findings. Survey non-responders were included in the main calculation of odds ratios with the conservative assumption that they had no unreported outcomes. Physiological and pharmacokinetic trials were also included in the main analysis, although

they may have had different objectives and outcomes compared to other therapeutic or preventative trials. However, excluding either the trials of non-responders or physiological and pharmacokinetic trials made little difference to the magnitude of bias.

Finally, the manner in which the levels of reporting were dichotomised may have affected the measure of outcome reporting bias. A sensitivity analysis comparing fully and partially reported outcomes to qualitatively and unreported outcomes resulted in higher degrees of bias across all cohorts and study designs. This effect was expected, because partially reported outcomes were the most heterogeneous group in terms of the potential for bias. Some outcomes may have been partially reported for reasons unrelated to statistical significance, such as a lack of awareness about how much detail should be included in a publication. This would be particularly relevant for the reporting of paired data, where the standard error of the mean difference may not be provided, or with survival data, where a Kaplan-Meier curve may be provided without the numbers at risk in the various time intervals.

It should be noted that the pooled odds ratio is an average value that cannot be universally applied to individual studies. The estimated magnitude of outcome reporting bias in each trial varied widely and cannot be reliably predicted for a given study.

Our findings of bias are consistent with the limited literature available on this topic. Previous publications expressing concern about selective outcome reporting within published trials have consisted of case studies showing that

the inclusion of unreported outcome data from published trials reduced the magnitude of the pooled treatment effect across trials.[111,112] Statistically significant findings have become non-significant with the inclusion of unreported data from published trials,[50] and have even reversed to become significant in the opposite direction.[112] Although these individual cases have limited external validity and may be subject to publication bias themselves, they illustrate the potential impact of unpublished outcomes on the conclusions drawn from meta-analyses.

The overall focus on significant results and *p*-values has been commonly observed and widely criticised by the research community.[6,28,92,93,95,118,172,173] A review of the scientific literature prompted researchers to comment that they "were depressed by the frequency of use of statistical significance as a measure of relative [scientific] importance."[174] Researchers often feel that studies without statistically significant results have little chance of being published in the face of a "prejudice against the null hypothesis."[25,54] Interestingly, the emphasis on *p*-values has not always been commonplace. In 1965, the editor of the *Journal of Experimental Psychology* commented that "the *p* level...is only one element in the persuasion."[69]

It is important to recognise that although statistical significance has received the most attention as a determining factor in the decision to report study outcomes, there are likely to be other factors that play a role in the selection process. These may include the novelty of results, whether the findings contradict or confirm prior beliefs, the commercial interests of sponsors, and the academic interests of investigators. Regardless of the multi-factorial nature

of selective reporting, statistically significant outcomes remain widely coveted by trialists as the "passport to publication."[175]

6.4.3. Risk factors for outcome reporting bias

The only significant factor in the univariate meta-regression analyses of 13 trial characteristics was the multicentre status of a study. In the exploratory multivariate analysis, the presence of multiple study centres and the failure to specify primary outcomes were both significantly associated with a lower magnitude of outcome reporting bias. The verification of outcome reporting by multiple study centres may reduce the chance of selective reporting in published trials. A previous review found a significant interaction between study publication bias and multicentre status, with higher degrees of bias among single centre studies,[43] while another review did not observe a significant association.[42] It is unclear as to why trials that define primary outcomes would be associated with increased bias, as this usually constitutes good trial reporting. The findings are exploratory and require confirmation.

No significant association was observed between outcome reporting bias and funding sources or other markers of reporting quality. However, it may be that unfavourable, industry-funded studies are simply unpublished as opposed to being published with selectively reported outcomes. Previous studies have not observed an interaction between study publication bias and commercial funding.[42] With regards to study quality, the reporting of random sequence generation, allocation concealment, and blinding have been shown to be unreliable markers of actual trial conduct, which precludes reliable conclusions from our analysis of these factors.[117]

6.4.4. Study limitations

The advantages of the three study cohorts were discussed in Chapter 4 along with potential sampling biases. Unrestricted by design, topic, or journal type, the PubMed and Ethics samples are representative of a relatively unselected and broad population of randomised trials funded by a variety of sources. The CIHR cohort, on the other hand, represents a population of large, government-funded, high quality trials whose methodology was subjected to rigorous peer review. The remainder of this chapter will discuss the potential biases arising in the process of data collection and analysis for trial outcomes, as well as their impact on the findings presented in the previous two chapters.

6.4.4.1. Limitations of data collection

The retrospective approaches applied to the trial cohorts may have introduced recall bias for Outcomes and Publications Survey questionnaires that solicited information about unreported outcomes in the three cohorts. For example, one trial protocol specified an outcome measurement up until two years of follow-up, but the publication presented significant results for this outcome at one month only despite a reported mean follow-up of over two years. The trialist explained that he "honestly can't remember whether [they] looked at two years or not." This problem of recall would have likely been more common for older trials in the Ethics and CIHR cohorts. Retrospective rather than prospective approaches were thought to be more valid overall for the thesis methodology because trialists may have been influenced by enrollment in a prospective study assessing their reporting practices. Also, from a feasibility standpoint, one would have had to wait many years for trial publication to occur. Median

times to publication of up to eight years have been observed in previous reviews of clinical research.[13]

Response bias may also have reduced the reliability of our findings, and a discussion of this issue was provided in Chapter 5 (Section 5.3.2). Response rates of 69%-74% were achieved in the Outcomes Surveys. Non-responders may have differed systematically from responders, and 25%-85% of survey responses were found to be unreliable in terms of identifying unreported outcomes. Since it is unlikely that trialists would over-report the number of omitted outcomes or exaggerate any biased selection of outcomes to report, the effect of response bias is likely to be to underestimate the true prevalence of incomplete outcome reporting, as well as the true magnitude of outcome reporting bias derived from our samples. Our findings can thus be viewed as a conservative estimate of selective outcome reporting.

A limitation of the PubMed cohort was that protocols were not available as an objective source of outcomes data. The heavy reliance on survey responses to identify unreported outcomes was found to be inadequate, resulting in a much lower and likely under-estimated prevalence of unreported outcomes compared to the Ethics and CIHR cohorts.

A limitation for Ethics trials was that Outcomes Survey data were not available, and the statistical significance of unreported outcomes was thus unknown. Similarly, the significance of many unreported outcomes was unknown for survey non-responders in the PubMed and CIHR cohorts. However, based on observations from the PubMed and CIHR cohorts, it can be assumed that

unreported outcomes had a high chance of being non-significant. Excluding these outcomes from the calculation of the odds ratio for outcome reporting bias would therefore act in a conservative direction to under-estimate the true extent of bias. This assumption is consistent with the finding that the odds ratio for bias was unchanged in the PubMed cohort and decreased in the CIHR cohort when unreported outcomes were excluded from the analysis.

Another potential limitation is that data collection in the Ethics and CIHR cohorts relied on the original trial protocols, but formal amendments may have been made at a later date. Formal amendments that were submitted to the Copenhagen and Frederiksberg Research Ethics Committee could not be comprehensively identified because they were filed separately from the original protocol and were not identifiable on the database. As a result, amendments were not incorporated into data collection for Ethics trials. Amendments that were submitted to the CIHR were reviewed, although it is unclear how many trialists would have felt it was necessary to submit formal amendments to the funding agency. Regardless, it would presumably be uncommon for amendments to have included changes to the specification of outcomes.

It is unclear as to why major changes would be made to trial outcomes, aside from practical measurement issues or new evidence invalidating the outcome. Modifications to protocols prior to trial commencement are usually unbiased, and should ideally be incorporated in the thesis data collection. On the other hand, amendments made after trial commencement have the potential for bias if they are data-driven, and therefore should not be incorporated in the data collection. Depending on the prevalence and timing of amendments to

outcome measures, the potential impact on the degree of outcome reporting bias is unclear.

Finally, data collection in the PubMed and CIHR cohorts was conducted by just one individual, which may have increased the potential for errors. However, inter-observer variability was minimised as a result, and trial reports in the PubMed and CIHR cohorts were reviewed twice by the same individual to double-check the accuracy of data extraction. Few errors (3% of outcomes) were identified when outcomes were reviewed by a second researcher for a random sample of 19 Ethics trials, which indicates that the error rate was low and that the data collection method was reliable.

6.4.4.2. Limitations of data analysis

Analyses were conducted at the trial level to account for the clustering of outcomes within each trial. It was assumed in the analyses that trials were independent from each other. However, as outlined in Chapter 4, several individual researchers published multiple trials, and reporting practices by the same author may have tended to be similar across different studies. 5/514 (1%) investigators appeared as contact authors for two trials each in the PubMed cohort, while 8/91 (9%) trialists published more than one trial each in the Ethics cohort. For the CIHR cohort, 7/41 (17%) trialists were listed as contact authors for 2 trials each. Our analyses did not account for any clustering effect between trials published by the same individual; the impact, if any, could be in either direction and would be specific to the individual case. The effect would likely be minimal, as the trials would have had different co-

investigators and were published in different journals with varying requirements for reporting.

A large proportion of trials in the three cohorts was excluded from the analysis of outcome reporting bias because odds ratios could not be calculated. This would be analogous to excluding a trial with no events in any study group from a meta-analysis, as useful information could not be derived for between-group comparisons. In all three cohorts, trials that were included in the calculation of odds ratios measured more outcomes with known statistical significance than excluded trials. This would be expected because trials with fewer outcomes would likely have less variability in statistical significance or levels of reporting across outcomes. The pooled odds ratio for outcome reporting bias is thus more representative of trials that measure larger numbers of outcomes. A higher proportion of cross-over trials was also observed among excluded trials in the PubMed and Ethics cohorts. Because of the type of outcome data required for meta-analysis of paired data, cross-over trials would have a higher chance of being excluded for not having any fully reported outcomes.

6.5. Chapter summary

Our results provide the first direct evidence of outcome reporting bias across broad cohorts of randomised trials. Data were collected from protocols, publications, and trialist surveys in order to evaluate the selective reporting of outcomes. Adequate reporting of trial outcomes was associated with statistical significance, such that fully reported outcomes had greater than two-fold higher odds of being significant compared to incompletely reported outcomes. The reasons commonly provided by trialists for omitting outcomes included a lack

of clinical importance, space constraints, and statistical non-significance. All of these justifications are likely biased, and lend additional supporting evidence for the tendency of trialists to selectively report statistically significant outcomes in their publications.

Chapter 7 - Consistency of outcomes between trial protocols and publications

7.1. Introduction

The planning and analysis of a methodologically sound randomised trial requires that clinically relevant outcomes be specified in advance of data collection and analysis. This practice helps in reducing the potential for spurious findings and distinguishes confirmatory from exploratory analyses.[94,129] Pre-specification is particularly important for primary outcomes because their designation implies that they should be given the most weight when interpreting trial results. Bias may arise if the specification, analysis, and reporting of primary outcomes are influenced by the results.

The previous two chapters presented evidence of widespread incomplete outcome reporting and its association with statistical significance. The present chapter will examine whether modifications are made to trial protocols in an attempt to achieve statistically significant primary outcomes. Using the Ethics and CIHR cohorts, outcomes specified *a priori* in trial protocols were compared to those reported in the final publications. The existence of discrepancies provided direct evidence of biased *post hoc* selection of primary outcomes to report in randomised trials.

7.2. Review of data collection

Outcomes were recorded from protocols and publications of trials in the Ethics and CIHR cohorts. Protocols and publications were compared with respect to the number of outcomes, the specification of each outcome as primary or non-

primary, and the analysis plans. Major inconsistencies were defined in advance to include the introduction of new primary outcomes, the failure to report the pre-specified primary outcome, changing the specification of pre-defined primary outcomes, changing the outcome used in the power calculation, and modifying the data type used in hypothesis testing. Publications were also reviewed for any mention of amendments to the original protocol.

7.3. Discrepancies in outcomes between protocols and publications

7.3.1. Prevalence of vague outcomes in trial protocols

A vague outcome was defined as one whose protocol description was not sufficiently detailed to describe a specific outcome measure. A discussion of the issues surrounding these outcomes was provided in Chapter 3 (Section 3.3). 29% (771/2694) of efficacy and 19% (173/919) of safety outcomes were vaguely described in protocols from the Ethics cohort. Corresponding percentages were lower in the CIHR cohort, with vague outcomes accounting for 12% (154/1233) and 8% (13/157) of efficacy and safety outcomes respectively. 64% (63/99) of Ethics and 44% (21/48) of CIHR trial protocols contained at least one vague efficacy outcome. 39% (27/69) and 12% (3/26) of protocols for the two cohorts respectively contained vague safety outcomes. When a specific outcome corresponding to a vague description was reported in the publication, it was not recorded as inconsistent with the protocol.

Of note, 6/102 (6%) Ethics protocols did not describe any specific outcomes. All were cross-over trials in the fields of physiology (n=4), gastroenterology (n=1), and psychiatry (n=1). At least one outcome was specified in every CIHR protocol.

7.3.1.1. Example

The following example from our study sample will illustrate a few of the problems presented by vague outcomes. The protocol for an industry-funded, multicentre trial comparing interventions for a prostate condition defined a specific score as one of its "main evaluation criteria". The analysis section of the protocol explained that the score would be analysed both as a "quantitative parameter" using analysis of variance, and as a "qualitative parameter' using a chi-square or Fisher exact test. In the final report, three outcomes were presented based on this symptom score. Two continuous primary outcomes were defined as the mean score at endpoint as well as the mean change from baseline. A third binary outcome was defined as the proportion of participants with 50% or greater improvement in their score from baseline to endpoint. All three outcomes were statistically significant. To be conservative in our data collection, these outcomes were not recorded as inconsistent with the protocol, but were listed as vaguely-specified because they were not adequately detailed in the protocol. However, the lack of protocol detail introduces the possibility that the cut-off for dichotomising the binary outcome was determined after examining the data. It is also possible that the choice to report both continuous outcomes was data-driven and would not have occurred if one or both were non-significant.

Table 7-1. Proportion of CIHR and Ethics trials with major discrepancies in the specification of primary outcomes and power calculation outcomes when comparing protocols and publications

Discrepancy in trial publications relative to protocols	Proportion (%) of trials with inconsistencies for ≥1 primary outcome	
	Ethics cohort	CIHR cohort
Changes to protocol-defined primary outcome[a]	**40/76 (53%)**	**16/48 (33%)**
Reported as non-primary	26/76 (34%)	11/48 (23%)
Omitted	20/76 (26%)	6/48 (13%)
New publication-defined primary outcome[b]	**21/63 (33%)**	**11/45 (24%)**
Changed from non-primary to primary	12/63 (19%)	4/45 (9%)
Not specified in protocol	11/63 (17%)	8/45 (18%)
ANY CHANGE TO PRIMARY OUTCOMES[c]	**51/82 (62%)**	**19/48 (40%)**
Change in power calculation outcome[d]	**10/38 (29%)**	3/36 (8%)
Change from outcome used in protocol	4/38 (11%)	2/36 (6%)
New power calculation (not reported in protocol)	6/38 (17%)	1/36 (3%)
ANY CHANGE TO PRIMARY OR POWER CALCULATION OUTCOMES[e]	**52/82 (63%)**	**19/48 (40%)**

[a,b,c]Among trials that defined primary outcomes in protocols,[a] publications,[b] or either[c]
[d]Among trials that reported a power calculation in the publication
[e]Among all above trials (a-d)

7.3.2. Prevalence of major inconsistencies in primary outcome specification

A large proportion of trials contained major discrepancies between protocols and publications in the specification of primary outcomes (Table 7-1). 62% (51/82) of Ethics trials and 40% (19/48) of CIHR trials contained major discrepancies in specification that were classified as changes to protocol-defined primary outcomes, or the appearance of new publication-defined primary outcomes. A total of 213 protocol- or publication-defined primary outcomes were identified as inconsistent in the Ethics cohort, 202 (95%) of

which were efficacy outcomes. 48 of 51 (94%) inconsistent primary outcomes measured efficacy data in the CIHR cohort. None of the trial reports in the Ethics and CIHR cohorts mentioned that an amendment had been made to primary outcomes.

All trials with major discrepancies in the Ethics cohort were verified by two individuals, and disagreements were resolved by consensus. Major corrections to the original data set were made for 3/259 (1%) primary outcomes in 3/52 (6%) trials, changing the overall prevalence of major discrepancies from 66% (54/82) to 62% (51/82). Minor corrections were made to 24/259 (9%) primary outcomes in 13/52 (25%) trials, but none of these affected the overall prevalence of major discrepancies in the cohort.

7.3.2.1. Changes to protocol-defined primary outcomes

One half of Ethics trials and one third of CIHR trials made major changes to protocol-defined primary outcomes (Table 7-1). These changes included trials that reported pre-specified primary outcomes as either secondary or unspecified outcomes in publications, which occurred in 34% (26/76) of Ethics and 23% (11/48) of CIHR trials. The major changes also included 26% (20/76) of Ethics trials and 13% (6/48) of CIHR trials that omitted pre-specified primary outcomes entirely from their publications. The denominators of these proportions correspond to trials that were eligible for the particular discrepancy, which in this case is the number of trials with protocol-defined primary outcomes.

