Assessing the role of optimal information size in systematic reviews

Josep Mª Garcia-Alamino

Kellogg College

Supervisors: Prof. Rafael Perera and Prof. Carl Heneghan

A thesis submitted for the degree of Doctor of Philosophy

Abstract

ASSESSING THE ROLE OF OPTIMAL INFORMATION SIZE IN SYSTEMATIC REVIEWS

Thesis submitted for the degree of Doctor of Philosophy

University of Oxford

Hilary Term 2018

Josep M Garcia-Alamino

Kellogg College

Background and objectives

The concept of optimum information size (OIS) was first proposed by Pogue and Yusuf in a 1997 article published in *Controlled clinical trials*, as "the minimum amount of information required in the collective literature for reliable conclusions about an intervention to be reached". OIS estimates, which are mainly based on standard sample size calculations, are influenced by several variables, including the control event rate (CER), effect size (ES) and heterogeneity.

The overall aims of this thesis are to provide an accurate overview of OIS as depicted in the literature, establish if and how OIS is being used, and assess awareness and perceptions regarding OIS estimation from benchmark bodies and authors of systematic reviews with meta-analyses.

Methods

The literature was systematically reviewed in order to investigate what information has been published to date on OIS and sequential analyses applied to systematic reviews with meta-analyses.

Published meta-analyses that included OIS estimates were reviewed in order to establish historical use and evolution of OIS in meta-analyses. On a more technical level, key statistical parameters — specifically, CER, ES and heterogeneity — were explored in terms of their impact on OIS estimates.

Finally, surveys were conducted, firstly of Cochrane Review Groups (CRGs), in order to determine opinions and policies regarding the OIS methodology, and secondly, of Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement authors and systematic review authors, in order to assess level of awareness and perceptions of OIS.

Results

Although OIS estimates are increasingly included in meta-analyses, most especially since 2010, only a small fraction of all meta-analyses published annually estimate OIS. The use of certain statistical parameters, particularly CER, ES and heterogeneity, is arbitrary and there are no guidelines or consensus in regard to which values to use. While the number of CRGs that had heard of and were considering using OIS estimates was substantially higher in 2016 than in 2010, ultimately only 50% of CRGs stated in 2016 that they were considering using OIS. It was found that few recently published meta-analyses achieved OIS, which would suggest that statistical power may have been inadequate to allow firm conclusions to be drawn. In terms of the range of values used for ES and heterogeneity, the results point to wide variability, although this variability is partially explained by outcome type. No clear and unequivocal relationship could ultimately be established for CER in relation to follow-up duration. While the concept of OIS generates substantial interest among

systematic review authors, PRISMA Statement authors tend to be more sceptical, possibly due to the lack of methodological development of OIS estimation.

Conclusion

OIS estimates are increasingly being used by authors of systematic reviews, yet there are no clear guidelines or consensus regarding the principal statistical assumptions that affect OIS estimates. Most PRISMA Statement authors and systematic review authors expressed an interest in knowing more about OIS, despite a lack of familiarity with the methodology.

Acknowledgements

This thesis is the culmination of several years of work that would not have been possible without the assistance of professional colleagues, friends and family. My first debt is to Professor Carl Heneghan for encouraging me to embark on this research into optimal information size (OIS). His faith in me, his enthusiasm and his guidance have been fundamental in motivating me from the outset. I also owe an enormous debt of gratitude to Professor Rafael Perera for his great patience in dealing with my many doubts and for making time to talk me through issues in person, via Skype or by email. Rafa has deepened my knowledge of technical aspects of this research topic and, in opening my mind to other ways of analysing and interpreting data, has taught me valuable lessons that will be useful to me in the future. Without the help and support he has so generously given me, it is difficult to imagine how I could have continued with and completed this thesis. In our work together, we have shared much that has enriched me personally and I will be eternally grateful for all his professional and personal support in helping me complete this journey.

Other people whom I wish to thank include Dr Xavi Miró, for his constant encouragement and particular interest in the progress of the research; Dr Manuel López-Cano, also for his encouragement but especially for his shared interest as a co-author in applying and investigating OIS; Dr Xavi Garcia, for ongoing support during the years invested in this project; Dr Toni Susín and Dr Malu Calle for their help with mathematical aspects of the research; Nèstor Sànchez for valuable assistance with R; and Ailish Maher for language and copyediting assistance. Special thanks goes to Dr Brigitte Espirac, who, as a former work colleague and now friend, was instrumental in fostering

my enthusiasm for this project from the outset and did everything possible to foster my professional development. I would also like to thank Dr Montse Bosque for professionally supporting me along the way, and Dr Joan Martí and Dr Xavier Bonfill, who opened doors for me in the area of evidence-based medicine.

Finally, this project would never have been completed had it not been for Magda, who has foregone many shared moments and experiences, whether travelling, going to the cinema or simply being together — because her husband was spending days at work but nights and weekends buried under mountains of papers, physically present but mentally absorbed in analysing reams of data. Magda also provided me with direction and support when I felt I was losing my way. I also want to thank my parents and my brothers for the support and encouragement they have always given me, most especially during a crucial period of ill-health during my childhood, and also for renewing my interest and motivation in studying and working. Special thanks to the health professionals who cared for me during that time, most especially María Canet, Dr Montse Melo, Dr Rosa Badía, Montse Moreno, Conchita Milián and Dr Oscar Asensio. Nor can I forget my paternal grandparents, and most especially my grandmother Angelita, for the material and moral support they have always given me in my studies and initial choice of career as a nurse; I still miss them, will always remember them with much love and know that they would be hugely proud of my achievements. And last but by no means least, my thanks to my sons Oriol and Joan. I can only apologise for the time away from them and the lost moments of play, but I hope they someday understand how important effort and perseverance are in achieving goals.

Personal statement

My interest in evidence-based medicine began when, on finishing undergraduate studies as a nurse, I was offered the opportunity to work as a part-time trial search coordinator in the Iberoamerican Cochrane Centre in Barcelona. In my task of exhaustively searching for and locating clinical trials for authors of Cochrane systematic reviews, it occurred to me whether it would not be possible to calculate a sample size for meta-analyses just as is done for clinical trials, so as to establish the number of clinical trials a review would need to ensure that it was adequately powered. Then one day I came across the article by Pogue and Yusuf published in 1998 in *The Lancet*. This publication sparked an interest in optimal information size (OIS) that led me to embark on the research project culminating in this thesis — the fruit of eight years of part-time evidence-based health care studies at the University of Oxford.

The practical and theoretical knowledge acquired in my years of research and the sense of personal achievement from having completed this thesis are invaluable to me. I respectfully hope my efforts will have a positive impact on authors of systematic reviewers and will foster more precise meta-analyses in support of evidence-based health care.

Dedication

This thesis is dedicated to my wife Magda, who has been a constant source of support and encouragement during the years dedicated to this research.

Table of Contents

List of tables

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of figures

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 6

List of abbreviations and acronyms

AMSTAR	A MeaSurement Tool to Assess Systematic Reviews
CER	control event rate
CI	confidence interval
CMG	Cochrane Methods Group
COMET	Core Outcome Measures in Effectiveness Trials
CONSORT	Consolidated Standards of Reporting Trials
CMG	Cochrane Methods Group
CRG	Cochrane Review Group
CTU	Copenhagen Trial Unit
DSMB	data and safety monitoring board
EBHC	evidence-based health care
ES	effect size
GRADE	Grading of Recommendations Assessment, Development and Evaluation
HDRS	Hamilton Depression Rating Scale
HR	hazard ratio
ln	natural logarithm
OIS	optimum information size
OR	odds ratio
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO	International Prospective Register of Systematic Reviews
RCT	randomized controlled trial
ROBIS	Risk of Bias in Systematic Reviews
RR	relative risk
RRR	relative risk reduction
SAMPL	Statistical Analyses and Methods in the Published Literature
SD	standard deviation
tRCT	truncated randomized controlled trial
TSA	Trial Sequential Analysis
WHO	World Health Organization

Chapter 1. Introduction

Is one meta-analysis enough?