Example

A single centre, government funded trial evaluated the efficacy of an anti-diabetic drug. A single primary outcome was clearly defined in the protocol as the response rate based on a specific, sustained percentage decrease in mean glycated hemoglobin. This outcome was also used in the power calculation. The trialist's survey response stated that the outcome was non-significant in a chi-square analysis and was thus omitted from the publication for this reason. In its place, a new outcome, the withdrawal rate from study treatment, was reported as a statistically significant primary outcome. No reference to the change in outcomes was made in the trial report. The power calculation outcome in the publication was described vaguely as "response rate," without further elaboration. This example illustrates an unambiguous case of biased manipulation in the selection and omission of primary outcomes based on statistical significance.

7.3.2.2. New publication-defined primary outcomes

Another common source of major inconsistencies was the introduction of new publication-defined primary outcomes (Table 7-1). Publications for one third of Ethics trials and one quarter of CIHR trials introduced at least one primary outcome that was not pre-specified as primary in the protocol. These discrepancies consisted of outcomes whose specification was changed from non-primary in protocols to primary in publications (19% of Ethics and 9% of CIHR trials), as well as the introduction of new primary outcomes that were not mentioned at all as outcomes in the protocols (17-18% of Ethics and CIHR trials). Among 10 Ethics trials that changed the specification of at least one outcome from unspecified to primary, 4 did not have any primary outcomes

defined in the protocols. All CIHR protocols defined at least one primary outcome.

Example

A multicentre, industry-funded trial of two antibiotics for the treatment of a systemic infection specified a single primary outcome in its protocol - the cure rate at 2 days. The trial publication did not report this outcome, but instead reported a new statistically significant primary outcome defined as the cure rate at 5 days. The statistical significance of the original omitted outcome is not known. The change in the defined time period under consideration for the primary outcome raises questions as to whether several time points were analysed to select the one giving the desired result.

7.3.2.3. Characteristics of trials with and without discrepancies

Trials with discrepancies in primary outcomes differed from those without discrepancies in several respects (Table 7-2). In the Ethics cohort, trials with discrepancies were more likely to be larger, multicentre, industry-funded, parallel group trials. A contrasting trend was observed in the CIHR cohort, with discrepancies being noted more frequently in smaller, single centre studies. By definition, no CIHR trial was fully funded by industry sources.

7.3.2.4. Further examination of discrepancies: Is there evidence for bias?

Trials with discrepancies in the specification of primary outcomes were examined in more detail to explore potential reasons behind the changes. The statistical significance of discrepant outcomes was assessed to determine

whether the inconsistencies tended to favour significant outcomes. Inconsistencies in each trial were classified as unanimously favouring either significant or non-significant primary outcomes, or both. In order to evaluate the impact of the changes on the overall trial interpretation, a second analysis was conducted by considering all primary outcomes defined in the publications, whether they were consistent or inconsistent with the protocol. The inconsistencies were classified as having changed the overall interpretation towards a statistically significant (positive) or non-significant (negative) direction, or as having had no effect on the overall direction.

Table 7-2. Characteristics of trials with and without major discrepancies in primary outcome specification between protocols and publications

	Design[a]		General Journal (%)	Full industry funding (%)	Multicentre (%)	Median sample size [IPR_{80}][b]	Median number of reported outcomes [IPR_{80}][b]
	Parallel (%)	Cross-over (%)					
ETHICS COHORT							
Discrepancies (n=51)	40 (78%)	11 (22%)	2 (4%)	31/50[c] (62%)	30 (59%)	80 [13-1000]	33 [14-78]
No discrepancies (n=51)	30 (59%)	19 (37%)	1 (2%)	25 (49%)	19 (37%)	41 [10-395]	27 [7-81]
CIHR COHORT							
Discrepancies (n=19)	15 (79%)	2 (11%)	7 (37%)	0	11 (58%)	141 [30-2143]	26 [6-63]
No discrepancies (n=29)	24 (83%)	1 (3%)	14 (48%)	0	21 (72%)	403 [64-4374]	25 [10-57]

[a] Other designs not listed
[b] Inner-80-percentile range
[c] Denominator corresponds to number of trials with known sources of funding

Direction of bias in discrepant trial outcomes

Ethics cohort

Among 51 Ethics trials with inconsistencies in primary outcomes, similar proportions favoured either significant (29%, 15/51) or non-significant primary outcomes only (27%, 14/51), while the remaining 43% (22/51) contained discrepancies favouring both or unclear directions (Table 7-3). The statistical significance of unreported outcomes was not known, as the Outcomes Survey was not conducted in this cohort.

Fifteen Ethics trials contained evidence of bias favouring statistically significant primary outcomes (Table 7-3). Nine of these trials reported non-significant, protocol-defined primary outcomes as unspecified (n=8 trials) or secondary (n=1 trial) outcomes in the reports. Two of the nine trials made these changes when another primary outcome was already statistically significant and reported as primary. Three additional trials did not specify any primary outcomes in their protocols, but changed the specification of a statistically significant outcome from unspecified in the protocol to primary in the publication. Another two trials introduced a significant primary outcome that was either previously unspecified (n=1 trial) or not described in the protocol (n=1 trial), while also omitting the protocol-defined primary outcome. The statistical significance of these omitted outcomes was not known. Finally, the remaining trial changed the specification of a non-significant outcome from primary in the protocol to unspecified in the publication, while introducing a new statistically significant primary outcome that had not been described at all in the protocol.

Table 7-3. Number of Ethics trials with major discrepancies in primary outcome specification stratified by statistical significance

Number of trials	Type of result[a]	Type of major discrepancy stratified by statistical significance											
		Changed from primary to non-primary			Omitted			Changed from non-primary to primary			New primary that was not defined in protocol		
		$p<0.05$	$p \geq 0.05$	NR[b]	$p<0.05$	$p \geq 0.05$	NR[b]	$p<0.05$	$p \geq 0.05$	NR[b]	$p<0.05$	$p \geq 0.05$	NR[b]
9	$p<0.05$		■										
3	$p<0.05$							■					
1	$p<0.05$						■	■					
1	$p<0.05$						■				■		
1	$p<0.05$		■								■		
5	$p \geq 0.05$	■											
3	$p \geq 0.05$	■					■						
2	$p \geq 0.05$											■	
2	$p \geq 0.05$						■						
1	$p \geq 0.05$						■		■				
1	$p \geq 0.05$								■				
2	Both										■	■	
1	Both							■	■				
1	Both							■	■			■	
2	Both		■						■				
1	Both	■						■					
1	Both	■									■		
2	Both	■	■				■						
1	Both	■	■										
9	Unknown						■						
1	Unknown			■									
1	Unknown									■			

■ indicates the presence of the particular type of discrepancy in a trial
[a]Favouring either statistically significant or non-significant outcomes, or both, or unknown
[b]NR = statistical significance not reported by publication or trialist

Fourteen Ethics trials contained evidence of bias favouring non-significant primary outcomes (Table 7-3). Eight trials changed the specification of statistically significant outcomes from primary in protocols to unspecified in publications. However, four of these eight trials had other protocol-defined primary outcomes that were already significant and reported as primary. Three of these eight trials also omitted a protocol-defined primary outcome. The remaining six trials introduced new non-significant primary outcomes that were either not defined at all (n=4 trials) or unspecified (n=2 trials) in protocols. Three of these six trials also omitted a protocol-defined primary outcome.

While 29/51 Ethics trials with discrepancies in primary outcome specification favoured either significant or non-significant outcomes, the other 22 trials were not biased in a clear direction in terms of statistical significance (Table 7-3). Eleven of these trials contained a mixture of discrepant outcomes that were significant and non-significant. Four trials introduced a mixture of significant and non-significant primary outcomes that were not defined in the protocol (n=2 trials) or were not specified as primary (n=1 trial), or both (n=1 trial). Another four trials changed an outcome from primary to unspecified while introducing a new primary outcome that was originally unspecified (n=3 trials) or not defined in the protocol (n=1 trial). The changed and introduced outcomes were all statistically significant in two of these trials, and all non-significant in the remaining two trials. Finally, three trials changed a mixture of significant and non-significant outcomes from primary to unspecified outcomes. Two of these three trials also omitted pre-specified primary outcomes.

The remaining eleven Ethics trials contained discrepant outcomes with unknown statistical significance. Nine trials omitted a protocol-defined primary outcome from the trial reports. One trial changed the specification of an outcome from primary to unspecified, while another trial converted an outcome from secondary to primary.

The assessment of Ethics trials was constrained by the absence of information regarding the statistical significance of omitted outcomes. However, if the unreported outcomes were assumed to be non-significant, as was the case with 71% and 99% of unreported outcomes in PubMed and CIHR trials respectively, then discrepancies in 45% (23/51) of Ethics trials would favour statistically significant primary outcomes, compared to only 16% (8/51) favouring non-significant ones. 39% (20/51) would have an unclear direction of bias.

CIHR cohort

For CIHR trials, the statistical significance of omitted outcomes and the reasons for not reporting them were known for Outcomes Survey responders. An evaluation of discrepancies among nineteen CIHR trials revealed evidence of bias favouring significant primary outcomes in 47% (9/19) of trials, compared to only 21% (4/19) favouring non-significant primary outcomes (Table 7-4). If omitted outcomes with unknown statistical significance were assumed to be non-significant, then discrepancies would favour significant outcomes in 53% (10/19) of trials, compared to 16% (3/19) favouring non-significant outcomes.

Table 7-4. Number of CIHR trials with major discrepancies in primary outcome specification stratified by statistical significance

Number of trials	Type of result[a]	Type of major discrepancy stratified by statistical significance											
		Changed from primary to non-primary			Omitted			Changed from non-primary to primary			New primary that was not defined in protocol		
		$p<0.05$	$p \geq 0.05$	NR[b]	$p<0.05$	$p \geq 0.05$	NR[b]	$p<0.05$	$p \geq 0.05$	NR[b]	$p<0.05$	$p \geq 0.05$	NR[b]
2	$p<0.05$		■										
1	$p<0.05$		■				■						
1	$p<0.05$										■		
1	$p<0.05$						■				■		
1	$p<0.05$						■	■					
1	$p<0.05$					■							
1	$p<0.05$		■			■							
1	$p<0.05$					■					■		
1	$p \geq 0.05$											■	
1	$p \geq 0.05$						■					■	
1	$p \geq 0.05$	■											
1	$p \geq 0.05$	■							■				
2	Both		■									■	
1	Both		■						■				
1	Both							■	■				
1	Both	■	■										
1	Unknown						■						

■ indicates the presence of the particular type of discrepancy in a trial

[a]Favouring either statistically significant or non-significant outcomes, or both, or unknown

[b]NR = statistical significance not reported by publication or trialist

Among nineteen CIHR trials with discrepancies in primary outcomes, nine contained evidence of bias to favour statistically significant primary outcomes (Table 7-4). Publications for three of the nine trials changed the specification of non-significant outcomes from primary to non-primary, with one trial also omitting a primary outcome of unknown significance. Three additional trials introduced significant primary outcomes that were either not defined (n=2 trials) or unspecified (n=1 trial) in protocols. Two of the three trials also omitted a pre-specified primary outcome, one of which was omitted due to a "lack of relevance" to the publications. The statistical significance was not provided for either omitted outcome. A further trial omitted a non-significant pre-specified primary outcome, with the explanation that the outcome results would be submitted in a future manuscript. Another trial changed the specification of two non-significant outcomes from primary to unspecified, while also omitting a third non-significant protocol-defined primary outcome. The reason provided for not reporting this primary outcome was a lack of both statistical significance and clinical importance. The final trial omitted a pre-specified primary outcome that was non-significant, while introducing a new primary outcome that was significant and not described in the protocol. The trialist stated that the primary outcome was omitted due to a lack of statistical significance.

Only four of the nineteen CIHR trials with discrepancies demonstrated evidence of bias favouring non-significant primary outcomes (Table 7-4). Two trials introduced new non-significant primary outcomes, with one of the trials omitting a pre-specified primary outcome as well. The statistical significance of the omitted outcome was not provided, but the trialist's intention was to report the result in a future manuscript. One trial changed a significant outcome from

primary to secondary in the report, although there was also another pre-specified primary outcome that was significant and reported as primary. The remaining trial changed a significant outcome from primary to unspecified, while also changing a non-significant outcome from secondary to primary.

Finally, six of the nineteen CIHR trials contained mixed or unclear evidence in terms of the direction of bias (Table 7-4). Three trials converted the specification of non-significant outcomes from primary to unspecified, while also introducing new non-significant primary outcomes that were either undefined (n=2 trials) or unspecified (n=1 trial) in the protocol. Another trial converted fourteen outcomes from secondary to primary, three of which were statistically significant. One additional trial changed the specification of three outcomes from primary in the protocol to unspecified in the report, one of which was significant. The remaining trial omitted a pre-specified primary outcome from the trial report, but the statistical significance was not provided.

Impact of discrepancies on the overall trial interpretation

The discrepancies in primary outcomes were also evaluated in terms of whether the inconsistencies affected the trial results overall. This analysis incorporated all publication-defined primary outcomes to determine the trial's overall interpretation as 'positive' or 'negative', regardless of whether the outcomes were consistent or inconsistent with the protocol. A 'positive' trial was defined as having at least one significant primary outcome, while a 'negative' trial had only non-significant primary outcomes. Trials with no primary outcomes specified in reports were considered to be 'unclear.' A major change was deemed to have affected the overall trial result if a negative trial

became an unclear or positive one (positive change towards statistical significance), or if a positive trial became unclear or negative (negative change towards non-significance).

In the Ethics cohort, discrepancies resulted in positive changes to overall results in 29% (15/51) of trials, compared to 24% (12/51) with changes in a negative direction. 47% (24/51) of trials remained unchanged. Of the twenty-four unchanged trials, twelve were positive, five were negative, and seven were unclear. If omitted outcomes were assumed to be non-significant, then the figures became 43% (22/51) of trials with positive changes and 20% (10/51) with negative changes.

Discrepancies in the CIHR cohort tended to favour statistically significant trials. Among the nineteen CIHR trials with discrepancies, overall results were changed in a positive direction in 32% (6/19) of trials, while 11% (2/19) were shifted in a negative direction. 58% (11/19) remained unchanged. Four of these unchanged trials had at least one significant primary outcome, six had no significant primary outcomes, and one had no primary outcomes defined in the publication. If omitted outcomes with unknown significance were assumed to be non-significant, then 37% (7/19) were shifted in a positive direction, compared to 5% (1/19) in a negative direction.

7.3.2.5. Sensitivity analysis

Primary outcomes in protocols and publications were identified mainly from explicit descriptions in the text. In the absence of explicit definitions, the power calculation outcome was used, followed by specific study objectives if

necessary. A sensitivity analysis was conducted to assess the impact of restricting the definition of primary outcomes to explicit specification only, without using the power calculations or objectives. A similar prevalence of discrepancies was observed among trials with protocol- or publication-defined primary outcomes. 65% (43/66) of Ethics trials overall contained inconsistencies in at least one primary outcome. 56% (36/64) changed or omitted a protocol-defined primary outcome, while 30% (14/47) introduced a new publication-defined primary outcome. CIHR trials were not affected because their primary outcomes were all defined explicitly.

7.3.2.6. Changes to protocol-defined power calculation outcomes

Major discrepancies were also common for outcomes reported in power calculations (Table 7-1). Among trials that described a power calculation in their publications, 29% of Ethics trials and 8% of CIHR trials changed the outcome stated to have been used to estimate the required sample size. 11% of Ethics and 6% of CIHR trials reported a different outcome from the one used in the protocol, while a further 17% of Ethics and 3% of CIHR trials reported a power calculation in their publications despite having described none in their protocol.

All 38 discrepant power calculations outcomes in the Ethics cohort were verified by two individuals. No corrections were necessary as there was 100% agreement between observers.

Example

A single centre, industry-funded trial evaluated the effects of an intervention on post-operative pain and nausea. The trial protocol described a power calculation based on the post-operative nausea grade, which was also the pre-specified primary outcome. However, in the publication, the power calculation was reported as having been based on the visual analogue pain score. The original outcome of nausea grade was omitted from the publication and its statistical result is not known. In its place, a new primary outcome was defined in the trial report based on the pain score over a specified time period, and was statistically significant. This example constitutes a clear case of biased reporting in which the primary outcome was not only altered, but the power calculation was also mis-represented as having been conducted in advance using a different outcome.

7.3.3. Major changes in data analysis

A major change in analysis was defined as alterations to the data type used in the analysis of primary outcomes, as well as changes in the type of analysis from descriptive tabulation to hypothesis testing. The identification of major changes was limited to outcomes with explicit descriptions of their data type or analysis in the protocols and publications.

Changes to the analysis of primary outcomes were uncommon. 5% (4/82) of Ethics trials contained at least one change (Table 7-5). Three of these trials involved alterations in the data type of seven primary outcomes from binary to continuous (n=1 trial), one outcome from continuous to binary (n=1 trial), and one outcome from binary to survival data (n=1 trial). The fourth trial changed

the analysis of a primary outcome from descriptive tabulation to a chi-square analysis. None of these outcomes was statistically significant.

Table 7-5. Proportion of CIHR and Ethics trials with major changes to primary outcome analyses specified in protocols and publications

Major change in data analysis	Number (%) of trials with major change for ≥1 efficacy or safety outcome	
	Ethics cohort (n=82 trials)[a]	CIHR cohort (n=48 trials)[a]
PRIMARY OUTCOMES	**4 (5%)**	**2 (4%)**
Change in data type[b]	3 (4%)	1 (2%)
Change from descriptive to hypothesis testing	1 (1%)	1 (2%)

[a]Number of trials with primary outcomes defined in protocol and/or publications
[b]Change between the following data types: continuous/ordinal versus binary/categorical versus survival data

4% (2/48) of CIHR trials made changes to the analysis of primary outcomes (Table 7-5). One trial changed the data type of two primary outcomes from continuous to ordinal, with corresponding changes in analysis from t-tests to statistically significant chi-square tests. The second trial changed the analysis plan for one primary outcome from a descriptive tabulation to a non-significant t-test.

Example

A multicentre, industry- and government-funded trial evaluated the therapeutic effects of drug interventions on changes in a quantitative cardiovascular variable. The single primary outcome specified in the protocol was statistically significant and reported in the publication. However, two new primary outcomes were also reported, both of which were significant. One of these new outcomes was an ordinal version of an omitted pre-specified continuous outcome. The protocol specified that the original continuous outcome would be analysed using a t-test, but the new primary outcome underwent a chi-

square analysis. It is unclear whether this change from a continuous to an ordinal variable was data driven because the significance of the original outcome was unknown.

A second case example deserves mention, even though it was not one of the trials that was identified under our definition of major discrepancies in analysis. A trial examined the effect of an intervention on patients recovering from a specific acute condition, and published two separate reports of trial results. In one publication, 2-sided significance tests were used, all of which were non-significant. However, in the other publication by the same authors for the same trial, 1-sided significance tests were used, and the lone significant outcome had a *p*-value of 0.03. It is well-known that 1-sided tests are rarely appropriate,[176,177] but even in instances where their use is believed to be warranted, it should be consistent across trial outcomes and publications. The use of both 1- and 2-sided tests for the same trial, coupled with the observation that the lone significant result would have been non-significant based on a 2-sided test, suggests that the choice of using 1- or 2-sided tests was data-driven and biased.

7.3.4. Prevalence of new outcomes reported in publications

A large proportion of trial publications introduced new outcomes of any specification that were not described in protocols (Table 7-6). 77% of Ethics trials and 96% of CIHR trials reported at least one new efficacy outcome, while 46% and 62% respectively reported new safety outcomes. The vast majority of these discrepancies involved a new variable being introduced, while 6%-7% of trials in Ethics and CIHR cohorts reported a new composite efficacy outcome

based on outcomes that had been specified individually in the protocol. 15% of trials in both cohorts introduced a new time point for a specified efficacy variable.

Table 7-6. Proportion of CIHR and Ethics trials with new outcomes based on discrepancies between protocols and publications

	Number (%) of trials with ≥1 new efficacy or safety outcome			
	Ethics cohort		**CIHR cohort**	
	Efficacy (n=99 trials)	**Safety (n=69 trials)**	**Efficacy (n=48 trials)**	**Safety (n=26 trials)**
Any new outcome	**76 (77%)**	**32 (46%)**	**46 (96%)**	**16 (62%)**
Different variable	74 (75%)	32 (46%)	43 (90%)	16 (62%)
Composite outcome	7 (7%)	0	3 (6%)	0
New time point	15 (15%)	2 (3%)	7 (15%)	0

The median number of new efficacy outcomes per trial was 4 [IPR_{80} 0-21] and 6 [1-25] in Ethics (n=99 trials) and CIHR (n=48 trials) cohorts respectively. Corresponding medians for safety outcomes were 0 [0-6] and 1 [0-3] in 69 and 26 trials respectively.

As with the odds ratio analysis for bias in incompletely reported outcomes, we conducted a similar analysis to assess whether the introduction of new outcomes was associated with statistical significance. The odds ratio for new outcomes being significant compared to protocol-defined outcomes was calculated for each trial, and pooled using a random effects model. A significant association was not consistently observed in trials overall or stratified by study design (Table 7-7).

New outcomes in the Ethics cohort were verified by two individuals for a random sample of 13 trials. Disagreements were resolved by consensus. Corrections were made to 13% (12/93) of new outcomes in 23% (3/13) of trials.

Table 7-7. Odds ratio (OR) for the association between new outcomes[a] and statistical significance

	Ethics cohort				CIHR cohort			
	Efficacy outcomes		Safety outcomes		Efficacy outcomes		Safety outcomes	
Trial design	n trials[b]	OR[c] (95% CI)	n trials[b]	OR[c] (95% CI)	n trials[b]	OR[c] (95% CI)	n trials[b]	OR[c] (95% CI)
All trials	56	1.4 (0.96-2.0)	18	1.3 (0.56-3.0)	38	1.0 (0.71-1.5)	7	3.4 (0.67-17)
Parallel group	42	1.3 (0.89-2.0)	17	1.6 (0.77-3.4)	30	1.1 (0.73-1.8)	5	7.8 (1.4-43)
Cross-over	13	1.6 (0.65-3.8)	0	N/A	2	0.32 (0.04-2.8)	1	0.25 (N/A)

[a]Not specified as outcomes in protocol but introduced in publication(s)
[b]Trials were excluded if odds ratio could not be calculated due to adjacent 'zero' cells in 2x2 table
[c]OR > 1 signifies that new outcomes have a higher odds of being statistically significant ($p < 0.05$) compared to outcomes specified in protocols

7.4. Discussion

The availability of trial protocols for the Ethics and CIHR cohorts provided a unique and objective source of information on the nature and number of trial outcomes, as defined *a priori*. Comparisons of protocol outcomes with those reported in publications revealed a high proportion of major discrepancies in primary and power calculation outcomes. Major changes to the analysis of primary outcomes were identified infrequently. A high prevalence of newly-introduced outcomes of any specification was observed. None of the discrepancies were explained in the trial reports. As endorsed by the revised CONSORT statement,[161] the reporting of amendments to the trial protocol should be routine, particularly when they involve primary study outcomes.

7.4.1. Prevalence of vague outcomes

The prevalence of vague outcome descriptions in protocols was high. Ethics protocols contained vague outcomes more frequently than CIHR protocols (64% versus 44% of trials for efficacy outcomes; 39% versus 12% of trials for safety outcomes), which may be indicative of higher quality protocols among CIHR trials. To be conservative, reported outcomes were not recorded as new outcomes if any vague description existed in the protocol. However, there was ample opportunity for researchers to select the specific data type, time point, and statistical test for such outcomes after examining the data because they were not bound by a pre-specified analysis plan.

7.4.2. Major discrepancies in primary outcome specification

The identification of major discrepancies was reliable and accurate. In a sensitivity analysis, the prevalence of major discrepancies observed in the Ethics cohort did not vary with the definition used to identify primary outcomes. Each discrepancy in the Ethics cohort was verified by two individuals, and the number of major corrections was low (1% of primary outcomes). If uncertainty existed, the benefit of the doubt was given to the trialists and an inconsistency was not recorded. By only verifying trials with discrepant primary outcomes, we erred on the side of being conservative. In other words, it was only possible to detect errors which, upon correction, would lead to more conservative estimates of the prevalence of discrepancies. However, we would not have detected the reverse situation, where outcomes were erroneously recorded as consistent with the protocol when they were in fact inconsistent. Based on the low number of corrections needed in trials with

discrepant outcomes, it is likely that few errors would have been found in the remaining trials with consistent primary outcomes.

Among studies that defined primary outcomes in their protocols or publications, almost two-thirds of Ethics trials and 40% of CIHR trials changed, omitted, or introduced at least one primary outcome. In CIHR trials, the overall impact of major discrepancies in outcome specification tended to favour statistically significant primary outcomes over non-significant outcomes. The impact was more equivocal in the Ethics cohort because the statistical significance of omitted outcomes was unknown. However, it is known that the vast majority of unreported outcomes in the PubMed and CIHR cohorts were non-significant. If unreported outcomes in the Ethics cohort were also assumed to be non-significant, then a tendency to favour significant results was observed. None of the trial reports mentioned that amendments were made to the protocols.

Trials with discrepancies in primary outcomes had different characteristics from those without discrepancies. A slightly higher proportion of Ethics trials with discrepancies were fully funded by industry sources, which is consistent with the higher proportion of multicentre trials and the larger sample sizes observed among these trials. Some sponsors with commercial interests are able to influence the content of trial publications, and may thus be inclined to make amendments that favour their product. Systematic reviews have found that industry funding is associated with larger estimates of treatment effect and with a higher frequency of conclusions favouring the sponsored treatment.[76,77,178,179] On the other hand, non-industry researchers also face academic pressure to

maximise their publications, which can be facilitated by having significant results in their trials.[6]

Opposite trends were observed in the CIHR cohort. CIHR trials with discrepancies were more likely to be smaller, single centre studies. With the absence of trials funded solely by industry in the CIHR cohort, it is possible that smaller sample sizes reflected lower quality studies, such that discrepancies may be more likely to occur. Single centre studies may have also been more amenable to major changes in primary outcomes because the decision had to be made by researchers from only one institution as opposed to the need for consensus from several centres.

Previous assessments of discrepancies in outcome specification are lacking. Individuals with access to both trial protocols and publications, such as those working at drug approval agencies, have occasionally observed differences in clinical endpoints and statistical analyses.[180] A recent trial of celecoxib in arthritis[181] generated controversy over its publication that described only favourable six month results, despite having conducted follow-up for a median of nine months.[182,183] No mention of the longer follow-up or the reason for data censoring was made in the trial report, although subsequent controversy prompted the explanation that differential loss to follow-up was the reason for the six month cut-off.

Our findings are consistent with a pilot study that observed discrepancies in the selection of outcomes to report between study protocols and publications. The pilot study examined the feasibility of systematically comparing trial protocols

to publications with the purpose of identifying inconsistencies in reported outcomes.[98] The authors reviewed protocols for fifteen clinical studies after obtaining consent from the researchers, including two randomised trials. The types of discrepancies observed were similar to those seen in our cohorts. One trial was found to have omitted one outcome and introduced three new outcomes, two of which were statistically significant and one of which had a *p*-value of 0.063. The second trial omitted two outcomes and introduced five new ones, three of which were significant. The remaining thirteen non-trials from their sample demonstrated similar inconsistencies.

Although the outcomes described in the pilot study were not limited to those specified as primary, the discrepancies were consistent with our findings. As found in the thesis cohorts, newly-introduced primary outcomes often consisted of a mixture of significant and non-significant results. This observation suggests that for some trials, factors other than statistical significance played a role in determining whether new outcomes were introduced in the publications.

Our findings are cause for major concern in the interpretation of both individual trial reports and systematic reviews. The main purpose of specifying primary outcomes and analyses in advance of trial commencement is to minimise the potential for misleading results.[129] The selection of these main outcomes should be based on their clinical importance, as determined *a priori*. Pre-specification is protective against the biased selection of results to report by committing the analyses to clinically relevant measures at specific time points prior to the results being known. The protective mechanism is no longer functional if pre-defined outcomes and analyses are subsequently changed, as

this introduces the potential for data-dredging and data-driven reporting of results.[87] As emphasized by the International Conference for Harmonisation guidelines, "redefinition of the primary variable after unblinding will almost always be unacceptable, since the biases this introduces are difficult to assess."[129]

7.4.2.1. Reasons for major discrepancies in outcome specification

Although there is little doubt that applying major changes to primary outcomes after trial commencement is methodologically unsound, the rationale behind them is less clear. Descriptions or explanations of deviations from pre-specified primary outcomes were not provided in any of the trial reports. It is thus unclear whether the intention of trialists was to deceive or whether this was simply an oversight in reporting.

A preference for statistically significant results is one obvious explanation for the inconsistencies in outcomes and analysis plans. This hypothesis is supported by our results, which suggest that discrepancies more often favour statistically significant outcomes over non-significant outcomes, leading to overall bias in favour of significant results. Among five CIHR trials whose survey responses provided reasons for omitting primary outcomes, two stated that the outcomes were omitted due to a lack of statistical significance. It has been well-established that trials with statistically significant results are more likely to be published compared to those with non-significant results.[6] Based on our findings, it is likely that similar bias applies to primary outcomes within published trials.

However, evidence of bias towards significant results was not observed in a large proportion of studies. Most of these trials contained discrepancies that produced a combination of significant and non-significant results, although several contained discrepancies that had unknown significance or favoured non-significant results alone. This suggests that a preference for statistically significant results is not the lone reason for changing the specification of primary outcomes. Our assessment methods may not have detected more subtle interactions, or the discrepancies may have occurred for reasons unrelated to statistical significance.

In terms of our assessment methods, it is possible that some changes were mis-classified as favouring non-significant primary outcomes because of the rigid cut-off of p=0.05 used to distinguish significant and non-significant results. Researchers may regard a p-value of 0.052 or 0.06 as sufficiently interesting to report. It is therefore possible that non-significant outcomes with p-values close to 0.05 may still tend to be reported if they are valued more than outcomes with p-values much higher than 0.05. Based on the definition applied in this thesis, however, such biased behaviour would have been attributed to a preference for non-significant results rather than statistically significant results.

Another possible explanation for the occurrence of discrepancies favouring non-significant outcomes is that our analysis did not distinguish which treatment group was favoured by the significant difference. Significant results may have been omitted if they favoured the non-sponsored comparison group. This could not be evaluated in the thesis cohorts because the direction of effect

was not known for unreported outcomes. These cases would have also been recorded in our analysis as a discrepancy that favoured non-significant results. Similarly, it may have been preferable in equivalence trials to have non-significant outcomes rather than significant ones, which would have also been recorded as a discrepancy that favoured non-significant results. However, equivalence trials were rare in the thesis cohorts.

A third explanation is that the results for other outcomes in the trial may have influenced whether statistical significance was seen as important for a particular outcome. For example, if one primary outcome was already significant, then the statistical significance of other pre-specified primary outcomes may have been less important relative to other competing considerations. One competing consideration is that a particular outcome may have been omitted if it was inconsistent with other outcomes within the same trial, regardless of its statistical result. Another consideration may pertain to the recommendation that only one or two primary outcomes be specified in a trial to avoid issues of interpretation and multiplicity.[129,161] For example, if several pre-specified primary outcomes were significant in a trial, it may have been desirable to report only one of them as primary in order to give the perception that it was the lone primary variable of interest. This was compatible with four of the eight Ethics trials and one of the two CIHR trials that changed a significant outcome from primary to non-primary. All five trials had several pre-specified primary outcomes that were significant, but only one of which was reported as primary while the others were reported as non-primary or omitted. These changes were recorded as favouring non-significant results in our analysis.

Some of the apparent changes in primary outcome specification may also be attributed to deficiencies in protocols rather than biased actions of researchers. In the Ethics cohort, four of ten trials that changed the specification of outcomes from unspecified to primary did not define any primary outcomes in their protocol. It is possible that the affected outcomes had been specified in advance as primary outcomes, but that the trialists failed to describe the specifications in the protocol. This explanation is not applicable to CIHR trials because they all defined primary outcomes in their protocols.

Finally, it is possible that some of the discrepancies in primary outcomes occurred for valid reasons that were not associated with the data analysis. In general, certain modifications to the original trial protocol may be warranted and potentially unbiased, such as the addition of study sites or alterations to inclusion criteria in order to accelerate recruitment of participants. However, major changes to primary outcomes are rarely justified. Exceptional circumstances can be defined in which it may be acceptable to omit protocol-defined primary outcomes, or to introduce new publication-defined primary outcomes. But there is no acceptable justification for changing the specification of outcomes from primary to non-primary, which occurred in over a third of Ethics trials and almost a quarter of CIHR trials.

The omission of a pre-defined outcome can be justified in three situations. The first is when amendments are made prior to trial commencement. This is the only situation in which bias can be definitively excluded, as the timing of the amendment precludes an association with data analysis and results. The omission of a pre-defined outcome after trial commencement can be justified

under two circumstances - either a practical or ethical obstacle prohibits its measurement, or new evidence invalidates its use as a reliable measure. In the latter circumstance, it may then be justified to replace the invalidated outcome with an appropriate new measure. An example of a valid omission was a CIHR trial publication that did not report a particular laboratory measurement because the blood test was withdrawn from the country during the trial. The omission was explained in the Outcomes Survey response but was not mentioned in the trial report. It is important to note that a justifiable reason for omitting outcomes does not exclude bias, and protocol changes must therefore be detailed in publications to enable a critical evaluation of their validity.

Valid reasons may also be proposed for the introduction of new primary outcomes in trial reports, which occurred in a third of Ethics trials and a quarter of CIHR trials. The first reason is the introduction of new outcomes as exploratory analyses, provided that they are explicitly defined as such in the report. These outcomes may have become relevant with the accumulation of new evidence from other studies. The reporting of new surrogate outcomes may also be justified in the context of replacing a primary outcome that was omitted for valid reasons.