1.1. Thesis overview

This thesis explores the evidence for optimal information size (OIS) and its use in meta-analyses of healthcare interventions, in view of the fact that growing numbers of systematic reviews with meta-analyses are published annually (1). Although it is broadly accepted that systematic reviews with meta-analyses provide the most reliable form of evidence for the effects of medical interventions (2), a number of studies have pointed to issues that undermine the validity of meta-analyses, mainly: that many meta-analyses — including those found in Cochrane reviews — yield false positive results due to insufficient information (3); that most meta-analyses do not have sufficient power to identify even moderate effects (4); and that statistical multiplicity resulting from repeated updates of meta-analyses results in unsubstantiated claims (5). The validity of the results reported by meta-analyses is thus called into question, which could affect their acceptance as a definitive source of reliable information regarding interventions.

One proposal that could improve the validity of meta-analysis results is to estimate OIS, defined as "the minimum amount of information required in the collective literature for reliable conclusions about an intervention to be reached" (6). An OIS estimate, which is based on standard sample size calculations, requires that the number of participants (information size) for a meta-analysis should

match the number of patients required for an adequately powered single randomized controlled trial (RCT) (7).

Sample size calculation and OIS are influenced by several key variables, including alpha and beta values (the probabilities of type I and type II error), power, the control event rate (baseline risk), homogeneity and effect size. Deciding which values to use for these variables, however, is difficult, and little guidance is currently available. Thus, values are typically based on those observed or estimated from traditional non-Bayesian meta-analyses. It would also seem relevant to estimate the number of randomized controlled trials that would ensure an OIS.

Further knowledge and debate regarding these issues is necessary to demonstrate the validity of OIS estimates and its acceptance by authors of systematic reviews, subject-area experts (like members of the Cochrane Collaboration) and authors of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement (8).

1.2. Thesis aim

The overall aim of this thesis is to provide an accurate overview of the OIS concept as described in the literature, to determine if and how it is being used and to collate views regarding the OIS methodology by authors of meta-analyses and authors of key statements. The thesis sets out to test 2 main hypotheses: (a) the use of OIS in systematic reviews with meta-analyses remains low; and (b) the methods for OIS estimates in meta-analysis, including estimates needed as part of the statistical assumptions, have not been broadly agreed by the research community.

1.3. Research objectives

The research objectives of this thesis are:

- To review the literature regarding methods for estimating OIS in meta-analyses.

- To review meta-analyses published from 2006 to 2015 that have included OIS estimates.
- To describe the statistical assumptions currently used to estimate OIS.
- To determine Cochrane Collaboration perceptions of OIS and the guidance available regarding OIS.
- To assess how different statistical assumptions regarding control event rate, effect size and heterogeneity affect OIS estimates.
- To quantify the number of meta-analyses published by the Cochrane Collaboration and in the top 5 medical journals that achieve OIS.
- To explore the relationship between the control event rate and the duration of follow-up in individual RCTs.
- To determine the level of acceptance of OIS among systematic reviewers and PRISMA Statement authors.

1.4. Overall structure of the thesis

Chapter 2. Randomized controlled trials, meta-analyses and optimal information size

Chapter 2 reviews the importance of sample size in RCTs and the implications of an inadequate sample size from the protocol stage to the publication of findings. Also discussed are the statistical implications — in terms of type I error, lack of power, etc — of inadequate sample size for systematic reviews with meta-analyses. Also reviewed is the historical evolution and use of sequential methods and their adaptation to the calculation of information size in systematic reviews with meta-analyses.

Chapter 3. Optimal information size in published meta-analyses

Chapter 3 reports findings for (a) a systematic review of methods and statistical assumptions used for OIS estimates in published meta-analyses (Cochrane and non-Cochrane) that have included this

estimate, and (b) a survey addressed to Cochrane Review Groups to collate awareness and perceptions of OIS and determine the guidance offered on OIS, if any.

Chapter 4. Impact of heterogeneity and effect size on optimal information size estimates

Chapter 4 focuses on effect size and heterogeneity and on understanding how different value assumptions regarding these parameters affect OIS estimates. Also quantified is the number of meta-analyses published by the Cochrane Collaboration and in the top 5 medical journals that achieve OIS.

Chapter 5. Impact of trial follow-up duration on the control event rate

Chapter 5 analyses a sample of published meta-analyses and individual RCTs within those meta-analyses to evaluate the possible impact of RCT follow-up on the control event rate. The control event rate is also evaluated for communicable and noncommunicable diseases and for pharmacological and non-pharmacological interventions.

Chapter 6. International survey: awareness and perceptions of optimal information size

Chapter 6 reports the results of a survey addressed to authors of published meta-analyses and authors of the PRISMA Statement. Authors were surveyed regarding their awareness and perceptions of OIS and about the possible inclusion of this methodology in the PRISMA Statement.

Chapter 7. Discussion and conclusions

Chapter 7 summarizes the findings of the thesis and its contribution to knowledge regarding use of OIS estimates in meta-analyses. As well as addressing the methodological strengths and limitations of the thesis, the final discussion provides some recommendations on how to improve use of OIS.

1.5. References

1. Ioannidis JP. The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. Milbank Q. 2016;94(3):485-514.
2. Clarke M, Brice A, Chalmers I. Accumulating research: a systematic account of how cumulative meta-analyses would have provided knowledge, improved health, reduced harm and saved resources. PLoS One. 2014; 9(7):e102670.
3. Brok J, Thorlund K, Gluud C, Wetterslev J. Trial sequential analysis reveals insufficient information size and potentially false positive results in many meta-analyses. J Clin Epidemiol. 2008; 61(8):763-9.
4. Jackson D, Turner R. Power analysis for random-effects meta-analysis. Res Synth Methods. 2017; 8(3):290-302. doi: 10.1002/jrsm.1240.
5. Polanin JR, Pigott TD. The use of meta-analytic statistical significance testing. Res Synth Methods. 2015;6(1):63-73.
6. Pogue JM, Yusuf S. Cumulating evidence from randomized trials: utilizing sequential monitoring boundaries for cumulative meta-analysis. Control Clin Trials. 1997;18(6): 580–593.
7. Wetterslev J, Imberger G. The information gap between the required and the actual accrued information size in the meta-analysis of antenatal magnesium sulphate to prevent cerebral palsy in preterm infants. BJOG. 2011; 118(7):890.
8. Page MJ, Moher D. Evaluations of the uptake and impact of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) Statement and extensions: a scoping review. Syst Rev. 2017; 6(1):263.

Chapter 2. Randomized controlled trials, meta-analyses and optimal information size

2.1. Aim and objectives

The aim of this chapter is to review and discuss the literature on sample size, hypothesis testing, power, measurement error (random error and bias) and statistical multiplicity as key concepts in randomized controlled trials (RCTs) from the perspective of their impact on systematic reviews with meta-analyses. Discussed in depth in relation to meta-analyses is the concept of the number of patients required to ensure an adequately powered study.

2.2. Randomized controlled trials

An RCT is the most rigorous way to determine whether a cause-effect relationship exists between a treatment and an outcome. Random allocation ensures that there are no systematic differences — in either known or unknown factors — between intervention groups that may affect outcomes. Double blinding, in which both the subjects and investigators are unaware of who is getting which specific treatment (1), ensures that the preconceived views of trial subjects and clinicians cannot systematically bias the assessment of outcomes. Intention-to-treat analyses maintain the advantages of random allocation, which may be lost if subjects are excluded through, for example, withdrawal or failure to comply (2).

Provided they are appropriately designed, conducted and reported, RCTs are the gold standard in evaluating healthcare interventions (3). However, a number of concerns have been raised regarding RCT methods, execution and reporting, with early studies indicating that key issues such as primary

outcomes, random sequence generation, allocation concealment and sample size calculation details have been inadequately described in more than half of the publications evaluated (4). Efforts to improve the reporting of RCTs in the 1990s eventually led to publication of the Consolidated Standards of Reporting Trials (CONSORT) Statement, whose most recent version is the CONSORT 2010 Statement.

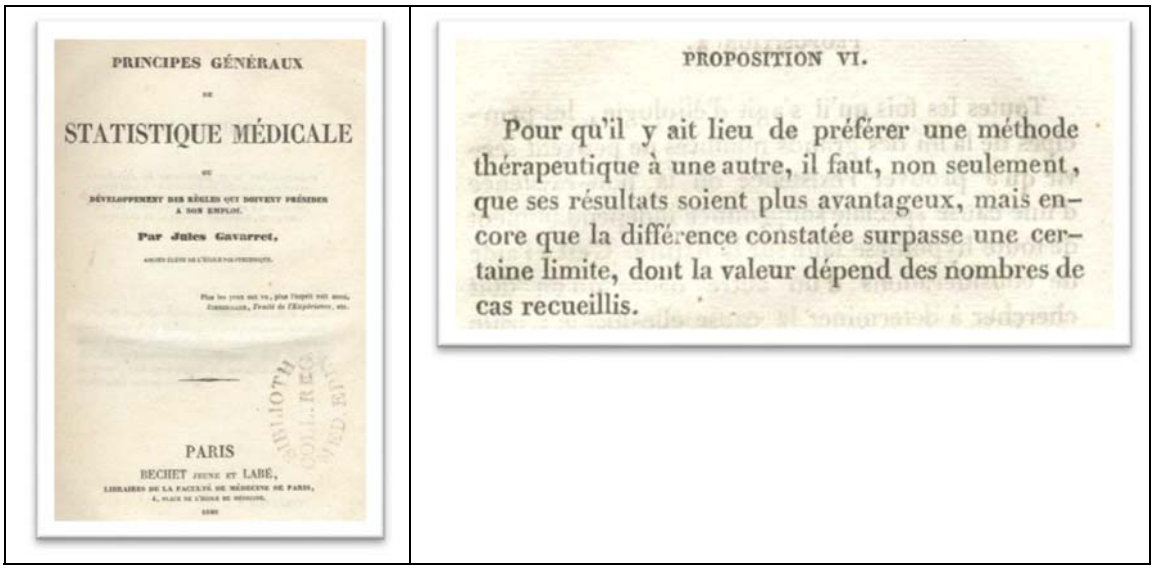
Sample size should be estimated in the planning phase of an RCT and should be included in the study protocol. One problem with RCTs is that the samples are often too small for an effect to be detected — as was demonstrated more than 30 years ago (5) and again more recently (6). This problem of underpowered RCTs can be addressed by systematically reviewing and conducting a meta-analysis of the information from multiple RCTs, given that, hypothetically, the power of a meta-analysis will be greater than that of any individual study (7).

Sample size for RCTs is calculated for the outcome that is considered the most relevant — i.e. the primary outcome — but also, in many cases, for secondary outcomes. Nonetheless, it has been observed that a specified sample size may be inadequate to assess secondary outcomes (8). Proper outcome selection is crucial when designing an RCT that aims to compare different interventions (9) and may also be useful to ensure that the sample size to evaluate the primary outcome has sufficient power. An initiative of the Core Outcome Measures in Effectiveness Trials (COMET) project (10) — supported by trialists, systematic reviewers, trial funders, trials registries and policy makers, among others — has as its objective to tackle the issue of inconsistency and outcome-reporting bias in RCTs by making it easier for results to be compared, contrasted and combined (as appropriate) using an agreed standardized set of outcomes. This set consists of the outcomes that should be measured and reported in all trials of a specific clinical condition (11). The COMET Core Outcome Set (COS) initiative may also help improve sample size estimation for RCTs.

2.2.1. Sample size, hypothesis testing and power

The most common goal of an RCT is to determine whether any difference exists in outcomes between an intervention (treatment) group and a control group, although other scenarios might be, for instance, confirming the equality or equivalence of intervention and control groups (12). Sample size for an RCT is calculated as the number of patients who must participate to obtain a reliable estimate of the magnitude of the effect. An RCT based on a small sample, however, lacks statistical power, which means that a positive treatment effect may not be found or, if found, be exaggerated (13). From a frequentist (as opposed to a Bayesian) perspective, sample size should be large enough to ensure that the chance of spurious differences occurring is quantifiably small. Moreover, the less dramatic the expected effect, the larger the sample should be to demonstrate a significant difference between intervention and control group results. Indeed, studies have demonstrated that small studies tend to show greater intervention effects than larger studies (14), although it is not entirely clear whether this is due to a true effect or just to random chance combined with publication bias. Sample size is crucial, ultimately, because confidence in the results of an RCT is directly related to the size of the sample (15).

Although it is difficult to identify the first clinical trials that included sample size estimates, some publications dating back to the 19th and early 20th century refer to the impact of the number of observations on confidence in the results of clinical studies (Figures 2.1 and 2.2).



PRINCIPES GÉNÉRAUX

DE

STATISTIQUE MÉDICALE

OU

DÉVELOPPEMENT DES RÈGLES QUI DOIVENT PRÉSIDER À SON EMPLOI.

Par Jules Gavarret,

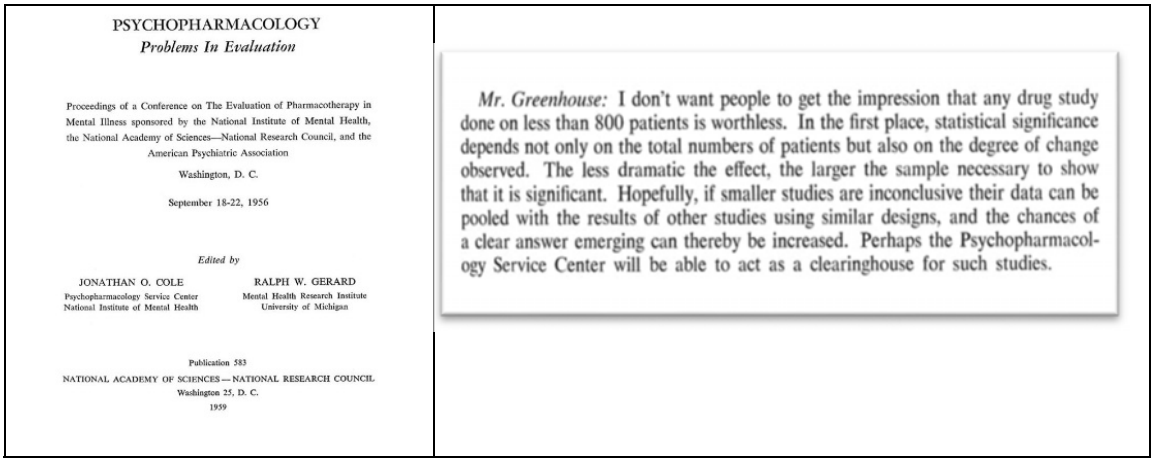
PARIS

BECHET JEUNE ET LABÉ,

PROPOSITION VI.

Pour qu'il y ait lieu de préférer une méthode thérapeutique à une autre, il faut, non seulement, que ses résultats soient plus avantageux, mais encore que la différence constatée surpasse une certaine limite, dont la valeur dépend des nombres de cas recueillis.

Figure 2.1. A publication on medical statistics (1840).
Gavarret LDJ. *Principes généraux de statistique médicale: ou développement des régles qui doivent présider à son emploi,* Paris: Bechet Jeune & Labé, 1840.
[Translation: *General principles of medical statistics: or the development of rules that must govern their use.* Proposition VI. To be able to decide in favour of one treatment method over another, it is not enough for the method to yield better results; the difference found must also exceed a certain limit, the extent of which is a function of the number of observations.]
Source: James Lind Library (http://www.jameslindlibrary.org/).



PSYCHOPHARMACOLOGY
Problems In Evaluation

Proceedings of a Conference on The Evaluation of Pharmacotherapy in Mental Illness sponsored by the National Institute of Mental Health, the National Academy of Sciences—National Research Council, and the American Psychiatric Association

Washington, D. C.

September 18-22, 1956

Edited by

JONATHAN O. COLE
Psychopharmacology Service Center
National Institute of Mental Health

RALPH W. GERARD
Mental Health Research Institute
University of Michigan

Publication 583
NATIONAL ACADEMY OF SCIENCES — NATIONAL RESEARCH COUNCIL
Washington 25, D. C.
1959

Mr. Greenhouse: I don't want people to get the impression that any drug study done on less than 800 patients is worthless. In the first place, statistical significance depends not only on the total numbers of patients but also on the degree of change observed. The less dramatic the effect, the larger the sample necessary to show that it is significant. Hopefully, if smaller studies are inconclusive their data can be pooled with the results of other studies using similar designs, and the chances of a clear answer emerging can thereby be increased. Perhaps the Psychopharmacology Service Center will be able to act as a clearinghouse for such studies.