It is therefore evident that there may be various complex reasons for making major changes to primary outcomes, and these are not necessarily associated with statistical significance alone. Additional factors likely play a role in rendering trial results more interesting and publishable through changes in major primary outcomes. Most of the reasons outlined are inadequate and

have the potential to introduce bias if they are data driven. However, some potentially valid reasons exist for making major changes to the specification of primary outcomes. As recognised decades ago, "an investigator who formulates or changes his hypothesis after or during the implementation of the study is, in fact, doing a retrospective study, with all the disadvantages that attach to such a study."[184] It is therefore crucial that investigators be transparent in describing all amendments to the original study protocol.

7.4.3. Discrepancies in power calculation outcomes

All discrepancies in power calculation outcomes for the Ethics cohort were verified by two individuals, and no errors were identified. This confirms the accuracy and reliability of the data extraction process. Formal protocol amendments that were submitted to the CIHR were reviewed, but these could not be retrieved for the Ethics cohort. It is possible that formal amendments detailing changes in power calculation outcomes were submitted for Ethics trials, but this would have presumably been uncommon.

29% of Ethics trials and 8% of CIHR trials used a new outcome in their published power calculations, or introduced calculations that had not been described in the protocols. Power calculations are useful in planning the sample size for a trial, and the details of an *a priori* calculation can be useful when interpreting trial results.[142] Firstly, the outcome used should logically correspond to the primary outcome in the trial. This was the basis for recording power calculation outcomes as the main outcome in the absence of an explicit description. Secondly, the reporting of a power calculation in publications may act as a surrogate measure of adequate trial planning.

Finally, it provides an indication of the magnitude of treatment effect that the investigators deem to be clinically relevant. This is particularly important for equivalence trials, where the clinically relevant difference is used to determine whether equivalence can be declared.

Because the interpretation of trial results relies to some extent on the accurate reporting of power calculations, changes must be made explicit in the publication. As with changes to outcome specification, the potential for bias is introduced when amendments to power calculations are made after trial commencement. Though the motive may not be to intentionally deceive, authors who fail to describe such modifications provide readers with an inaccurate basis upon which to interpret trial results.

Prior evidence of changes to power calculations is limited to case reports. One case was identified by chance when the same individual who reviewed the protocol of a trial submitted for funding was also asked to review the final manuscript submitted for publication in the *British Medical Journal*. The reviewer noticed that a larger clinically relevant difference was used in the power calculation reported in the manuscript compared to the protocol, which enabled the trial to appear more adequately powered to support its non-significant result.[185]

While it is widely recognised that the choice of numbers to use in power calculations is often arbitrary,[186] the selection of an outcome to use implies that it is clinical important. Therefore, as with primary outcomes, changes to the power calculation outcome cannot be justified unless feasibility issues or new

evidence renders its use impractical or meaningless. For example, in the case of surrogate endpoints, more valid measures may be developed during the course of a trial. However, such valid reasons would be rare, and it would still be necessary to document the original power calculation and the reasons for making any amendments in the final trial report.

Regardless of the reason for changing power calculation outcomes, the potential for bias cannot be excluded. It is therefore important that any modifications be described in trial reports. Unfortunately, this was not done for any of the trials with power calculation outcome discrepancies in the Ethics and CIHR cohorts.

7.4.4. Major changes in data analysis

The identification of discrepancies in data analysis was difficult. The process involved a comparison of unreported outcomes to reported outcomes within each trial in order to identify identical variables with different data types. The lack of detail in protocol descriptions of outcome data types and analyses presented difficulties in evaluating consistency in these parameters. Efforts to be conservative whenever possible likely resulted in an underestimate of the true number of changes in outcome data types or analyses.

Major changes in data analysis for primary outcomes were observed in only 4%-5% of trials. None of the trial reports mentioned amendments to the original protocol. To our knowledge, this is the first estimate of the prevalence of such discrepancies in a cohort of studies. As with the pre-specification of

outcomes, the pre-specification of analysis plans helps to protect against misleading findings from data-driven analyses.[129]

Almost all of the changed analyses resulted in outcomes that were not statistically significant, but bias based on significance could not be ascertained because the results of the original analysis were unknown. Changes in the data type may have been made because the original analysis produced results that were less easily explained, not as cosmetically desirable, or even in the opposite direction to what was desired. The rationale for analysing a primary outcome using a hypothesis test when it was pre-specified to undergo only descriptive tabulation is also unclear, as the modified primary outcomes in both trials with this discrepancy were non-significant.

Subjectivity in data analysis has been documented in both observational research and randomised trials.[187] Diverse interpretations of the same data can result depending on the selection of analyses to report, creating the potential for bias. By definition, analyses that are not pre-specified are exploratory and this should be noted in the publication. Otherwise, readers may be deceived as to the true nature of the trial findings.

7.4.5. Prevalence of new outcomes reported in publications

New outcomes were verified by two individuals in a random sample of Ethics trials. Few corrections were made to the original data set, indicating that the data extraction process was reliable.

In addition to inconsistencies in primary and power calculation outcomes, the introduction of new outcomes of any specification was common. Over three-quarters of trials reported an average of 4-5 new efficacy outcomes per trial. None of the trial reports mentioned that these outcomes had not been specified in the protocol. These observations are consistent with a previous pilot study.[98] The analysis of outcomes that were not pre-specified in the protocol increases the chance for spurious results. Data-dredging in an attempt to obtain a desirable result is therefore scientifically unsound.

There was no strong evidence of an association between the introduction of new outcomes and statistical significance. As mentioned for primary outcomes, it is possible that the discrepancies reflect inadequately detailed protocols rather than the introduction of new outcomes *post hoc*. However, the method of identifying new outcomes was conservative because outcomes that were vaguely described in the protocol were not considered to be inconsistent. Other factors that may have played a role in the introduction of new outcomes are similar to those discussed for primary outcomes, including the direction of significance, consistency with other pre-specified trial outcomes, and the introduction of important exploratory analyses or new surrogate measures. As with the other deviations from protocols, it is important for the trial report to identify outcomes that were not pre-specified so that an assessment of bias can be made.

7.5. Chapter summary

Major discrepancies between primary outcomes in protocols and publications were observed in a large proportion of randomised trials from the Ethics and

CIHR cohorts. Discrepancies were biased to favour statistically significant primary outcomes. The introduction of new primary or non-primary outcomes was widespread but not associated with statistical significance.

Several potential explanations exist for the high prevalence of major discrepancies in primary outcomes, both related and unrelated to statistical significance. Although some of the discrepancies may be have occurred for valid reasons, such circumstances would not be expected to arise frequently and would not in themselves exclude bias. Failure to describe these changes in the trial report constitutes a mis-representation of the investigators' pre-specified hypotheses about the effects of the intervention - this "sin of omission" should be considered a breach of scientific conduct.[88,188] Any deviations from protocols must be documented and explained in trial publications to enable their incorporation into the overall interpretation of the study.

Chapter 8 - Implications, solutions, and future directions

8.1. Introduction

The first seven thesis chapters presented the background, methods, results, and discussion for an assessment of selective outcome reporting in three cohorts of published randomised trials. This final chapter will discuss the implications of the findings and present possible ways of addressing the problems identified. Future directions for research as well as conclusions will then be presented.

8.2. Summary of key findings

Although it is widely believed to exist, the selective reporting of outcomes has not previously undergone comprehensive review across cohorts of randomised trials. Direct evidence of outcome reporting bias has been limited to case reports with limited generalisability to the overall trial literature.

The results of this thesis provide reliable evidence of significant deficiencies in outcome reporting within published trials. The reporting of outcomes was frequently found to be not only incomplete, but also biased and inconsistent with trial protocols. A median of up to a third of efficacy outcomes and half of safety outcomes per parallel group trial were found to be inadequately reported for meta-analysis, with a median of 4-5 efficacy and 2-3 safety outcomes per trial being omitted entirely from publications. Full reporting of outcomes was consistently biased to favour statistically significant results, with a greater than two-fold increase in odds for a fully reported outcome being significant relative to an incompletely reported outcome. Furthermore, discrepancies between

primary outcomes specified in protocols and publications were observed in approximately one half of randomised trials, with some evidence of further bias in favour of significant results. These findings have serious implications for the conduct and reliability of systematic reviews.

8.3. Implications for systematic reviews

8.3.1. Efforts to obtain incompletely reported data from trialists

The high prevalence of outcome reporting deficiencies observed across the three trial cohorts is unacceptable for systematic reviewers who must rely for the most part on published data to derive reliable pooled effect estimates. Incompletely reported outcomes cannot be included in meta-analyses without making unverifiable assumptions about the nature of the data. In addition to introducing potential bias, incomplete reporting causes an excessive amount of work for systematic reviewers who must then attempt to acquire the necessary data from trialists.

While partially and qualitatively reported outcomes are inadequate for meta-analysis, they at least inform the reviewer of their existence. If subsequent attempts to obtain aggregate data for these outcomes are unsuccessful, then their exclusion from meta-analyses can be interpreted with the results of the review. The existence of unreported outcomes, however, cannot be determined reliably from trial publications alone unless they happen to be mentioned in the Methods section. Systematic reviewers should routinely examine the Methods sections in detail for mention of unreported outcomes.

The Outcomes Survey results for the PubMed cohort indicate that adequate response rates (69%) can be achieved when soliciting information about unreported outcomes, although responses that denied their existence were often unreliable. Trials with full industry funding were less likely to have a survey response than those with partial or no industry sponsorship. However, useful information was obtained from trialists who revealed unreported outcomes - many of which were considered to be primary and clinically important. It should be noted that the PubMed cohort consisted of recently published trials, and lower response rates in general may be expected for systematic reviews involving older trials.

8.3.2. Effect on the reliability of pooled effect estimates

The reliability of systematic reviews depends to a large extent on the identification and availability of all existing data. The evidence presented in this thesis demonstrates that outcome reporting bias results in an under-representation of fully reported, statistically non-significant results within published trials. There is further potential for bias in relation to discrepancies that were frequently observed between primary outcomes specified in protocols and publications. As a result, there would be a tendency for systematic reviews to over-estimate the effects of interventions based on data provided in the published literature. This bias would be in addition to that created by the selective publication of entire studies (study publication bias).

The suppression of non-significant findings, and the resulting over-estimation of treatment effects, create problems for healthcare professionals and policy-makers who rely to a large extent on published evidence to guide their

practice. The worst possible situation would be one in which the over-representation of statistically significant outcomes leads to the promotion of ineffective or potentially harmful interventions. Perhaps more commonly, a marginally beneficial therapy may be considered to be of more value than it deserves. These misguided perceptions could have negative consequences for patients, researchers, and healthcare systems.

From the patient's perspective, several problems may arise. Firstly, patients may be subjected to adverse effects from ineffective therapies and healthcare policies. For example, a trial demonstrating the lack of efficacy and potential harm of routine hospitalisation in uncomplicated twin pregnancies remained unpublished for seven years, resulting in an ongoing policy that was potentially detrimental to patients.[88] Secondly, the introduction of more effective therapies may be delayed by enthusiasm for an inferior one. In the example of the pregnancy trial, alternative policies may not have been investigated if studies demonstrating the harm of routine hospitalisation were not published, or if particular outcomes were selectively suppressed. Finally, the failure to fully disseminate study findings constitutes a breach of the implied contract between investigators and trial participants.[189] A patient's decision to participate in a trial is often driven by an altruistic desire to help future patients and advance medical knowledge.[190-192] These expectations will not have been fulfilled in their entirety if the results of a published study are reported in an incomplete or biased manner.

Misleading conclusions drawn from biased literature would also have a negative effect on researchers. New trials may be initiated based on

hypotheses generated from a biased collection of outcomes. Studies may also be initiated to address questions that have previously been answered but simply not reported in publications. This results in unnecessary duplication of research activities. Perhaps more commonly, new trials may not be initiated if earlier published studies found primarily significant effects. It has been suggested that non-significant studies contradicting previous significant results, trials that show smaller effect sizes than previous ones, or confirmatory trials may have a lower chance of being accepted for publication.[27,193-195]

Furthermore, outcome reporting bias indirectly perpetuates the pervasive belief among researchers that statistically significant results are more interesting and more likely to yield publication. An awareness of this bias may continue to encourage the practice of data-dredging and multiple comparisons with the hope that at least one will produce a statistically significant result.

An over-estimation of the efficacy or safety of an intervention would also be detrimental to the healthcare system. Resources may be wasted on policies that are much less effective clinically or economically than anticipated. Dealing with the potentially adverse consequences of such policies on patients may further increase costs. In addition, trust in the system may be eroded if such mistakes continue to occur.

The problems outlined for patients, researchers, and the healthcare system are amplified in the case of meta-analysis, where conclusions would appear spuriously convincing due to larger combined sample sizes and increased precision. Pooled estimates of effect would, however, be biased towards a

significant effect. It should be noted that the problematic effects of outcome reporting bias are in addition to and in the same direction as those created by study publication bias. Even if all trials were to be published, thereby eliminating study publication bias, there remains a large potential for bias within published studies in the selection of outcomes to report. Due to the negative implications for both the medical community and patients, under-reporting of research is considered unethical and constitutes scientific misconduct.[189] This applies to both suppression of results of whole trials and dishonest selective reporting within publications.

8.4. Implications for trialists

Our comprehensive review of outcomes and analyses specified in trial protocols reveals important deficiencies in the amount of detail provided. Vagueness was a particular problem in the definition of trial outcomes, thus enabling a range of possible interpretations and analyses with *post hoc* selection of favourable results. In order to minimise bias through pre-specification of outcomes, it is crucial that trialists define specific outcomes *a priori*, including the particular variable, time point(s) of interest, and the data type to be used for analysis, such as binary, continuous, or survival data.[129] When there are several ways of measuring a particular variable, one should be selected as primary. This also applies to the pre-specification of specific time points of interest when repeated measures are conducted.[168]

Our results also demonstrate that the reporting of trials is generally poor. Details of important methodological features such as random sequence generation, allocation concealment, and attrition were inadequately reported in

66%-81% of recent PubMed trials. Also, the reporting of 31%-48% of efficacy and 50%-60% of safety outcomes on average in each trial was inadequate for meta-analysis. Furthermore, deviations from protocols occurred in over a half of trials, but none was acknowledged in the publications.

It is clear from our findings that major improvements remain to be made in the reporting of randomised trials. Firstly, details of important methodological features should be provided, as inadequate reporting has been found to be associated with inflated estimates of treatment effect.[131] Secondly, trialists and journals should bear in mind that individual trials will ideally be incorporated in a subsequent systematic review to obtain an overall estimate of treatment effect. It is therefore crucial that sufficient data be provided for all primary and secondary outcomes such that their inclusion in future meta-analyses is possible. Special attention must be given to the full reporting of paired data, which was found to be particularly poor. Finally, deviations from trial protocols should be avoided. If they occur, they must be described in the published reports so that readers can assess the potential for bias.

8.5. Addressing bias and deficiencies in outcome reporting

Researchers have only recently begun to address the issue of outcome reporting bias, with the current study providing the first empirical evidence across cohorts of trials. As with study publication bias, it may prove to be difficult to implement an adequate solution to this important problem. The areas that could be explored include methods to detect or correct for outcome reporting bias; but most importantly, it is necessary to adopt methods to minimise its occurrence.

8.5.1. Detection of outcome reporting bias

While partially and qualitatively reported outcomes can be readily identified from trial publications, difficulties arise when attempting to identify unreported outcomes for a systematic review. Unreported outcomes in published trials can be detected in some cases by examining trial reports and contacting trialists. Firstly, evidence from individual publications can suggest the presence of unreported outcomes if some are mentioned in the Methods but not the Results sections of reports. Secondly, bias that is not obvious in a single trial report may subsequently be uncovered when the entire sample of trials in a review is examined as a whole. Suspicions have indeed been raised by systematic reviewers who note that some common, clinically important outcomes are reported by the majority of trials but not by a few others, with a preference to report results that are statistically significant.[101,103-105] Finally, trialists can be contacted to enquire about the presence of any unreported outcomes and to obtain the necessary data for all incompletely reported outcomes.

Beyond these actions, there does not appear to be a reliable method of detecting outcome reporting bias. Diagnostics have been developed in an attempt to elucidate the presence of study publication bias. They include funnel plot analyses, the trim-&-fill plot, and various selection models.[6] As these methods are based on a single study outcome, it should be possible to adapt them to detect unreported outcomes in published trials. For example, if the funnel plot is asymmetric for a particular outcome, then trials that did not report this outcome may be examined for compatibility with the missing segments of the funnel plot. This could be repeated for each outcome of

interest. However, such diagnostics lack power, and a positive test may be due to outcome reporting bias, study publication bias, true heterogeneity, or other 'small study' effects.[39]

If incompletely reported outcomes are identified but no additional data are obtained from trialists, selection models can be used to estimate the potential effects of outcome reporting bias. One selection model has been proposed for use in sensitivity analyses.[100] The drawback of this method is the assumption that only the outcome with the smallest *p*-value is reported, which does not reflect actual practice. Also, selection models are often based on unverifiable assumptions, which limits their use in systematic reviews.