Figure 2.2. Discussion of sample size in a publication on psychopharmacology (1959).
Cole JO, Gerard RW (eds). *Psychopharmacology. Problems in Evaluation.* National Academy of Sciences, Publication 583, Washington DC, 1959, p 371.
Source: James Lind Library (http://www.jameslindlibrary.org/).

Estimating a statistically appropriate sample size is directly related to the concepts of hypothesis testing and false positive and false negative results (type I and II errors, respectively) and also to the statistical power of an RCT.

An RCT begins with a hypothesis, which can be either a null hypothesis (H_0) or an alternative hypothesis (H_1). The null hypothesis usually reflects no effect, whereas the alternative hypothesis reflects an effect in either direction. When the null hypothesis is true but is rejected, a type I error is said to occur, and when the null hypothesis is false but is not rejected, a type II error is said to occur. Type I and II errors are also commonly referred to as false positives and false negatives, respectively. While trialists compile information and evidence from their study to reject or not reject their a priori null hypothesis, they will generally want to observe results that reject their null hypothesis.

Statistical testing is used to quantify whether there is sufficient evidence to refute a null hypothesis (16). Different types of statistical tests — e.g. the paired t-test, independent t-test or analysis of variance (ANOVA) — are used depending on the type of data in the study (17). Type I and type II errors are both quantified as statistical probabilities, as alpha (α) and (β) levels of significance, respectively. Typically used alpha values — 0.05 and 0.01 (18) — are deliberately small, to avoid false positive findings. The beta level of significance is conventionally set to 0.20 (or sometimes 0.10) to reflect a less than 20% (or 10%) chance of a false negative conclusion. By convention, a type I error is tolerated less than a type II error, because a false positive inference — i.e. rejecting the null hypothesis when it is true — has more serious consequences than a false negative inference (16).

Statistical power is intimately related to type I and II error, as it reflects the probability of obtaining a statistically significant result when there is a genuine difference in treatments (19); in other words, power represents the chance of detecting a specified effect if it really does exist. The power of a

test, conventionally set to 80%, is the pre-study probability of a test rejecting the null hypothesis when it is false (20). Power is therefore inversely related to the probability (i.e. 20%) of making a type II error (21). A simple definitional equation is, therefore, power= 1-β.

The power of a statistical test depends on several factors, some uncontrollable (e.g. effect size (ES) and population variance), and others controllable, namely, sample size and the alpha value (16). Assuming all other parameters to be constant, lowering the alpha value will decrease the probability of a type I error but increase the probability of a type II error. On the other hand, increasing the sample size will enhance power and reduce the risk of a type II error (17) — but will also result in a costlier RCT.

2.2.2. Sample size: from protocol to publication

Sample size calculations are widely recognized to have an important impact on planning, interpretation and conclusions for RCTs (22). The CONSORT 2010 Statement indicates that these calculations must be reported and justified in published articles. Yet sample size calculations are often erroneous, inadequately reported and frequently based on inaccurate assumptions (23).

Problems with sample size in RCTs frequently start in the planning phase, with estimates as recorded in protocols often differing from what actually features in publications. In their classic 2008 study of a comprehensive cohort of RCTs, An-Wen Chan et al. (22) compared sample size calculations and data analysis methods as described in protocols and the corresponding publications. They identified a high frequency of unacknowledged discrepancies and poor reporting, finding that a mere 15.7% of trials fully and consistently reported requisite components of sample size calculations in both protocols and publications. Charles et al. (24) analysed RCTs reported in 6 general medical journals published in 2005 and 2006, finding that only about 34% adequately described sample size calculations; they also found, for 62% of the RCTs, that the primary outcome reported in the

protocol (on which the sample size calculation was theoretically based) differed from the primary outcome referred to in the final report.

2.2.3. Adaptive designs and truncated trials

Although sample size is usually estimated before an RCT begins (i.e. the RCT is based on a fixed sample size), in some cases an adaptive design is used to allow for a variable sample size. In such cases, several interim analyses are performed until a given ES is achieved.

One form of adaptive design, called sequential design (or sequential analysis), has been used in confirmatory RCTs for several decades, as documented in a 25-year review by Todd (25). A sequential design provides a hypothesis-testing framework for decision making regarding early stoppage of an RCT, whether because a benefit has been identified (26) or because of futility (27). This kind of adaptive design has ethical and economic advantages as it means that costly RCTs can be stopped as soon as there is enough evidence to answer the research question (26). Such RCTs are called 'truncated trials' (tRCTs) (26). The interim analyses conducted for this kind of RCT usually have built-in statistical stopping rules, with sample size estimated using some kind of sequential analysis technique.

2.2.4. Measurement error and missing data

A crucial principle in research is to minimize random error and systematic error (also known as bias), i.e. increase precision and accuracy, respectively.

Random (or non-systematic) error is outside the control of the researcher as it is unpredictable and unavoidable. Precision is limited by random error and usually may be improved by repeating measurements, whose precision reflects the level of agreement between several measurements of the same quantity. Increased sample size is the main recommendation to avoid the imprecision resulting from random error (28).

Systematic error (bias) — often difficult to detect even by experienced researchers — is predictable error that is not due to chance but systematically shifts in a single direction from the true value. In terms of accuracy, it reflects how close measurements are to the true value of the quantity being measured. Systematic error results from flaws in the design, conduct or analysis of a study that may lead to underestimation or overestimation of the true intervention effect (29).

2.3. Systematic reviews with meta-analyses

Systematic reviews, which are broadly considered to provide the most reliable form of evidence regarding the effects of a medical intervention (30), aim to identify all studies answering a particular question, appraise their quality and summarize their results using a rigorous scientific methodology. Clinicians often use them to keep up to date in their field and they are also essential for the development of clinical practice guidelines. Certain funding agencies require systematic reviews to be conducted as a justification for additional research (31).

Meta-analyses are frequently included in systematic reviews as a statistical approach to summarizing the results of individual studies (32), with the aim being to combine findings so as to increase the statistical power and precision of the effect estimate (33).

In 1992, Chalmers et al. (34) introduced the term ‘cumulative meta-analysis’ to describe a statistical procedure to retrospectively calculate, each time the results of a further trial in a series became available, summary estimates for the results of similar trials. Cumulative meta-analyses have challenged initially favourable results observed in individual RCTs or earlier meta-analyses. A recent work by Clarke et al. (35) highlighted how a cumulative meta-analysis of existing evidence might have led to better decision making and choices about new RCTs over decades of research, concluding that this would have resulted in “an earlier uptake of effective interventions in practice,

less exposure of trial participants to less effective treatments and reduced waste resulting from unjustified research".

2.3.1. False positives and power

A growing body of evidence shows that statistically significant results in early meta-analyses had a high proportion of type I errors (36). A recent simulation study showed that including primary studies with false-significant effects — especially when numerous and large — may bias effect estimates in a meta-analysis, resulting in type I error rates failing to meet the pre-specified nominal level, usually 5% (37).

Type I error is typically fixed at 0.05 to ensure a smaller than 5% chance of drawing a false positive conclusion. However, only 32% of meta-analyses preserve type I error risk below 5%, according to Imberger (38). Furthermore, in underpowered meta-analyses subject to continuous updating, the risk of type I error tends to be higher than the conventional probability of 5% (36).

The problem of sufficient statistical power in meta-analyses may be even greater than suggested when power analyses intended for single trials are used. Imberger et al. (38) found that only 12% of meta-analyses had power ≥80% and that as many as 50% had power ≤50%. Turner et al. (39) showed that underpowered RCTs made up the entirety of the evidence in most Cochrane meta-analyses, with 70% of the corresponding studies having less than 50% power to detect a 30% relative risk reduction.

Power is highly correlated with the cumulative number of patients and events, with empirical studies suggesting that more evidence accumulated over time may point to many apparently large intervention effects being substantially overestimated in early studies (40). Imberger et al. (36), in their analysis of 100 meta-analyses (selected from 4,736 Cochrane reviews) meeting specific inclusion criteria, demonstrated that the majority of published Cochrane meta-analyses were

underpowered. Brok et al. (41) demonstrated that 39% of the meta-analyses published by the Cochrane Neonatal Group had statistically significant but potentially false positive results, concluding that many of those meta-analyses were substantially underpowered. AlBalawi et al. (42) found that 45% of high-quality Cochrane meta-analyses for heart disease were too small to detect or reject a 25% relative treatment effect; they also estimated that 17% of meta-analyses with apparently statistically significant results reported potentially false positive findings. The same authors concluded that a statistically significant or non-significant result — even for a methodologically robust high-quality meta-analysis — does not guarantee conclusive findings. In their large meta-epidemiological study, Dechartres et al. (43) found smaller trials to have significantly larger estimates of treatment effects and also that effect estimates in meta-analyses differed solely based on trial sample size, with, on average, stronger effect estimates in small-to-moderately sized trials than in larger trials.