8.5.2. Prevention of outcome reporting bias

With few exceptions, possible methods of preventing selective outcome reporting are similar to those that have been proposed for over a decade to minimise study publication bias. Clearly it is more desirable to prevent the occurrence of such biases than to attempt to detect and correct for them afterwards - particularly when the diagnostic methods are often not specific nor sensitive, and when the methods of adjustment are based on assumptions that cannot be verified. General causes of selective reporting that can be drawn from our study and the literature include the limited space available for trial reporting in journals, an overall preference for statistically significant results, and the absence of adequate mechanisms to monitor the completeness of trial reporting. The available solutions can thus be divided into three broad approaches to address each of these causes: the expansion of available journal space; eliminating the overall preference for significant results in

medical research; and prospective trial registration with publication of protocols.

8.5.2.1. Journal space

Restricted journal space was commonly cited by trialists as a reason for omitting outcomes. If a large number of outcomes were measured, then limited space may preclude the full reporting of many of them, particularly those with non-significant results. A logical solution to this cause of outcome reporting bias is to expand the space available to fully report trial data.

The advent of electronic journals has presented opportunities for this expansion.[170] Some print journals with online resources, such as the *British Medical Journal*, have allowed authors to present additional data on their website. Trialists are then able at least to mention all outcomes in the print version, and provide full data for incompletely reported outcomes in the online version. The introduction of online journals, such as those published by *BioMed Central*, also helps to overcome the hindrance of space restrictions in print manuscripts. Some journals, such as *Clinical Chemistry* and *Neurology*, even allow the publication of raw data from papers they publish.[196] Unfortunately, the majority of the 271 journals that published trials in the PubMed cohort do not offer these supplementary online resources, nor could they do so with available resources. The use of the internet to display additional information and protocols should thus be expanded in order to help improve the completeness of trial outcome reporting.

8.5.2.2. Changing views about p<0.05

It is well-known that statistically significant results are favoured over non-significant results.[6] Many trialists in our cohort stated that they omitted some of the outcomes in their publications because of non-significant findings. Researchers surveyed in other studies have also failed to submit some of their studies to journals for the same reason.[53,54]

This preference for significant results has persisted for decades. However, it is important that the focus on the *p*-value cut-off of 0.05 be shifted to reflect the fact that an outcome with $p=0.047$ in a statistical test is hardly more significant than one with $p=0.053$. Unfortunately, the dichotomisation of significant versus non-significant outcomes has contributed to biased reporting of study results, both between- and within-studies. Confidence intervals and *p*-values should be interpreted on a spectrum of probabilities rather than dichotomising based on arbitrary thresholds.

Attitudes can change, however, only if journals convince trialists that statistically significant results are not necessary for publication of adequately-powered studies. In fact, it has been argued that the results of a study should not even play a role in the peer review process, since studies with a sound methodological design and important topic should be published regardless of the findings.[197] Peer review of the study protocol alone should thus be sufficient for the decision to publish.[198] *The Lancet* has taken the lead by reviewing protocols of randomised trials and systematic reviews at study inception, and then publishing summaries of selected protocols online.[199] The

journal also makes a commitment to peer review the primary study results for publication.

Others have proposed the creation of a "Journal of negative results",[200] which would eliminate the requirement for statistically significant results in order to be published. Such a journal would help to directly reduce study publication bias more than outcome reporting bias, but it would be an indirect step towards eliminating the "prejudice against the null hypothesis".[25] Although it would be useful to systematic reviewers, the journal would likely not succeed commercially. In addition, such a journal is based on the simplistic but problematic notion that trials can be easily classified as 'positive' or 'negative.'

8.5.2.3. Prospective trial registration

One suggestion that has been proposed for over a decade to address study publication bias is the registration of trials.[50,170,201,202] Prospective trial registration has several advantages. Firstly, it informs systematic reviewers of all existing trials so that unpublished studies can be identified. Although study publication bias would not be avoided unless unpublished data were comprehensively obtained, at least the potential for bias would be known based on the amount of missing data. The amount and nature of missing data is an important consideration in assessing the validity of the review's results. Secondly, trial registration informs researchers of ongoing trials, such that collaboration rather than duplication can occur. Funding and ethics institutions would also benefit from an awareness of ongoing trials in order to assess the need for a new proposed study. Thirdly, potential trial participants would be

able to access information about ongoing studies that they could enter, which may increase recruitment rates.

Several trial registers currently exist. In 1997, the United States Food and Drug Administration Modernization Act legislated a requirement for prospective registration of all trials examining serious or life-threatening conditions; in response, the National Institutes of Health created a publicly-accessible clinical trials registry (www.ClinicalTrials.gov).[203] The United Kingdom (UK) National Research Register was established for studies of interest to the National Health Service, while the UK Medical Research Council has created its own register.[170] The Australian government announced in 2001 that it was establishing a public register of trials by requiring that all public hospitals report information on drug trials in their annual reports.[204] In another positive step towards reducing publication bias, two pharmaceutical companies, Glaxo Wellcome and Schering Health Care, agreed in 1996 and 1998 respectively to register their ongoing clinical trials.[205] Other companies have unfortunately not followed suit.

A *meta*Register of controlled trials was created to link over 20 international trial registers (www.controlled-trials.com), but it is still not comprehensive.[206] A similar online resource provides access to hundreds of registers for ongoing trials (www.trialscentral.org). One difficulty in searching across current registers is the lack of standardised identification numbers for individual trials. This is being addressed by the unique International Standard Randomized Controlled Trials Number (ISRCTN) as well as other numbering systems.

While all of these steps are encouraging, several obstacles exist. The lack of mandatory registration means that many trials remain susceptible to study publication bias. Furthermore, resources are lacking to administer comprehensive international trial registers, even though many ethics and funding bodies already enter information about approved studies into a computer database. A recent setback was the European Commission's decision not to fund an application for resources to promote trial registration in Europe.[202] Finally, it is important to note that trial registration alone does not promote accurate and complete reporting of trials.

8.5.2.4. Publication of protocols

Rationale for publication of protocols

A problem with the current system of research publication is that trialists cannot be held accountable for accurate and complete trial reporting due to the lack of an adequate monitoring mechanism. Even if all trials were to be registered and study publication bias were to be addressed by identifying unpublished studies, outcome reporting bias within published trials would still exist unless complete protocols were also available in the public domain. Publication of protocols is therefore the most appropriate mechanism of addressing the incomplete and selective reporting of trial outcomes.

In addition to having the previously mentioned advantages of trial registration, the availability of trial protocols in the public domain would ensure that any unacknowledged *post hoc* deviations from pre-specified outcomes and analyses could be readily identified. In particular, unreported or newly-introduced outcomes would be known. This would surely serve as a deterrent

to data dredging and selective reporting, and investigators would be more likely to report any changes to the original protocol in order to justify their actions.[183,207] The identification of unreported outcomes and amendments would also provide valuable information to consider when assessing the validity of findings from trials and systematic reviews.

Despite its many benefits, however, the availability of trial protocols would not necessarily prevent the occurrence of partially or qualitatively reported outcomes. The most feasible means of encouraging the full reporting of outcomes remains the expansion of journal space.

Enforcement of protocol publication

Venues for protocol publication include online journals, websites of either print journals or sponsors, and trial registers. In practice, however, protocols are rarely published. In order to ensure comprehensive protocol publication, it is important that compliance be enforced by institutions with the authority to regulate research and its dissemination. These would include institutional or ethics review boards, funding bodies, and journals. Review boards and funding agencies have an ethical obligation to ensure that the results of studies they approve are fully disseminated.[124] Ethics and funding approval should thus be contingent on both protocol publication and a commitment to publish study results. Some have also suggested that a failure to publish should be taken into account by committees when reviewing future applications by the same investigators.[124]

Journals could also adopt part of the responsibility by requiring that all studies submitted for publication have an identification number confirming that its protocol was registered and made publicly available at inception. The International Committee of Medical Journal Editors expressed concern as early as 1989 over the potential for selective reporting of outcomes within a study.[208] Since then, editors have advocated trial and protocol registration,[207,209,210] but have not yet made acceptance for publication contingent on this action. Though not comprehensive, *The Lancet*'s decision to publish some trial protocols while assigning them a unique International Standard Randomized Controlled Trial Number on the *meta*Register of Controlled Trials is a step in the proper direction. The journal's *Information for Authors*[199] also requires that protocols for randomised trials be submitted along with manuscripts for review, although this request is not strictly enforced according to one of the editorial staff (David McNamee, personal communication, 2003).

Calls for the submission of protocols along with manuscripts have previously been rejected. The editor of the *New England Journal of Medicine* has argued against the need for such submissions; suggesting that discrepancies would be rare because "reports of clinical trials are supposed to describe the initial protocol adequately and mention any subsequent modifications. Failure to do so constitutes a breach of scientific conduct."[188] The editor of the *British Medical Journal* rejected a similar request by stating that funding bodies and research ethics committees already scrutinise protocols, and that the extra workload for editorial staff may detract from the quality of service to authors.[211] However, our results demonstrate that such confidence is misplaced; important

modifications to protocol outcomes are common and remain unacknowledged in trial reports.

Although journals remain a logical venue for protocol publication, it would be impossible for them to publish all trial protocols. As a possible alternative venue for protocol publication, trial registers could be expanded to enable the publication of full protocols online. Trial sponsors could also offer to publish their protocols prospectively on their websites. This would not only help to address the issue of selective reporting, but would also increase the credibility of industry-sponsored research. For certain ongoing trials where commercial interests may be compromised by the availability of protocols in the public domain, protocol publication could be delayed until study completion.

8.5.2.5. Summary

Although methods may be developed to detect and correct for outcome reporting bias, they remain inadequate at present. It is evident that preventative measures constitute the most desirable means to address the problem of selective reporting within published trials. Journal space constraints could be eased through the use of online resources in order to encourage the full reporting of all outcomes, while editors could help to counteract the widely-held view that statistically significant results are more publishable than non-significant ones.

The most effective preventative measure that can practically be implemented and enforced by ethics or funding bodies consists of prospective protocol publication. Mandatory publication would not only help to identify unreported

outcomes, but would also enable the identification of *post hoc* amendments and probably deter such biased practices. Unfortunately, prospective protocol publication would not address the issue of partially or qualitatively reported outcomes; until journals are alerted to this issue and increased space for manuscripts is available, systematic reviewers must continue to rely on trialists to provide additional data for these incompletely reported outcomes upon request.

8.6. Future research and unanswered questions

8.6.1. Further research using trial protocols

The thesis methodology and its results highlight the advantages of using trial protocols to evaluate the biased reporting of medical research. Trial protocols provide an *a priori* outline of important methodological features, and the documentation of deviations could be the focus of future studies. However, our study also highlighted the difficulties in obtaining access to protocols and in interpreting vagueness in their description of methods. It is therefore important that institutions such as review boards and funding agencies understand the ethical and scientific rationale for allowing this type of research involving study protocols. Providing access to protocols would not only enable confirmation of our findings in other trial cohorts, but would also allow for extensions to other aspects of trial methodology.

8.6.2. Unanswered questions

This thesis has found a high prevalence of deficiencies in outcome reporting across broad cohorts of randomised trials. However, several questions remain

that warrant further examination. Perhaps the most important issue is whether outcome reporting bias has a significant impact on the results of systematic reviews. This has not been assessed empirically aside from individual case examples. In addition, trialists were not asked about the reasons for discrepancies between protocols and publications. A follow-up survey with this question would provide some indication of the validity of the observed inconsistencies. Furthermore, the issue of selective reporting of analyses based on particular groups of participants could be assessed. This would include subgroup analyses, intention-to-treat analyses, and comparisons between certain study groups in a multi-arm trial. Finally, future studies can investigate whether trials published in journals with online publication space are less prone to selective outcome reporting than those limited to print versions.

8.7. Conclusions

Several important conclusions can be drawn from the results of this thesis. Firstly, it is evident that outcomes within published trials are often inadequately reported for meta-analysis. This poor reporting is influenced by journal space restrictions and is associated with non-significant outcomes, resulting in an over-representation of statistically significant findings in the medical literature. The preference for the most 'interesting' results based on statistical significance in a single trial publication must be balanced with a more global view in terms of the necessary data for future systematic reviews. It is therefore important that all primary and secondary outcomes are specified *a priori* and then fully reported in online or print publications.

Secondly, trialists often change, omit, or introduce new primary outcomes relative to the original trial protocol. While exploratory analyses can be important for generating hypotheses for future studies, trial conclusions should be limited to confirmatory analyses that are specified *a priori*. *Post hoc* modifications have a strong potential for bias, and should be avoided. At the very minimum, these amendments must be explicitly described in trial publications to enable an informed interpretation of study validity. Failure to do so constitutes scientific misconduct.

Finally, prospective and mandatory publication of trial protocols remains the most appropriate mechanism to deter and detect outcome reporting bias. Until this process becomes universal and enforceable through journals as well as ethics and funding bodies, surveys of trialists enable systematic reviewers to obtain important information about unreported outcomes for at least some studies.

Based on the overall findings, it is clear that outcome reporting in publications of randomised trials is often incomplete, biased, and inconsistent with protocols. The selective reporting of outcomes therefore remains a potentially serious threat to the reliabilty of systematic reviews.

Appendix 1. Initial PubMed search strategy for randomised controlled trials published in December, 2000[115]

1) randomized controlled trial [pt]
2) randomized controlled trials [mh]
3) random allocation [mh]
4) double blind method [mh]
5) single blind method [mh]
6) #1 OR #2 OR #3 OR #4 OR #5
7) #6 NOT (animal [mh] NOT human [mh])
8) limit 00/12/01 – 00/12/31

Appendix 2. Final modified PubMed search strategy for randomised controlled trials published in December, 2000[114]

1) randomized controlled trial [pt]
2) controlled clinical trial [pt]
3) randomized controlled trials [mh]
4) random allocation [mh]
5) double blind method [mh]
6) single blind method [mh]
7) cross-over studies [mh]
8) multicenter study [pt]
9) #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8
10) #9 NOT (animal [mh] NOT human [mh])
11) limit 00/12/01 – 00/12/31

Appendix 3. Publications Survey questionnaire for the CIHR cohort

Canadian Institutes of Health Research | Instituts de recherche en santé du Canada

"[*Title*]"
Principal investigator: Dr. [*Name*]

1) Please indicate all source(s) of trial funding: ☐ Government ☐ Industry/Commercial ☐ None ☐ Other _____

2) Have any publications reported your trial results? ☐ Yes - how many? ____ (proceed to **Section A**) ☐ No (proceed to **Section B**)

Section A - For PUBLISHED study

3) We would be grateful if you could please provide us with copies of all resulting publications. Alternatively, please provide the journal citations or send them by e-mail: _____

_____

_____

4) Were there any outcomes that were measured in your trial for the purpose of comparisons between randomized groups, but were **not reported** in any of the above publications? Please exclude baseline characteristics as well as data collected for administrative purposes.

☐ Yes
☐ No (Thank you for your time)

Section B - For UNPUBLISHED study

5) Please indicate current trial status: ☐ Completed ☐ Ongoing ☐ Not started ☐ Started but abandoned due to _____

6) **If completed**, were there any statistically significant inter-group differences in main or subgroup analyses for
a) Any primary outcome(s) ☐ Yes ☐ No b) Any non-primary outcome(s) ☐ Yes ☐ No

7) **If completed**, why was the study unpublished? ☐ Manuscript being prepared ☐ Results not statistically significant
☐ Submitted but rejected ☐ Too few subjects recruited ☐ Other _____

Thank you for your help. When completed, please return by e-mail: achan@cihr-irsc.gc.ca; or fax: (613) 957-8782; or mail: Dr. Karmela Krleza-Jeric, 410 Laurier Avenue W, 9th Floor, Address Locator 4209A, Ottawa, ON, K1A 0W9

Appendix 4. Text of cover letter for the Publications Survey of CIHR trialists

Dear Dr. [*Name*],

Re: "[*Title*]"

The Canadian Institutes of Health Research (formerly the Medical Research Council of Canada) is currently evaluating the outcomes and impact of the randomized controlled trials that it funds. To facilitate this important venture, Drs. Karmela Krleza-Jeric, An-Wen Chan, and myself are conducting a short survey of all trialists who submitted successful applications from 1990-98.

The above-mentioned trial (submitted in [*Year*]) has been included in our sample. I would greatly appreciate your help in answering 3-4 brief questions in the attached survey. Your answers will be recorded in our database in a confidential manner. Any reports arising from this study will present only aggregate data such that no individual study or trialist can be identified.

We would greatly appreciate receiving your response by December 20, 2002. Please return the completed survey by e-mail to **achan@cihr-irsc.gc.ca**. Alternatively, you can fax or post your response using the contact details provided on the questionnaire.

Please do not hesitate to contact either Karmela (613-957-6130), An-Wen (613-941-4598) or myself should you have any questions or concerns.

Thanking you in advance,

Isabelle Schmid
Deputy Director,
Knowledge Translation Programs
Canadian Institutes of Health Research
Ph: (613) 954-6643

Appendix 5. Protocol data extraction form for Ethics and CIHR cohorts (adapted from electronic version)

ADMINISTRATIVE INFORMATION **Study Number** _____

Protocol number _____

Title _____

Primary Investigator Name Dr./Mrs./Mr./Ms._____

Address_____

E-mail _____@_____

Ph _____ Fax _____

Co-Investigators Name Dr./Mrs./Mr./Ms._____

Name Dr./Mrs./Mr./Ms._____

TRIAL CHARACTERISTICS

1) Intervention: ☐ Drug ☐ Surgery/procedure ☐ Equipment ☐ Lifestyle/Counselling

2) Specialty field _____

3) Study Design ☐ Parallel ☐ Factorial ☐ Cross-over ☐ Other _____

4) Estimated sample size _____

5) Power calculation outcome _____

6) Number of Study Groups ☐ 2 ☐ 3 ☐ 4 ☐ Other _____

7) Data Collection Sites ☐ Single centre ☐ Multicentre

8) Funding ☐Commercial ☐Partial commercial ☐Government ☐Private ☐None ☐Other___

9) Blinding ☐ Participant ☐ Assessor ☐ Investigator ☐ Caregiver ☐ Data analyst
☐ Single ☐ Double ☐ Triple ☐ Unclear

SPECIFIED OUTCOMES

Outcome	Safety/ Efficacy	Data type	Specification	Statistical test to be used for analysis

Appendix 6. Operational definitions for recording of trial characteristics

1) Main intervention of interest

- *Drug interventions* consisted of a chemical substance administered to study participants.
- *Surgical/procedural interventions* involved a manual procedure being performed on the participants.
- *Devices* applied to participants would include equipment such as a new type of ventilator.
- *Lifestyle/counselling interventions* consisted of modifications to participants' lifestyles (including exercise/diet) or attendance at counselling/information sessions.

2) Specialty field

- Trials were placed into one of 26 specialty fields based on the condition or population being investigated, as well as the intervention:

- Alternative medicine
- Anaesthesia
- Cardiology
- Critical Care
- Dentistry
- Dermatology
- Endocrinology
- Gastroenterology
- Geriatrics
- Haematology/Immunology
- Infectious diseases
- Nephrology
- Neurology
- Obstetrics/Gynecology
- Oncology
- Ophthalmology
- Otolaryngology
- Paediatrics
- Pharmacology
- Physiology
- Psychiatry/Psychology
- Radiology
- Respirology
- Rheumatology
- Surgery
- Urology

- It is recognised that this classification is subjective and not mutually exclusive. When applicable, classification as physiological or pharmacological studies took precedence over the other specialty designations because they were deemed to be potentially different from other trials. Physiological trials examined the effects of interventions in healthy volunteers, while pharmacological studies examined pharmacokinetics as their primary focus.

3) Study design

- *Parallel group studies* randomised each participant to one of the treatment arms.
- *Cross-over trials* exposed every participant to each treatment and control intervention in a random sequence.
- *Factorial trials* randomised participants to intervention A only, B only, both A and B concurrently, or neither.
- *Split-body trials* randomised separate body parts within each participant to the intervention arms.
- *Cluster trials* randomised groups of participants to the study arms.

4) Study sample size

- Defined as the total number of subjects randomised in the study.
- Individuals were only counted once in cross-over and split-body designs.

5) Power calculation outcome

- The outcome used in the power calculation was recorded

6) Number of comparison groups

- Defined by the number of study groups to which a study participant could potentially be randomised.

7) Number of data collection sites

- *Single centre trials* collected data from one study site.
- *Multicentre trials* collected data from two or more study sites.

8) Funding sources

- *Commercial funding* was defined as the provision of all trial resources by a for-profit organisation.
- *Partial commercial funding* referred to the provision of some study resources by industry sources, such as study drugs or placebo.
- *Government/Private funding* came from non-commercial sources.

9) Blinding

- *Adequate* if one or both of the study participants and outcome assessors had no knowledge of the participants' group allocation, or if the report stated that the trial was blinded, or single-/double-/triple-blind.
- *Inadequate* if neither the participants nor the assessor were blinded.
- *Unclear* if not described.

Appendix 7. Data extraction form for trial reports (adapted from electronic version)

ADMINISTRATIVE INFORMATION **Study Number** _____

1) Study Citation _____

2) Contact Details: Name Dr./Mrs./Mr./Ms._____

Address_____

E-mail _____@_____

Ph _____ Fax _____

ELIGIBILITY

4) Described as randomised ☐ Yes ☐ No (Exclude)

5) Intervention: ☐ Yes ☐None (Exclude)

6) Human subjects ☐ Yes ☐ No (Exclude)

7) Primary report (PubMed cohort only) ☐ Yes ☐ No (Exclude)

8) Economic outcomes/Diagnostic test accuracy ☐ Yes (Exclude) ☐ No

TRIAL CHARACTERISTICS

9) Journal type ☐ General ☐ Specialty _____

10) Type of report ☐ Full Publication ☐ Short report ☐ Letter

11) Specialty field _____

12) Intervention: ☐ Drug ☐ Surgery/procedure ☐ Equipment ☐ Lifestyle

13) Study Design ☐ Parallel ☐ Factorial ☐ Cross-over ☐ Other _____

14) Total sample size (all randomised subjects) _____

15) Number of Study Groups ☐ 2 ☐ 3 ☐ 4 ☐ Other _____

16) Data Collection Sites ☐ Single centre ☐ Multicentre

17) Funding ☐Commercial ☐Partial commercial ☐Government ☐Private ☐None ☐Other_____

18) Blinding ☐ Participant ☐ Assessor ☐ Investigator ☐ Caregiver ☐ Data analyst
☐ Single ☐ Double ☐ Triple ☐ Unclear

QUALITY OF REPORTING

	Adequate (give details)	Inadequate	Unclear
19) Generation of random sequence			
20) Allocation concealment			
21) Defined primary outcome(s)			
22) Power calculation			
23) Handling of attrition			

OUTCOMES

Outcome	Safety/ Efficacy	Data type	Specification	Level of reporting	Analysis used	$p < 0.05$ (If yes, give direction)

Appendix 8. Operational definitions for the quality of reporting of methodological features and the characteristics of primary trial publications

Parameter	Recorded as
Reporting of power calculation	▪ *Adequate* if a power calculation was mentioned with any amount of data in the report ▪ *Inadequate* if none was reported
Specification of primary outcome(s)	▪ *Adequate* if they defined the primary outcome(s) explicitly, or used a main outcome in the power calculation or objectives (see Section **2.3.2.2**) ▪ *Inadequate* if not defined as above
Reporting of random sequence generation	▪ *Adequate* if a truly random method was described, including a random number table, computer-generated random sequence, coin toss, draw of numbers from a container, or variation thereof ▪ *Unclear* if the method was not described
Reporting of allocation concealment	▪ *Adequate* if the individual enrolling trial participants was kept unaware of the randomisation sequence in advance through the use of central randomisation, independent preparation of drugs in sequential unmarked containers, sealed opaque envelopes, post-enrollment randomisation such as a coin toss, or variations thereof ▪ *Inadequate* if the allocation sequence was predictable or known prior to patient enrollment ▪ *Unclear* if not described
Reporting of attrition	▪ *Adequate* if all randomised patients with available data were analysed in their assigned groups (intention-to-treat); and if loss to follow-up was explicitly detailed for each study group ▪ *Inadequate* if the analysis was not intention-to-treat, or if some collected data were excluded for reasons other than loss to follow-up, or if losses to follow-up were not detailed
Journal type	• *General medical journals* published trials from all specialties • *Specialty journals* published trials in a specific field
Type of report	• *Short reports/letters* were classified as this type of publication by the journal ▪ *Full reports* were considered to be full-length publications by the journals

Appendix 9. Preliminary Outcomes Survey questionnaire sent to trialists as a pilot study

"«Title»"
«Journal», January 2001

1. Were there any outcomes that were measured in the trial but not reported in the published paper (please tick below)? Please exclude data collected solely for administrative purposes. ☐ Yes ☐ No

2. **If yes**, what were these outcomes, were they pre-specified, what was their statistical and clinical significance, and what was the primary reason for not reporting each of these outcomes? Please use the table below (continue on next page if needed).

STUDY __	Outcome Type as per protocol[a] (Mark one)			$p < 0.05$ in any comparison (Mark one)		Clinical Importance* (Mark one)			Primary Reason(s) for Not Reporting (Please mark all that apply)				
Outcome	1°	2°	U	Yes	No	1	2	3	Journal space limit	Not clinically important	Not statistically significant	Reported elsewhere	Other (Please specify)

[a] 1° = primary outcome; 2° = secondary outcome; U = unspecified outcome (not defined as primary/secondary)
* 1 = no/little clinical importance; 2 = moderate clinical importance; 3 = high clinical importance

Thank you for your help. Please fax to +44 1865 226962 (United Kingdom) when completed, or mail to:
Dr. An-Wen Chan, Centre for Statistics in Medicine, Institute of Health Sciences, Old Road, Oxford, OX3 7LF, UK

Appendix 10. Outcomes Survey cover letter sent to PubMed trialists explaining the nature of the study and the request for information

CENTRE FOR STATISTICS IN MEDICINE
UNIVERSITY OF OXFORD

Institute of Health Sciences, Old Road, Oxford OX3 7LF, United Kingdom
E-mail: an-wen.chan@ndm.oxford.ac.uk Phone: +44 1865 226996 Fax: +44 1865 226962

«Address»

[Date], 2002

Dear «Contact_Name»,

I am conducting methodological research for my doctorate at the University of Oxford under the supervision of Professor Douglas Altman and Jonathan Deeks. One of the topics of special interest is the reporting of outcomes in trial publications. As part of my study, I am surveying all trialists who published randomised controlled trials in the month of December, 2000, and I would greatly appreciate your assistance.

You recently published an interesting trial in «The» *«Journal»*, entitled "**«Title»**", which I would like to include in our sample. I hope that you will be willing to help us by answering a few questions. Your answers will be confidential, and any publications arising from this study will only report aggregate data such that no individual study can be identified. Attached please find a survey and a list of outcomes mentioned in your paper.

I hope that you will spare the time needed to answer the questions on the enclosed sheet, and then fax or mail it back (no cover letter required). Please do not hesitate to contact me with any questions or concerns.

With my thanks in advance,

Dr. An-Wen Chan, MD
Rhodes Scholar, DPhil. Candidate

Appendix 11. Final Outcomes Survey questionnaire sent to trialists in the PubMed cohort

Note: Question 2 was included only if this information was lacking from the trial report

"[*Title*]"

[*Journal*], December 2000

1. Please tick the source(s) of funding for the study: ☐ Industry ☐ Government ☐ Private ☐ None ☐ Other:_____

2. Was the study data collected from a single site or more than one site (please tick): ☐ Single Centre ☐ Multicentre

3. Were there any outcomes that were measured in the trial but **not reported** in the published paper (please mark below)? Please exclude baseline data as well as data collected solely for administrative purposes. ☐ Yes ☐ No

4. **If yes**, what were these outcomes, were they pre-specified, what was their statistical and clinical significance, and what was the primary reason(s) for not reporting each of these outcomes? Please use the table below (continue on next page if needed).

STUDY __ **Outcome**	Outcome Type as per protocol[a] (Mark one)			$p < 0.05$ in any inter-group comparison		Clinical Importance* (Mark one)			Primary Reason(s) for Not Reporting (Please mark all that apply)					
	1°	2°	U	Yes	No	1	2	3	Journal imposed space limit	Authors' concern about space	Not clinically important	Not statistically significant	Reported elsewhere (give citation)	Other (Please specify)

[a] 1° = primary outcome; 2° = secondary outcome; U = unspecified outcome (not defined in advance as primary/secondary)

* 1 = no/little clinical importance; 2 = moderate clinical importance; 3 = high clinical importance

Thank you for your help. Please fax to +44 1865 226962 (United Kingdom) when completed, or mail to:
Dr. An-Wen Chan, Centre for Statistics in Medicine, Institute of Health Sciences, Old Road, Oxford, OX3 7LF, UK

Appendix 12. Modified Outcomes Survey cover letter text sent by e-mail to CIHR trialists to explain the nature of the study and the request for information

[*Address*]

[*Date*], 2002

Dear Dr. [*Surname*],

Re: '[*Trial title*]'

As mentioned previously, the Canadian Institutes of Health Research is currently evaluating the outcomes and impact of the randomised controlled trials that it has funded. Thank you for your helpful response to our initial survey. Drs. Karmela Krleza-Jeric, An-Wen Chan, and myself are now conducting the follow-up survey of all trialists who submitted successful applications to MRC/CIHR from 1990-98.

As you may recall, the above-mentioned trial (submitted in [*Application Date*]) has been included in our sample. From a review of your study protocol and publications, we have incorporated unreported outcomes into a table in the attached questionnaire. I would greatly appreciate your help in completing the checklist in the table, and adding any other unreported outcomes. Your answers will be recorded in our database in a confidential manner. Any reports arising from this study will present only aggregate data such that no individual study or trialist can be identified.

We would greatly appreciate receiving your response before [*Date*]. Please return the completed table by e-mail to achan@cihr-irsc.gc.ca. Alternatively, you can fax or post your response using the contact details provided on the questionnaire.

Please do not hesitate to contact either Karmela (613-957-6130) or myself should you have any questions or concerns.

Thanking you in advance,

Isabelle Schmid
Deputy Director,
Knowledge Translation Programs
Canadian Institutes of Health Research
Ph: (613) 954-6643

Appendix 13. Modified Outcomes Survey questionnaire sent to CIHR trialists

Canadian Institutes of Health Research / Instituts de recherche en santé du Canada

In the table below, we have listed outcomes that were mentioned in your protocol for comparisons **between randomised groups** but not reported in any of the publications we reviewed. Please complete the table for each outcome. Please add any additional **unreported outcomes** that were measured in your trial for comparisons between randomised groups (exclude baseline data as well as data collected solely for administrative purposes).

STUDY __	$p < 0.05$ in any inter-group comparison		Primary reason(s) for **not reporting** inter-group comparisons (Please mark all that apply)					
Unreported Outcome	Yes	No	Journal imposed space limit	Authors' concern about space	Not statistically significant	Not clinically important	Not intended for inter-group comparisons	Other (Please specify)
Outcome 1								
Outcome 2								
Outcome 3								
Outcome 4								
Outcome 5								
etc.								

Thank you for your help. When completed, please return to Dr. Karmela Krleza-Jeric by e-mail: KKrleza-Jeric@cihr-irsc.gc.ca; or fax: (613) 957-8782; or mail: 410 Laurier Avenue W, 9th Floor, Address Locator 4209A, Ottawa, ON, K1A 0W9

Appendix 14. E-mail messages sent as first reminders with and without questionnaires attached for PubMed trialists

a) With attachment

Dear Dr. [*Name*],

This is a friendly reminder about the short survey I sent to you last month regarding your interesting trial entitled "[*Title*]" ([*Journal*], Dec. 2000). Please ignore the rernainder of this message if you have recently submitted your response or are in the process of doing so, and I will look forward to receiving your response.

To summarise briefly, I am conducting methodological research for my doctorate at the University of Oxford under the supervision of Professor Douglas Altman and Jonathan Deeks. I am interested in the reporting of outcomes in trial publications. As part of my study, I am surveying all trialists who published randomised controlled trials in the month of December, 2000.

I hope that you will be able to spare the time needed to answer a few questions relating to your trial. Please find the survey attached to this message as a Microsoft Word document. Alternatively, I can fax or mail it to you.

Please do not hesitate to contact me with any questions or concerns.

With my thanks in advance,
An-Wen Chan

b) Without attachment

Dear Dr. [*Name*],

This is a friendly reminder about the short survey I sent to you last month regarding your interesting trial entitled "[*Title*]" ([*Journal*], Dec. 2000). Please ignore the remainder of this message if you have recently submitted your response or are in the process of doing so, and I will look forward to receiving your response.

To summarise briefly, I am conducting methodological research for my doctorate at the University of Oxford under the supervision of Professor Douglas Altman and Jonathan Deeks. I am interested in the reporting of outcomes in trial publications. As part of my study, I am surveying all trialists who published randomised controlled trials in the month of December, 2000.

I hope that you will be able to spare the time needed to answer a few questions relating to your trial. If so, please reply to this message with a simple "yes" and I will e-mail you the survey as a Word file attachment. Alternatively, I can fax or mail it to you.

Please do not hesitate to contact me with any questions or concerns.