2.3.2. Statistical multiplicity

The more analyses that are done, the more likely it is that some tests will be found to be statistically significant even though there may be no true effect (44). Statistical multiplicity, which refers to the use of more than one null hypothesis test, increases the risk of false positive findings (45, 46). Statistical multiplicity therefore poses a challenge when trying to check the reliability and comparability of conclusions in RCTs and in meta-analyses (47).

Although the concept of statistical multiplicity has been keenly debated in the context of single RCTs, it has received little attention in the context of systematic reviews with meta-analyses (47). Meta-analyses are commonly updated when new trials are published, i.e. analyses are repeated on accumulating data and significance testing is repeated, which, if done with the conventional p-value

criterion (typically two-sided at the 5% significance level), is prone to exacerbate the risk of type I error (48) (Figure 2.3).

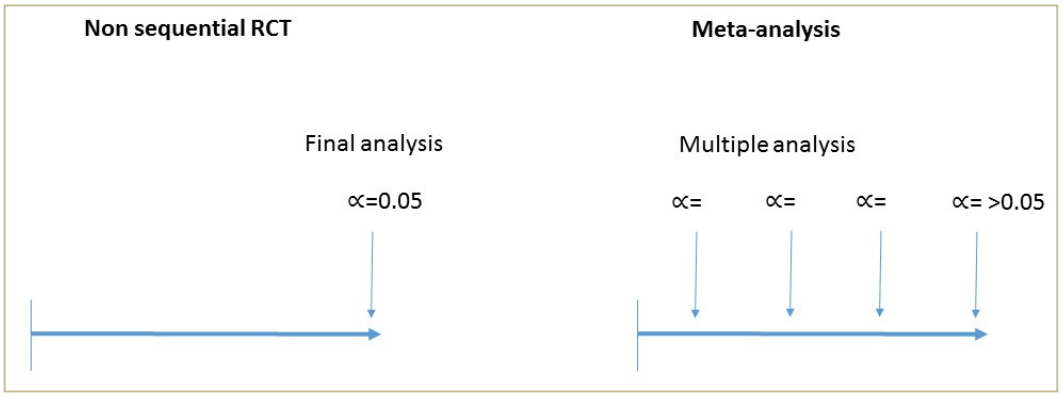


Figure 2.3. Significance testing for an RCT and for a meta-analysis.
In the non-sequential RCT there is a single calculation of alpha at the end of the study. In the meta-analysis the calculation of alpha is updated each time a new RCT is included. Because the alpha value is 'spent' in 4 analyses, the final alpha value (type I error) is higher than 0.05.

In 2011, Imberger et al. (47) noted that the issue of statistical multiplicity was largely being ignored in both Cochrane and non-Cochrane reviews, which would suggest that apparently conclusive meta-analysis results may be inconclusive (49). Table 2.1 summarizes the main causes of statistical multiplicity in systematic reviews.

Multiple tests of statistical significance may be performed because multiple outcomes of a meta-analysis are being evaluated or because a meta-analysis includes newly published RCTs. The higher probability of type I error (50) is analogous to the increased risk of error present when interim analyses are performed for individual trials.

Table 2.1. Some causes of statistical multiplicity in systematic reviews.

Cause	Example
Multiple outcomes	Mortality, cardiac event, cerebrovascular disease
Multiple time points	1 week, 3 months, 6 months, time to event
Multiple effect measures	Dichotomous, continuous
Subgroup analyses	High-risk patients, low-risk patients
Multiple looks at accumulating data	Updated data

Although research syntheses may combine findings from hundreds of studies, they are not immune to inflated type I errors when many statistical tests are conducted without adequate control over the error rate (50). Figure 2.4 shows, using a logarithmic representation, how the probability of type I error increases as the number of statistical tests increases.

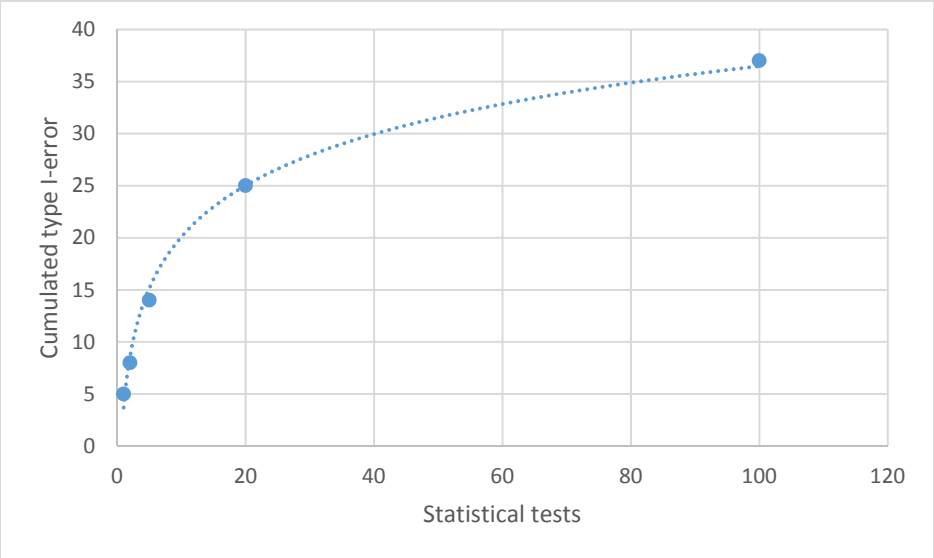


Figure 2.4. Type I error probabilities by the number of statistical tests.

Source: Adapted from Wetterslev 2017 (72).

Tendal et al. (51) found, for 18 of 19 studied meta-analyses, that there was a multiplicity of data in at least one trial report within each meta-analysis that frequently resulted in substantial variations in the pooled results.

As mentioned, the risk of erroneous results due to repeated analyses of accumulating data are well recognized for RCTs, for which reason, the CONSORT Statement recommends methods to minimize systematic error, one of which is the use of interim monitoring boundaries. However, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement (www.prisma-statement.org) does not include any recommendation about how to account for the increased risk of systematic error arising from multiple looks at data, and neither does the *Cochrane Handbook for Systematic Reviews of Interventions* (52) in its current version 5.1.0 which, however, is currently undergoing a major update, with version 6 planned for release sometime during 2018.

2.3.3. Reporting: quality and risk of bias

A number of tools exist to evaluate the risk of bias and the quality of systematic reviews with meta-analyses, of which the most important are PRISMA, AMSTAR, ROBIS and GRADE, briefly described in turn below.

The PRISMA Statement, published in 2009 to replace a previous tool called QUORUM, aims to improve the reporting and critical appraisal of systematic reviews and meta-analyses, although it is not recommended for the actual quality assessment of systematic reviews (53). However, in its 27-item checklist, even though items #15 and #16 address 'risk of bias' (publication bias, selective reporting, etc.) and 'additional analyses' (sensitivity or subgroup analyses, meta-regression), respectively, there is no mention – in any item – of the risk associated with type I and II errors in meta-analyses.

AMSTAR (A MeaSurement Tool to Assess Systematic Reviews) (54, 55), published in 2007, assesses the quality of systematic reviews via a checklist of 11 items. It is the most frequently mentioned tool for assessing systematic review quality (56) and has also recently been validated specifically to perform assessments of the methodological quality of systematic reviews in healthcare (57). Like

PRISMA, however, AMSTAR does not address how to assess the risk of bias related to type I and II errors in meta-analyses.

The ROBIS (Risk of Bias in Systematic Reviews) tool (58) published in 2016, is considered the first rigorously developed specific tool to assess the risk of bias in systematic reviews. Again, however, no item refers to the evaluation of type I and II error risk.