With my thanks in advance,
An-Wen Chan

Appendix 15. Statistical methods and problems with meta-regression

Statistical methods

Statistical methods for meta-regression can be based on random effects or fixed effects models. Random effects models have been advocated because they account for potential residual heterogeneity, as opposed to the fixed effects assumption that heterogeneity is fully explained by the covariates examined.[122,212]

The estimate of residual heterogeneity used in random effects models can be calculated for weighted regression analyses using a variety of methods. The Bayesian approach is one of the only methods that considers the imprecision in estimating residual heterogeneity, but the choice of non-informative priors can be problematic[213] and the effect has been shown to be small in practice.[122,214] Maximum likelihood methods are asymptotically efficient, but may result in a downward bias of the estimate of residual heterogeneity. Restricted maximum likelihood methods are therefore more appropriate and were used for our analysis in Stata.[122]

Problems with meta-regression

Meta-regression has several drawbacks that limit its reliability.[122,212] Firstly, the results are observational in nature because participants are not randomised between trials. As with epidemiological studies, the potential for confounding must therefore be considered. Secondly, associations observed using average data across trials may not be applicable to individuals within trials - this is known as the ecological fallacy.[215] Thirdly, false positive findings may occur if a large number of covariates are investigated relative to the number of trials, or if the covariates are determined *post hoc*. We therefore pre-specified thirteen trial characteristics, and our meta-regression analysis was restricted to the larger PubMed cohort of 161 trials.

References

(1) Bunker JP, Frazier HS, Mosteller F. Improving health: measuring effects of medical care. *Milbank Q* 1994;72:225-258.

(2) Burgers JS, Grol R, Klazinga NS, Makela M, Zaat J. Towards evidence-based clinical practice: an international survey of 18 clinical guideline programs. *Int J Qual Health Care* 2003;15:31-45.

(3) Egger M, Smith GD. Meta-Analysis. Potentials and promise. *BMJ* 1997;315:1371-1374.

(4) Oxford Centre for Evidence-based Medicine Levels of Evidence. Oxford Centre for Evidence-based Medicine. http://www.cebm.net/levels_of_evidence.asp#levels (accessed 15-6-2003).

(5) Egger M, Jüni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. *Health Technol Assess* 2003;7(1).

(6) Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases. *Health Technol Assess* 2000;4(10).

(7) Villar J, Carroli G, Belizan JM. Predictive ability of meta-analyses of randomised controlled trials. *Lancet* 1995;345:772-776.

(8) LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. *N Engl J Med* 1997;337:536-542.

(9) Cappelleri JC, Ioannidis JP, Schmid CH, de Ferranti SD, Aubert M, Chalmers TC et al. Large trials vs meta-analysis of smaller trials: how do their results compare? *JAMA* 1996;276:1332-1338.

(10) Egger M, Smith GD. Bias in location and selection of studies. *BMJ* 1998;316:61-66.

(11) Cook DJ, Reeve BK, Guyatt GH, Heyland DK, Griffith LE, Buckingham L et al. Stress ulcer prophylaxis in critically ill patients. Resolving discordant meta-analyses. *JAMA* 1996;275:308-314.

(12) Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. *JAMA* 1998;279:281-286.

(13) Stern JM, Simes RJ. Publication bias: evidence of delayed publication in a cohort study of clinical research projects. *BMJ* 1997;315:640-645.

(14) Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991;337:867-872.

(15) Tierney JF, Clarke M, Stewart LA. Is there bias in the publication of individual patient data meta-analyses? *Int J Technol Assess Health Care* 2000;16:657-667.

(16) Gøtzsche PC. Reference bias in reports of drug trials. *Br Med J (Clin Res Ed)* 1987;295:654-656.

(17) Dickersin K. The existence of publication bias and risk factors for its occurrence. *JAMA* 1990;263:1385-1389.

(18) Egger M, Dickersin K, Davey Smith G. Problems and limitations in conducting systematic reviews. In: Egger M, Davey Smith G., Altman D, editors. Systematic reviews in healthcare: meta-analysis in context. 2nd edition. London: BMJ Publishing Group, 2001. p. 43-68.

(19) Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence b(i)ased medicine--selective reporting from studies sponsored by pharmaceutical industry: review of studies in new drug applications. *BMJ* 2003;326:1171.

(20) Hahn S, Williamson PR, Hutton JL, Garner P, Flynn EV. Assessing the potential for bias in meta-analysis due to selective reporting of subgroup analyses within studies. *Stat Med* 2000;19:3325-3336.

(21) Felson DT. Bias in meta-analytic research. *J Clin Epidemiol* 1992;45:885-892.

(22) Rosenthal R. Publication bias, retrieval bias and pipeline effects. Discussion of the paper by Begg and Berlin. *J R Stat Soc [Ser A]* 1988;151:419-463.

(23) Sackett DL. Bias in analytical research. *J Chronic Dis* 1979;32:51-63.

(24) Sterling TD. Publication decisions and their possible effects on inferences drawn from tests of significance - or vice versa. *J Am Stat Assoc* 1959;54:30-34.

(25) Greenwald AG. Consequences of prejudice against the null hypothesis. *Psychol Bull* 1975;82:1-20.

(26) Bozarth JD, Roberts RR. Signifying significant significance. *Am Psychol* 1972;27:774-775.

(27) Smart R.G. The importance of negative results in psychological research. *Can Psychol* 1964;5:225-232.

(28) Sterling TD, Rosenbaum WL, Weinkam JJ. Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *Am Stat* 1995;49:108-112.

(29) Csada RD, James PC, Espie RHM. The file drawer problem of nonsignificant results – does it apply to biological research? *Oikos* 1996;76:591-593.

(30) Mulward S, Gøtzsche PC. Sample size of randomized double-blind trials 1976-1991. *Dan Med Bull* 1996;43:96-98.

(31) Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. *JAMA* 1994;272:122-124.

(32) Moscati R, Jehle D, Ellis D, Fiorello A, Landi M. Positive-outcome bias: comparison of emergency medicine and general medicine literatures. *Acad Emerg Med* 1994;1:267-271.

(33) Davidson RA. Source of funding and outcome of clinical trials. *J Gen Intern Med* 1986;1:155-158.

(34) Hubbard R, Armstrong JS. Publication bias against null results. *Psychol Rep* 1997;80:337-338.

(35) Thornley B, Adams C. Content and quality of 2000 controlled trials in schizophrenia over 50 years. *BMJ* 1998;317:1181-1184.

(36) Lochner HV, Bhandari M, Tornetta P, III. Type-II error rates (beta errors) of randomized trials in orthopaedic trauma. *J Bone Joint Surg Am* 2001;83-A:1650-1655.

(37) Freedman KB, Back S, Bernstein J. Sample size and statistical power of randomised, controlled trials in orthopaedics. *J Bone Joint Surg Br* 2001;83:397-402.

(38) Dickinson K, Bunn F, Wentz R, Edwards P, Roberts I. Size and quality of randomised controlled trials in head injury: review of published studies. *BMJ* 2000;320:1308-1311.

(39) Egger M, Davey Smith G, Schneider M, Minder CE. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629-634.

(40) Berlin J, Begg CB, Louis TA. An assessment of publication bias using a sample of published clinical trials. *J Am Stat Assoc* 1989;84:381-392.

(41) McPherson K. Discussion of paper by Begg and Berlin: Publication bias: a problem in interpreting medical data. *J Am Stat Assoc* 1988;151:445-447.

(42) Dickersin K, Min YL, Meinert CL. Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. *JAMA* 1992;267:374-378.

(43) Dickersin K, Min YI. NIH clinical trials and publication bias. *Online J Curr Clin Trials* 1993;Doc No 50:4967.

(44) Dickersin K. How important is publication bias? A synthesis of available data. *AIDS Educ Prev* 1997;9:15-21.

(45) Misakian AL, Bero LA. Publication bias and research on passive smoking: comparison of published and unpublished studies. *JAMA* 1998;280:250-253.

(46) Dickersin K, Olson CM, Rennie D, Cook D, Flanagin A, Zhu Q et al. Association between time interval to publication and statistical significance. *JAMA* 2002;287:2829-2831.

(47) Dickersin K, Chan S, Chalmers TC, Sacks HS, Smith H, Jr. Publication bias and clinical trials. *Control Clin Trials* 1987;8:343-353.

(48) Sommer B. The file drawer effect and publication rates in menstrual cycle research. *Psychol Women Q* 1987;11:233-242.

(49) Coursol A, Wagner EE. Effect of positive findings on submission and acceptance rate: a note on meta-analysis bias. *Prof Psychol Res Pr* 1986;17:136-137.

(50) Simes RJ. Publication bias: the case for an international registry of clinical trials. *J Clin Oncol* 1986;4:1529-1541.

(51) Stewart LA, Parmar MK. Meta-analysis of the literature or of individual patient data: is there a difference? *Lancet* 1993;341:418-422.

(52) MacLean CH, Morton SC, Ofman JJ, Roth EA, Shekelle PG. How useful are unpublished data from the Food and Drug Administration in meta-analysis? *J Clin Epidemiol* 2003;56:44-51.

(53) Timmer A, Hilsden RJ, Cole J, Hailey D, Sutherland LR. Publication bias in gastroenterological research - a retrospective cohort study based on abstracts submitted to a scientific meeting. *BMC Med Res Methodol* 2002;2:7.

(54) Weber EJ, Callaham ML, Wears RL, Barton C, Young G. Unpublished research from a medical specialty meeting: why investigators fail to publish. *JAMA* 1998;280:257-259.

(55) Rotton J, Foos PW, Vanmeek L, Levitt M. Publication practices and the file drawer problem - a survey of published authors. *J Soc Behav Pers* 1995;10:1-13.

(56) Scherer RW, Dickersin K, Langenberg P. Full publication of results initially presented in abstracts. A meta-analysis. *JAMA* 1994;272:158-162.

(57) De Bellefeuille C, Morrison CA, Tannock IF. The fate of abstracts submitted to a cancer meeting: factors which influence presentation and subsequent publication. *Ann Oncol* 1992;3:187-191.

(58) Kupfersmid J, Fiala M. A survey of attitudes and behaviors of authors who publish in psychology and education journals. *Am Psychol* 1991;46:249-250.

(59) Mahoney MJ. Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognit Ther Res* 1977;1:161-175.

(60) Epstein WM. Confirmational response bias among social work journals. *Sci Tech Hum Values* 1990;15:9-37.

(61) Nylenna M, Riis P, Karlsson Y. Multiple blinded reviews of the same two manuscripts. Effects of referee characteristics and publication language. *JAMA* 1994;272:149-151.

(62) Ernst E, Resch KL. Reviewer bias: a blinded experimental study. *J Lab Clin Med* 1994;124:178-182.

(63) Ernst E, Resch KL. Reviewer bias against the unconventional? A randomized double-blind study of peer review. *Complement Ther Med* 1999;7:19-23.

(64) Resch KL, Ernst E, and Garrow J. Does peer review favor the conceptual framework of orthodox medicine? A randomized controlled study. (The International Congress on Biomedical Peer Review and Global Communications. Prague, 1997). In: *The Cochrane Methodology Review Database, The Cochrane Library*, Issue 1, 2000. Oxford: Update Software.

(65) Abbot NC, Ernst E. Publication bias: direction of outcome less important than scientific quality. *Perfusion* 1998;11:182-184.

(66) Ceci SJ, Peters DP. Peer review - a study of reliability. *Change* 1982;14:44-48.

(67) Anonymous. Manuscript guideline. *Diabetologia* 1984;25:6A.

(68) Hawkins DF. Clinical trials - meta-analysis, confidence limits and 'intention to treat' analysis. *J Obstet Gynecol* 1990;10:259-260.

(69) Melton A.W. Editorial. *J Experimental Psychol* 1962;64:553-557.

(70) Information for authors. Am J Psych. http://ajp.psychiatryonline.org/misc/ifora.shtml (accessed 15-6-2003).

(71) Kerr S, Tolliver J, Petree D. Manuscript characteristics which influence acceptance for management and social science journals. *Acad Manage J* 1977;20:132-141.

(72) Garfunkel JM, Ulshen MH, Hamrick HJ, Lawson EE. Effect of institutional prestige on reviewers' recommendations and editorial decisions. *JAMA* 1994;272:137-138.

(73) Olson CM, Rennie D, Cook D, Dickersin K, Flanagin A, Hogan JW et al. Publication bias in editorial decision making. *JAMA* 2002;287:2825-2828.

(74) Hemminki E. Study of information submitted by drug companies to licensing authorities. *BMJ* 1980;280:833-836.

(75) Bardy AH. Bias in reporting clinical trials. *Br J Clin Pharmacol* 1998;46:147-150.

(76) Bekelman JE, Li Y, Gross CP. Scope and Impact of Financial Conflicts of Interest in Biomedical Research: A Systematic Review. *JAMA: The Journal of the American Medical Association* 2003;289:454.

(77) Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003;326:1167.

(78) Rochon PA, Gurwitz JH, Simms RW, Fortin PR, Felson DT, Minaker KL et al. A study of manufacturer-supported trials of nonsteroidal anti-inflammatory drugs in the treatment of arthritis. *Arch Intern Med* 1994;154:157-163.

(79) Rennie D. Thyroid storm. *JAMA* 1997;277:1238-1243.

(80) Schulman KA, Seils DM, Timbie JW, Sugarman J, Dame LA, Weinfurt KP et al. A national survey of provisions in clinical-trial agreements between medical schools and industry sponsors. *N Engl J Med* 2002;347:1335-1341.

(81) Lauritsen K, Havelund T, Laursen LS, Rask-Madsen J. Withholding unfavourable results in drug company sponsored clinical trials [letter]. *Lancet* 1987;1:1091.

(82) Dong BJ, Hauck WW, Gambertoglio JG, Gee L, White JR, Bubp JL et al. Bioequivalence of generic and brand-name levothyroxine products in the treatment of hypothyroidism. *JAMA* 1997;277:1205-1213.

(83) Nathan DG, Weatherall DJ. Academic freedom in clinical research. *N Engl J Med* 2002;347:1368-1371.

(84) Canadian Coordinating Office for Health Technology Assessment. Annual report 1997-1998. Ottawa: CCOHTA, 1998.

(85) Millstone E, Brunner E, White I. Plagiarism or protecting public health? *Nature* 1994;371:647-648.

(86) Felson DT, Anderson JJ, Meenan RF. Time for changes in the design, analysis, and reporting of rheumatoid arthritis clinical trials. *Arthritis Rheum* 1990;33:140-149.

(87) Mills JL. Data torturing. *N Engl J Med* 1993;329:1196-1199.

(88) Chalmers I. Underreporting research is scientific misconduct. *JAMA* 1990;263:1405-1408.

(89) Bunn F, Alderson P, and Hawkins V. Colloid solutions for fluid resuscitation (Cochrane Review). In: *The Cochrane Library*, Issue 4, 2002. Oxford: Update Software.

(90) Davey Smith G., Egger M. Meta-analysis. Unresolved issues and future developments. *BMJ* 1998;316:221-225.

(91) Williamson PR, Marson AG, Tudur C, Hutton JL, Chadwick D. Individual patient data meta-analysis of randomized anti-epileptic drug monotherapy trials. *J Eval Clin Pract* 2000;6:205-214.

(92) Pocock SJ, Hughes MD, Lee RJ. Statistical problems in the reporting of clinical trials. A survey of three medical journals. *N Engl J Med* 1987;317:426-432.

(93) Begg CB, Berlin J. Publication bias and dissemination of clinical research. *J Natl Cancer Inst* 1989;81:107-115.

(94) Tannock IF. False-positive results in clinical trials: multiple significance tests and the problem of unreported comparisons. *J Natl Cancer Inst* 1996;88:206-207.

(95) Gøtzsche PC. Bias in double-blind trials. *Dan Med Bull* 1990;37:329-336.

(96) Meinert CL, Tonascia S, Higgins K. Content of reports on clinical trials: a critical review. *Control Clin Trials* 1984;5:328-347.

(97) Gøtzsche PC. Methodology and overt and hidden bias in reports of 196 double-blind trials of nonsteroidal antiinflammatory drugs in rheumatoid arthritis. *Control Clin Trials* 1989;10:31-56.

(98) Hahn S, Williamson PR, Hutton JL. Investigation of within-study selective reporting in clinical research: follow-up of applications submitted to a local research ethics committee. *J Eval Clin Pract* 2002;8:353-359.

(99) Gøtzsche PC. Multiple publication of reports of drug trials. *Eur J Clin Pharmacol* 1989;36:429-432.

(100) Hutton JL, Williamson PR. Bias in meta-analysis due to outcome variable selection within studies. *Appl Stat* 2000;49:359-370.

(101) Dickson R, Awasthi S, Demellweek C, and Williamson P. Anthelmintic drugs for treating worms in children: effects on growth and cognitive performance (Cochrane Review). In: *The Cochrane Library*, Issue 4, 2002. Oxford: Update Software.

(102) Fernando MA, Balasuriya S. Effect of Ascaris lumbricoides infestation on growth of children. *Indian Pediatr* 1983;20:721-731.

(103) Whitehead, C., Moss, S., Cardno, A., and Lewis, G. Antidepressants for people with both schizophrenia and depression. In: *The Cochrane Library*, Issue 4, 2002. Oxford: Update Software.

(104) Tinnion ON and Hanlon M. Acellular vaccines for preventing whooping cough in children (Cochrane Review). In: *The Cochrane Library*, Issue 4, 2002. Oxford: Update Software.

(105) Vickers A, Ohlsson A, Lacy JB, and Horsley A. Massage for promoting growth and development of preterm and/or low birth-weight infants (Cochrane Review). In: *The Cochrane Library*, Issue 4, 2002. Oxford: Update Software.

(106) Vickers AJ and Smith C. Homoeopathic Oscillococcinum for preventing and treating influenza and influenza-like syndromes (Cochrane Review). In: *The Cochrane Library*, Issue 4, 2002. Oxford: Update Software.

(107) Felson DT, Anderson JJ, Meenan RF. The comparative efficacy and toxicity of second-line drugs in rheumatoid arthritis. Results of two metaanalyses. *Arthritis Rheum* 1990;33:1449-1461.

(108) Mandel EM, Rockette HE, Bluestone CD, Paradise JL, Nozza RJ. Efficacy of amoxicillin with and without decongestant-antihistamine for otitis media with effusion in children. Results of a double-blind, randomized trial. *N Engl J Med* 1987;316:432-437.

(109) Cantekin EI, McGuire TW, Griffith TL. Antimicrobial therapy for otitis media with effusion ('secretory' otitis media). *JAMA* 1991;266:3309-3317.

(110) Rennie D. The Cantekin affair. *JAMA* 1991;266:3333-3337.

(111) West RR, Jones DA. Publication bias in statistical overview of trials: example of psychological rehabilitation following myocardial infarction [abstract]. In: Proceedings of the 2nd International Conference on the Scientific Basis of Health Services and 5th Annual Cochrane Colloquium; 1997 Oct.8-12, Amsterdam, 82.

(112) McCormack K, Scott NW, Grant AM. Outcome reporting bias and individual patient data meta-analysis: a case study in surgery [abstract]. In: Abstracts for workshops and scientific sessions, 9th International Cochrane Colloquium; 2001 Oct. 9-13; Lyon, France, 34-35.

(113) PubMed overview. National Center for Biotechnology Information. http://www.ncbi.nlm.nih.gov/entrez/query/static/overview.html (accessed 15-6-2003).

(114) Robinson KA, Dickersin K. Development of a highly sensitive search strategy for the retrieval of reports of controlled trials using PubMed. *Int J Epidemiol* 2002;31:150-153.

(115) Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ* 1994;309:1286-1291.

(116) Lefebvre C, Lusher A. Re-designing and updating the Cochrane Highly Sensitive Search Strategy for identifying reports of trials in MEDLINE [abstract]. In: Abstracts for workshops and scientific sessions, 9th International Cochrane Colloquium; 2001 Oct. 9-13; Lyon, France, 77.

(117) Hill CL, LaValley MP, Felson DT. Discrepancy between published report and actual conduct of randomized clinical trials. *J Clin Epidemiol* 2002;55:783-786.

(118) Sterne JA, Davey SG. Sifting the evidence-what's wrong with significance tests? *BMJ* 2001;322:226-231.

(119) Edwards P, Roberts I, Clarke M, DiGuiseppi C, Pratap S, Wentz R et al. Increasing response rates to postal questionnaires: systematic review. *BMJ* 2002;324:1183.

(120) Ioannidis JP, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. *JAMA* 2001;285:437-443.

(121) DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177-188.

(122) Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Stat Med* 1999;18:2693-2708.

(123) Giammona M, Glantz SA. Poor statistical design in research on humans: the role of committees on human research. *Clin Res* 1983;31:572-578.

(124) Mann H. Research ethics committees and public dissemination of clinical trial results. *Lancet* 2002;360:406-408.

(125) Blunt J, Savulescu J, Watson AJ. Meeting the challenges facing research ethics committees: some practical suggestions. *BMJ* 1998;316:58-61.

(126) Savulescu J, Chalmers I, Blunt J. Are research ethics committees behaving unethically? Some suggestions for improving performance and accountability. *BMJ* 1996;313:1390-1393.

(127) Department of Health. Local research ethics committees (HSG(91)5). London: Department of Health, 1991.

(128) Medical Research Council. Personal information in medical research, MRC Ethics Series. London: Medical Research Council, 2001.

(129) ICH Harmonised Tripartite Guideline: Statistical principles for clinical trials. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use

(February 1998, E9). http://www.ich.org/ich5e.html#Design (accessed 15-6-2003).

(130) Berger VW, Bears JD. When can a clinical trial be called 'randomized'? *Vaccine* 2003;21:468-472.

(131) Jüni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ* 2001;323:42-46.

(132) Fossati R, Confalonieri C, Apolone G, Cavuto S, Garattini S. Does a drug do better when it is new? *Ann Oncol* 2002;13:470-473.

(133) Devereaux PJ, Manns BJ, Ghali WA, Quan H, Lacchetti C, Montori VM et al. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. *JAMA* 2001;285:2000-2003.

(134) Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001;357:1191-1194.

(135) Halpern SD, Karlawish JH, Berlin JA. The continuing unethical conduct of underpowered clinical trials. *JAMA* 2002;288:358-362.

(136) Chan KBY, Man-Son-Hing M, Molnar FJ, Laupacis A. How well is the clinical importance of study results reported? An assessment of randomized controlled trials. *Canadian Medical Association Journal* 2001;165:1197.

(137) Pua HL, Lerman J, Crawford MW, Wright JG. An evaluation of the quality of clinical trials in anesthesia. *Anesthesiology* 2001;95:1068-1073.

(138) Schumm LP, Fisher JS, Thisted RA, Olak J. Clinical trials in general surgical journals: are methods better reported? *Surgery* 1999;125:41-45.

(139) Ah-See KW, Molony NC. A qualitative assessment of randomized controlled trials in otolaryngology. *J Laryngol Otol* 1998;112:460-463.

(140) Slim K, Bousquet J, Kwiatkowski F, Pezet D, Chipponi J. Analysis of randomized controlled trials in laparoscopic surgery. *Br J Surg* 1997;84:610-614.

(141) Bhandari M, Richards RR, Sprague S, Schemitsch EH. The quality of reporting of randomized trials in the Journal of Bone and Joint Surgery from 1988 through 2000. *J Bone Joint Surg Am* 2002;84-A:388-396.

(142) Altman D, Moher D, Schulz KF. Peer review of statistics in medical research: Reporting power calculations is important [letter]. *BMJ* 2002;325:491.

(143) Milton J, Logan S, Gilbert R. Well-known signatory does not affect response to a request for information from authors of clinical trials: a randomised controlled trial. In: Abstracts for workshops and scientific sessions, 9th International Cochrane Colloquium; 2001 Oct. 9-13; Lyon, France, 28.

(144) Hetherington J, Dickersin K, Chalmers I, Meinert CL. Retrospective and prospective identification of unpublished controlled trials: lessons from a survey of obstetricians and pediatricians. *Pediatrics* 1989;84:374-380.

(145) Nieminen P, Isohanni M. Bias against European journals in medical publication databases. *Lancet* 1999;353:1592.

(146) Zielinski C. New equities of information in an electronic age. *BMJ* 1995;1480-1481.

(147) Pich J, Carne X, Arnaiz JA, Gomez B, Trilla A, Rodes J. Role of a research ethics committee in follow-up and publication of results. *Lancet* 2003;361:1015-1016.

(148) Cook DJ, Guyatt GH, Ryan G, Clifton J, Buckingham L, Willan A et al. Should unpublished data be included in meta-analyses? Current convictions and controversies. *JAMA* 1993;269:2749-2753.

(149) Moher D, Fortin P, Jadad AR, Jüni P, Klassen T, Le Lorier J et al. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet* 1996;347:363-366.

(150) Jüni P, Holenstein F, Sterne J, Bartlett C, Egger M. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. *Int J Epidemiol* 2002;31:115-123.

(151) Egger M, Zellweger-Zähner T, Schneider M, Junker C, Lengeler C, Antes G. Language bias in randomised controlled trials published in English and German. *Lancet* 1997;350:326-329.

(152) Silagy CA, Middleton P, Hopewell S. Publishing protocols of systematic reviews: comparing what was done to what was planned. *JAMA* 2002;287:2831-2834.

(153) Higgins J, Thompson S, Deeks J, Altman D. Statistical heterogeneity in systematic reviews of clinical trials: a critical appraisal of guidelines and practice. *J Health Serv Res Policy* 2002;7:51-61.

(154) Roberts I, Schierhout G. The private life of systematic reviews. *BMJ* 1997;315:686-687.

(155) Naylor CD. Meta-analysis and the meta-epidemiology of clinical research. *BMJ* 1997;315:617-619.

(156) Johansen HK, Gøtzsche PC. Problems in the design and reporting of trials of antifungal agents encountered during meta-analysis. *JAMA* 1999;282:1752-1759.

(157) Hampton JR, Julian DG. Role of the pharmaceutical industry in major clinical trials. *Lancet* 1987;2:1258-1259.

(158) Blumenthal D, Campbell EG, Anderson MS, Causino N, Louis KS. Withholding research results in academic life science. Evidence from a national survey of faculty. *JAMA* 1997;277:1224-1228.

(159) Davidoff F, DeAngelis CD, Drazen JM, Hoey J, Hojgaard L, Horton R et al. Sponsorship, Authorship, and Accountability. *JAMA: The Journal of the American Medical Association* 2001;286:1232.

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

(161) Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;134:663-694.

(162) Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA* 1996;276:637-639.

(163) Devereaux PJ, Manns BJ, Ghali WA, Quan H, Guyatt GH. The reporting of methodological factors in randomized controlled trials and the association with a journal policy to promote adherence to the Consolidated Standards of Reporting Trials (CONSORT) checklist. *Control Clin Trials* 2002;23:380-388.

(164) Egger M, Juni P, Bartlett C. Value of flow diagrams in reports of randomized controlled trials. *JAMA* 2001;285:1996-1999.

(165) Moher D, Jones A, LePage L. Does the CONSORT statement improve the quality of reports of randomized trials: a controlled before and after evaluation? *Manuscript* 1999.

(166) Elbourne DR, Altman DG, Higgins JP, Curtin F, Worthington HV, Vail A. Meta-analyses involving cross-over trials: methodological issues. *Int J Epidemiol* 2002;31:140-149.

(167) O'Neill RT. Secondary endpoints cannot be validly analyzed if the primary endpoint does not demonstrate clear statistical significance. *Control Clin Trials* 1997;18:550-556.

(168) Phillips A, Ebbutt A, France L, Morgan D. The International Conference on Harmonization guideline "Statistical principles for clinical trials": Issues in applying the guideline in practice. *Drug Information J* 2000;34:337-48.

(169) Davis CE. Secondary endpoints can be validly analyzed, even if the primary endpoint does not provide clear statistical significance. *Control Clin Trials* 1997;18:557-560.

(170) Chalmers I, Altman DG. How can medical journals help prevent poor medical research? Some opportunities presented by electronic publishing. *Lancet* 1999;353:490-493.

(171) Andersen B, Holm P. Problems with p. Significance testing in medical research. Basle: Roche, 1984.

(172) Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. *Br Med J (Clin Res Ed)* 1986;292:746-750.

(173) Kupfersmid J. Improving what is published. A model in search of an editor. *Amer Psychologist* 1988;43:635-642.

(174) Kruskal W, Majors R. Concept of relative importance in recent scientific literature. *Am Statistician* 1989;43:2-6.

(175) Mainland D. Statistical ritual in clinical journals: is there a cure?--I. *BMJ* 1984;288:841-843.

(176) Bland JM, Altman DG. Statistics Notes: One and two sided tests of significance. *BMJ* 1994;309:248.

(177) Moye LA, Tita ATN. Defending the Rationale for the Two-Tailed Test in Clinical Research. *Circulation* 2002;105:3062.

(178) Yaphe J, Edman R, Knishkowy B, Herman J. The association between funding by commercial interests and study outcome in randomized controlled drug trials. *Fam Pract* 2001;18:565-568.

(179) Wahlbeck K, Adams C. Beyond conflict of interest. Sponsored drug trials show more-favourable outcomes. *BMJ* 1999;318:465.

(180) Siegel JP. Editorial review of protocols for clinical trials. *N Engl J Med* 1990;323:1355.

(181) Silverstein FE, Faich G, Goldstein JL, Simon LS, Pincus T, Whelton A et al. Gastrointestinal toxicity with celecoxib vs nonsteroidal anti-inflammatory drugs for osteoarthritis and rheumatoid arthritis: the CLASS study: A randomized controlled trial. Celecoxib Long-term Arthritis Safety Study. *JAMA* 2000;284:1247-1255.

(182) Gottlieb S. Researchers deny any attempt to mislead the public over JAMA article on arthritis drug. *BMJ* 2001;323:301.

(183) Hawkey CJ. Journals should see original protocols for clinical trials. *BMJ* 2001;323:1309.

(184) Porter AM. Clinical research in general practice. *Br Med J* 1971;1:600-602.

(185) Mires G, Williams F, Howie P. Randomised controlled trial of cardiotocography versus Doppler auscultation of fetal heart at admission in labour in low risk obstetric population. Commentary: changes between protocol and manuscript should be declared at submission. Commentary: research governance must focus on research training. Commentary: Approach to power calculations has to be realistic. *BMJ* 2001;322:1457-1460.

(186) Bacchetti P. Peer review of statistics in medical research: the other problem. *BMJ* 2002;324:1271-1273.

(187) Anonymous. Subjectivity in data analysis. *Lancet* 1991;337:401-402.

(188) Relman AS. Editorial review of protocols for clinical trials (reply). *N Engl J Med* 1990;323:1355.

(189) Antes G, Chalmers I. Under-reporting of clinical trials is unethical. *Lancet* 2003;361:978-979.

(190) Welton AJ, Vickers MR, Cooper JA, Meade TW, Marteau TM. Is recruitment more difficult with a placebo arm in randomised controlled trials? A quasirandomised, interview based study. *BMJ* 1999;318:1114.

(191) Sugarman J, Kass NE, Goodman SN, Perentesis P, Fernandes P, Faden RR. What patients say about medical research. *IRB* 1998;20:1-7.

(192) Mattson ME, Curb JD, McArdle R. Participation in a clinical trial: the patients' point of view. *Control Clin Trials* 1985;6:156-167.

(193) Charlton B. Think negative. *New Scientist* 1987;116:72.

(194) Zelen M. Guidelines for publishing papers on cancer clinical trials: responsibilities of editors and authors. *J Clin Oncol* 1983;1:164-169.

(195) Shuster JJ. Publishing clinical trials and dealing with external comparisons. *Control Clin Trials* 1998;19:269-270.

(196) Altman DG. Poor-quality medical research: what can journals do? *JAMA* 2002;287:2765-2767.

(197) Angell M. Negative studies. *N Engl J Med* 1989;321:464-466.

(198) Newcombe RG. Towards a reduction in publication bias. *Br Med J (Clin Res Ed)* 1987;295:656-659.

(199) Information for authors. The Lancet. http://thelancet.com/authorinfo (accessed 15-6-2003).

(200) Maxwell C. Clinical trials, reviews, and the Journal of Negative Results. *Br J Clin Pharmacol* 1981;11:15-18.

(201) Dickersin K. Report from the panel on the Case for Registers of Clinical Trials at the Eighth Annual Meeting of the Society for Clinical Trials. *Control Clin Trials* 1988;9:76-81.

(202) Tonks A. A clinical trials register for Europe. *BMJ* 2002;325:1314-1315.

(203) McCray AT. Better access to information about clinical trials. *Ann Intern Med* 2000;133:609-614.

(204) Current Controlled Trials Ltd. Current Controlled Trials. http://www.controlled-trials.com (accessed 15-6-2003).

(205) Sykes R. Being a modern pharmaceutical company: involves making information available on clinical trial programmes. *BMJ* 1998;317:1172.

(206) Manheimer E, Anderson D. Survey of public information about ongoing clinical trials funded by industry: evaluation of completeness and accessibility. *BMJ* 2002;325:528-531.

(207) Godlee F. Publishing study protocols: making them more visible will improve registration, reporting and recruitment. *BMC News Views* 2001;2:4.

(208) Statements from the Vancouver Group. International Committee of Medical Journal Editors. *BMJ* 1989;299:1394-1395.

(209) Horton R, Smith R. Time to register randomised trials. The case is now unanswerable. *BMJ* 1999;319:865-866.

(210) Tonks A. Registering clinical trials. *BMJ* 1999;319:1565-1568.

(211) BMJ Editor. Major journals should peer review trials at protocol stage (reply). *BMJ* 1997;314:1691-1692.

(212) Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and interpreted? *Stat Med* 2002;21:1559-1573.

(213) Natarajan R, Kass RE. Reference Bayesian methods for generalized linear mixed models. *J Am Stat Assoc* 2000;95:227-237.

(214) Hardy R, Thompson SG. A likelihood approach to meta-analysis with random effects. *Stat Med* 1996;15:619-629.

(215) Morgenstern H. Uses of ecological analysis in epidemiologic research. *Am J Public Health* 1982;72:127-130.

(216) Feinstein AR. X and ipr_p: an improved summary for scientific communication. *J Chronic Dis* 1987;40:283-288.