Finally, GRADE (Grading of Recommendations Assessment, Development and Evaluation) (59) is a rating system for the quality of evidence and strength of recommendations in guidelines that uses several indicators, among them, study limitations, imprecision, inconsistency of results, indirectness of evidence and publication bias (60). GRADE is the first quality rating tool that specifically proposes a method to evaluate imprecision, namely, optimal information size (OIS), stating as follows: "If the optimal information size criterion is not met, rate down for imprecision, unless the sample size is very large." (61).

2.3.4. Inclusion of truncated trials in meta-analyses

It has been suggested that the inclusion of tRCTs in meta-analyses could overestimate ES (62). Bassler et al. (63) conducted a systematic review investigating the impact of tRCTs on meta-analyses, commenting that, when tRCTs are considered, it should be borne in mind that although meta-analyses should serve to correct overestimates from individual tRCTs, pooled results from currently available trials may tend to overestimation. The risk of overestimation will be particularly high when the following 3 conditions co-occur:

1) The tRCTs have a relatively small number of events (e.g. <200).
2) There is a substantial difference in relative risk (RR) between the tRCTs and the non-tRCTs (e.g. RR<0.7).
3) The tRCTs account for a substantial proportion (>20%) of studies included in the meta-analysis.

When all 3 conditions co-occur, Bassler et al. (63) recommend the use of sensitivity analyses that include and that exclude tRCTs. However, a subsequent simulation analysis published in 2016 (64) recommended not excluding tRCTs from meta-analyses evaluating treatment effects, concluding that: “The pooled effects from meta-analyses that include truncated RCTs do not show a problem of upward bias even if the truncated RCT freezes future experiments.”

2.3.5. Big data and real-world evidence

An important consideration with RCTs is any potential for significant savings in accrual time and financial costs (65). Obtaining data has, to date, been complex from the perspective of including patients, whether because of restrictive inclusion criteria, extended inclusion periods, ethical dilemmas, high costs or little external validity of the data (66).

However, new technologies are allowing access to large volumes of real-life data, which, provided they are analysed correctly, constitute potentially reliable real-world evidence (67). A big data scenario — broadly defined as “a new generation of technologies and architectures, designed to extract value from large volumes of a wide variety of data by enabling high-velocity capture, discovery and analysis” (68) — is coming to the fore in biomedicine that offers potential for complementing data from clinical trials. A related concept is real-world evidence, which refers to information on healthcare derived from various sources other than traditional clinical research, whether electronic health records, billing data, disease registries or even data gathered through personal devices and health applications (69).

The fact that access to real-world evidence is being made possible by current and potential sources of information and new statistical approaches to data analysis (70) would suggest that precise calculations of sample size may not be so important in the future (71). However, it needs to be borne

in mind that vast datasets would also contain a great deal of noise, and this logically, would reduce the validity of the data and any conclusions based on them.

2.4. Optimal information size and sequential design

2.4.1. Optimal information size

As pointed out by Wetterslev et al. (72), "most meta-analyses in systematic reviews, including Cochrane ones, do not have sufficient statistical power to detect or refute even large intervention effects. This is why a meta-analysis ought to be regarded as an interim analysis on its way towards a required information size." Pogue and Yusuf (73, 74), in their 1998 proposal to apply interim analysis techniques to meta-analyses, put forward the concept of OIS, to be estimated in a similar way to sample size for comparative RCTs. They defined OIS — using the word 'optimum' rather than 'optimal' — as follows: "We propose the concept of optimum information size (OIS) as the minimum amount of information required in the collective literature for reliable conclusions about an intervention to be reached" (73). Since then, several meta-analyses have been published that use this methodology or a modified version of it (see, for instance, 75, 76, 78 and 79).

2.4.2. Sequential design

The classic frequentist approach to an RCT — called non-sequential design — is to determine the number of participants in a study beforehand (fixed sample size). An alternative approach is sequential design, in which the total number of participants is not known a priori but is determined as the study progresses. In sequential design, statistical analyses are performed periodically and new observations are included or the study is terminated depending on the significance of results at each analysis point. Sequential design has ethical and financial advantages over non-sequential design, as it rationalizes sample size and study duration and is, therefore, potentially more cost-effective.

Although sequential methods were first used for statistical approaches to quality control in manufacturing processes from around the 1920s, modern sequential analysis theory is attributed to Abraham Wald, who, in a book published in 1947, defined sequential analysis as follows (80):

> Sequential analysis is a method of statistical inference whose characteristic feature is that the number of observations required by the procedure is not determined in advance of the experiment. The decision to terminate the experiment depends, at each stage, on the results of the observations previously made. A merit of the sequential method, as applied to testing statistical hypotheses, is that test procedures can be constructed which require, on the average, a substantially smaller number of observations than equally reliable test procedures based on a pre-determined number of observations (p.1).

The innovative statistical inference method developed by Wald — initially restricted to wartime research and production but released to the public in 1945 (81) — was the sequential probability ratio test. This test proved very useful for hypothesis testing, given that it aimed to mitigate the increased risk of type I error resulting from repeated testing of a growing number of observations. Since it was only suitable for continuous evaluation as new outcomes resulted (i.e. sample size was not bounded), this test was eventually specifically adapted for use in RCTs, as in its original format it would have required results to be updated every time a new patient was included (80).

Sequential methods, based on what is called adaptive sampling, have been used in RCTs for over 50 years (82). A pioneer in the field, particularly for comparative studies, was Peter Armitage (83) who, in 1958, wrote that “the number of published reports of clinical trials involving sequential analysis could be counted on the fingers of both hands.” In 1975, Armitage published *Sequential Medical Trials* (83), underpinned by the core notion that if, for a comparative study with 2 arms and a two-sided significance test, a significant alpha level was obtained after each observation or pair of

observations, then the trial could be stopped and the difference between the 2 arms reported. Since then, numerous methodological articles on sequential design and sequentially designed RCTs have been published.

Since repeated statistical testing increases the risk of type I error, adjustments must be made to ensure that the probability is maintained at a pre-specified level (84). The statistical methods used in sequential methods therefore must ensure that interim analyses can be performed without increasing type I error.

2.4.3. Group sequential analysis

The problem with updating data each time a new patient was included in a study led Pocock (85) to propose a method called group sequential analysis, based on recruiting patients into equal-sized groups so that the decision to stop or continue the trial would be based on the results of significance testing of accumulated data after groups of the same size were evaluated. This approach requires pre-determination of a maximum number (K) of interim analyses, in such a way that each analysis will be performed as soon as a pre-established quantity of data becomes available. For each interim analysis a significance level (α') of less than 0.05 (or a specific value, e.g. 0.01) is set. Thus, if no significant differences are found in the first interim analysis, the study continues until another group of patients of the same size is accrued and the analysis is repeated. If no significant differences are found after completing the pre-determined number of interim analyses, then the study ends with a negative result (because the alternative hypothesis (H1) is rejected). This approach provided a great impetus to sequential methods, mainly because it provided specific guidelines that met requirements regarding type I error and statistical power.

An improvement on the Pocock (85) approach was developed by O'Brien and Fleming (86), who proposed the alpha spending function for interim analyses. This spending function was based on a

decreasing sequence of alpha values, so that much stronger evidence of an effect would be required to stop a trial early. According to this statistical technique, which results in a monotonic curve, data are reviewed periodically and the study is terminated early once a treatment proves superior. The types of boundaries produced by this alpha spending function were first proposed for equal increments of the information fraction (86,87). However, a major drawback with this method when applied to meta-analyses is that the number of tests to be performed and the number of patients to be evaluated between any 2 analyses has to be decided at the outset (85).

To overcome these problems, Lan and DeMets (88) described a procedure to compute a flexible discrete boundary for each interim analysis that would allow varying numbers of patients, i.e. unequal or even unpredictable group sizes. In this approach, an increasing alpha spending function is specified that characterizes the rate at which the total alpha is spent. Boundary values are the same for all the interim analyses according to the Pocock (85) method, whereas they are different for the O'Brien-Fleming method (86).

Figure 2.5 illustrates the differences between the Pocock (85) and O'Brien-Fleming (86) methods.

The methods developed by Pocock (85), O'Brien and Fleming (86) and Lan and DeMets (88) have been mainly described and used under the frequentist paradigm, but can also be applied, with some adjustments, under a Bayesian paradigm (89).

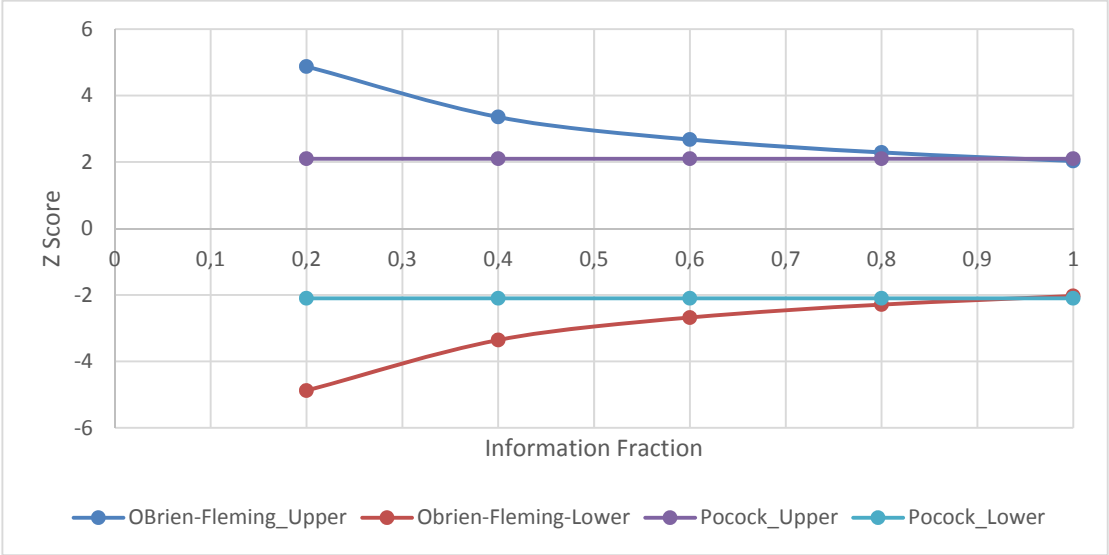


Figure 2.5. Comparison of the Pocock and O'Brien-Fleming interim analysis methods. Monitoring boundaries are simulated with Fortran software for the following scenario: interim analyses K = 5 and monitoring boundaries = two-sided symmetric. Note how boundary values for all interim analyses are the same for the Pocock method but different for the O'Brien-Fleming method.

2.4.4. Sequential methods in meta-analyses

Although the application of sequential methods is well established for individual RCTs, this is much less the case for meta-analyses. In cumulative meta-analyses that include new trials as they appear, therefore, the fact that the effect estimate is updated potentially increases the risk of type I error.

Several techniques for applying sequential methods to meta-analyses include frequentist methods, such as the alpha spending approach (73), trial sequential analysis (72) and Whitehead's (89) triangular test, not to mention other methods, such as the law of the iterated logarithm, stochastic curtailment, the semi-Bayes procedure and Bayesian methods (90).

Lan and DeMets's (88) alpha spending function can be readily adapted to yield approximate monitoring boundaries for cumulative meta-analyses. The fact that the rate at which alpha is spent by past decisions while remaining independent of future decisions makes it possible to adapt this method to cumulative meta-analyses, given that the number of new RCTs that may be included in the future cannot be predicted.

As mentioned earlier, Pogue and Yusuf (73,74) proposed using OIS to calculate the number of patients required for an adequately powered meta-analysis. To obtain the OIS they calculated the alpha value, power, control event rate (CER, i.e. the proportion of control group patients with the outcome of interest) and ES (i.e. the quantified size of the difference between treatment groups) — but not heterogeneity (i.e. the variation in outcomes between studies). They then calculated the two-sided Lan-DeMets monitoring boundaries to obtain the cumulative test statistic Z. If this statistic crossed the upper or lower boundary — as illustrated in Figure 2.6 — then an effect could be concluded in favour of or against the intervention, respectively (74).

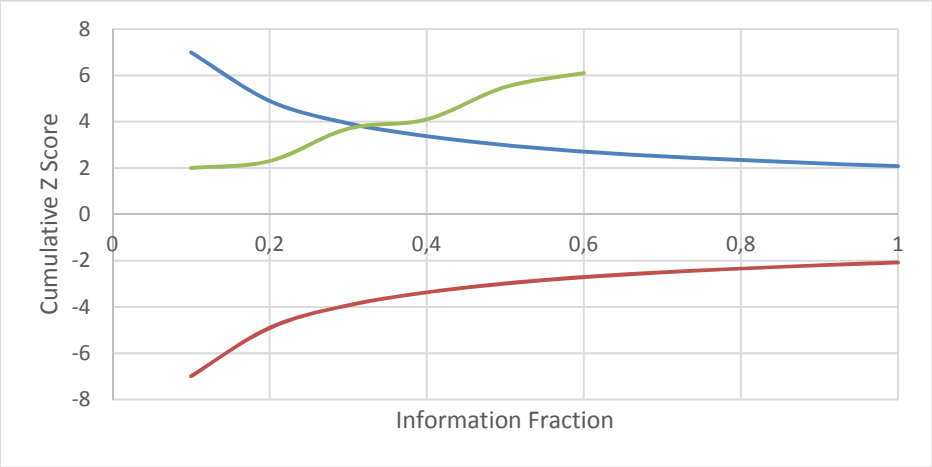


Figure 2.6. O'Brien-Fleming alpha spending function.
Two-sided symmetric monitoring boundaries, with interim analyses K=10 and an overall significance level of 0.05. The cumulative Z score (green line) — calculated after each interim analysis of each RCT — crossing the upper monitoring boundary (blue) indicates an effect in favour of the intervention.

Whitehead (91), who described a sequential design using the triangular test method, performed multiple simulation analyses using fixed-effects and random-effects models, demonstrating that the latter will often require more patients than the former, and most especially if the random component is large relative to the treatment difference. However, concluding that heterogeneity estimation using this method may not be ideal, Whitehead (91) proposed using more sophisticated methods, such as the Bayesian approach and a prior distribution.

Wetterslev et al. (92), drawing on previous work by Pogue and Yusuf (73,74), used an approach called trial sequential analysis, in which the required number of participants is calculated on the basis of pre-specified type I and II errors, power, ES and CER; heterogeneity is also included in order to adapt this approach to random-effects meta-analyses and Lan-DeMets (88) boundaries are used, also adapted for use in random-effects meta-analyses. Wetterslev et al. (93) also developed a diversity measure called D^2, intended to be comparatively more conservative and less biased than the classic inconsistency measure (I^2) of heterogeneity. Note in passing, however, that Kulinskaya (94) considered both these measures of heterogeneity to be faulty, suggesting that estimates of heterogeneity in a random-effects model cannot be considered reliable when the number of trials is small. In the trial sequential analysis approach, the cumulative Z-curve is constructed with each cumulative Z-value calculated after including a new RCT and the estimated sample size is achieved once the Z-curve crosses the upper monitoring boundary (95), as illustrated in Figure 2.7.

In reality, it can be difficult to know which assumptions are the most appropriate when making OIS estimates for meta-analyses (96). One of the main criticisms of OIS estimation in a cumulative meta-analysis is that it is a 'moving target', in that the inclusion of a new trial changes estimates of the key CER and heterogeneity parameters (72).

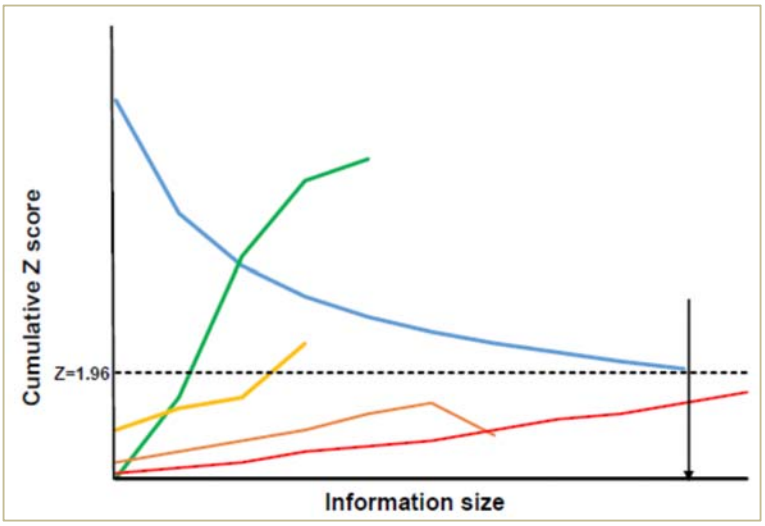


Figure 2.7. Upper half of a two-sided trial sequential analysis.
The blue curve is the upper monitoring boundary and the green, yellow, orange and red curves are cumulative Z-curves from 4 different meta-analyses. The black arrow on the right indicates the OIS point. The yellow curve crossing Z=1.96 indicates a 'traditionally' significant result. The green curve crossing the monitoring boundary before reaching the OIS indicates reliable evidence adjusted for random error. The orange curve not crossing Z=1.96 indicates the absence of evidence if the OIS is not achieved. The red curve not crossing Z=1.96 indicates no pre-defined intervention effect if the OIS is achieved. The monitoring boundary moves upwards and to the right as CER decreases, or ES decreases, or heterogeneity increases.
Source: Adapted from Brok et al., 2008 (41).

In a 2010 study that compared Wetterslev's trial sequential analysis approach with Whitehead's boundaries approach, Van der Tweel et al. (97) found that the latter generally requires less information than the former and also noted that: "(a) the power can be quantified; (b) stopping for futility is an option; (c) no (arbitrary) estimate of total information size is necessary beforehand; (d) point and interval estimates are adjusted for the multiple testing; and (e) gains in efficiency can be achieved, both for efficacy and for futility, and thus ethical and economic benefits can be obtained."

A noteworthy problem with trial sequential analysis is that the heterogeneity value (in this case, D^2) cannot be considered fully reliable. To estimate heterogeneity more effectively in sequential random-effects meta-analyses, Higgins et al. (90) adopted a semi-Bayesian approach and used Whitehead's monitoring boundaries. In a simulation analysis, these authors compared 5 different methods for estimating heterogeneity, demonstrating that the most balanced and straightforward

approach is to use a prior information distribution of heterogeneity with a semi-Bayesian approach, as it offers reasonable false positive and coverage properties. These authors also called for further research to characterize the degree of heterogeneity that can be anticipated in meta-analyses with particular clinical and methodological features (81).

Most sequential methods require a lower p-value for earlier analyses than for later analyses. Shuster et al. (98) defended — as being relatively more aggressive towards early stopping — using Pocock's (85) approach for sequential meta-analyses, given that the p-value required for stoppage is the same for each look if the sequential nature of the looks is ignored. The same authors also argued against methods like those of Wetterslev (95) and Higgins et al. (90), because they violated the independent increment premise in that a decrease in total information may be observed even as trials are added. They concluded in favour of the Pocock (85) approach — with carefully planned increments of studies — as a robust method for conducting a prospective group sequential meta-analysis of a series of RCTs (98).

Contrasting with all previous such methods based on frequentist and semi-Bayesian approaches, Spence et al. (99) — demonstrating the potential of their approach by conducting a simulation study — used a fully Bayesian sequential meta-analysis with a prior information distribution of heterogeneity and with ES parameters and stopping decisions or recommendations based on posterior probabilities calculated directly from the ES posterior distribution.

Frequentist methods mainly have limitations regarding the estimation of heterogeneity, especially at the outset of the meta-analysis. However, the drawback with fully Bayesian approaches is the lack of direct control over type I and II errors that is a feature of traditional frequentist methods. Nonetheless, Bayesian treatment of heterogeneity and ES remains a good alternative to balancing conclusions regarding effect, futility and the need for further RCTs.

2.5. Software

Many different methods for performing sequential meta-analyses and calculating information size are available and, likewise, several different softwares are available that range from specifically designed software to standard online sample size calculators (some with no option for including heterogeneity as a parameter).

TSA (Trial Sequential Analysis) is a specific software created by the Copenhagen Trial Unit (CTU). Applying the O'Brien-Fleming (95) method, it calculates the required information size for a meta-analysis based on either a random-effects or fixed-effects model. It allows for inclusion of the heterogeneity parameter as either I^2 or D^2 (93). It also calculates boundaries and produces graphical representations as output (Figure 2.8).

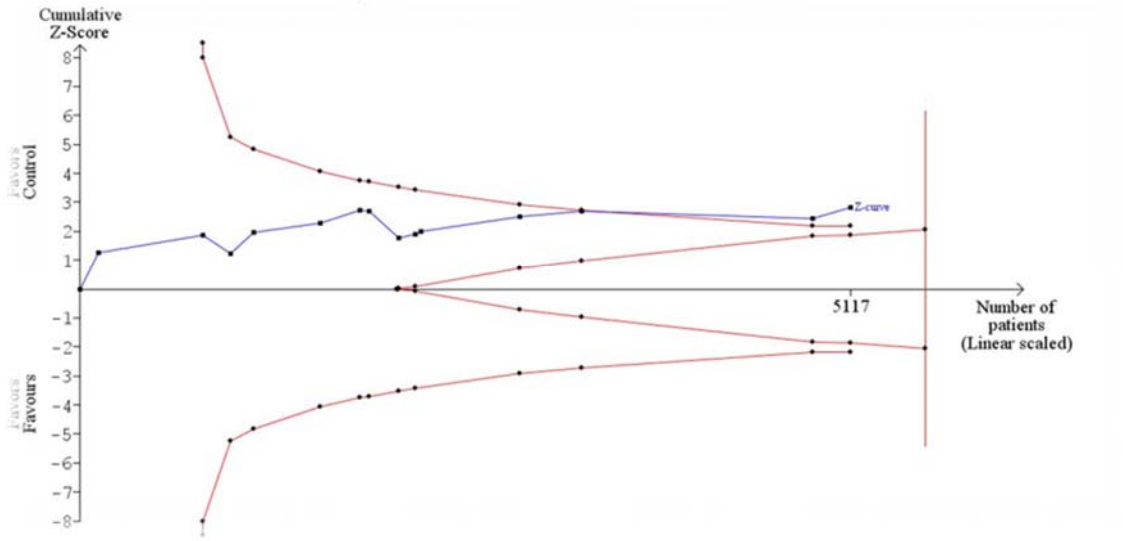


Figure 2.8. TSA software output.
The cumulative Z-curve (indicated in blue) crossing the upper boundary (red) indicates that the meta-analysis results are conclusive. The boundary below the upper boundary (also red) indicates the futility area. The blue line crossing this second boundary indicates that there is sufficient information to confirm that the intervention is not effective. The lower part of the graph represents the control group. The blue line crossing the boundary (red line) in the futility area indicates that there is sufficient information to confirm that the control group intervention is not effective. The blue line crossing the second boundary (red line) in this area indicates that there is sufficient information to confirm that the control group intervention is effective.

For STATA (www.stata.com), a standard data analysis and statistical software, Miladinovic et al. (100) described a new command, metacumbounds, to estimate trial sequential monitoring boundaries in cumulative meta-analyses. Their approach is based on the Lan-DeMets method for estimating sequential boundaries for individual RCTs and the ldbounds package applied in R statistical software. Using the STATA software and the metan command, metacumbounds plots the Lan-DeMets bounds, Z-values and p-values obtained from both fixed-effects and random-effects cumulative meta-analyses.

PEST (Planning and Evaluation of Sequential Trials; www.mps-research.com/PEST), developed by John Whitehead, simulates sequential RCTs to illustrate how a trial might proceed under various user-specified conditions. It calculates stopping rules and performs a series of interim analyses, at each of which the trial might be stopped and a conclusion drawn. However, although still available for use, updates to this software have been stopped due to a lack of finance for the project.

R (from the R Project for Statistical Computing; www.r-project.org) is an open-source software environment and programming language for statistical computing and graphics that has been used for semi-Bayesian and fully Bayesian sequential meta-analyses. Higgins et al. (90) wrote the function to perform analyses using the semi-Bayesian approach and Spence et al. (99) wrote the function to perform fully Bayesian analyses.

2.6. Literature gaps and hypothesis generation

Systematic reviews with meta-analyses are crucial to decision making in healthcare and to suggest directions for future research. Nonetheless, the literature review highlights the fact that meta-analysis results are often inconclusive or even questionable due to problems of power and related statistical issues. Although OIS estimation is a possible approach to improving the decision-making

value of meta-analyses, there is a substantial evidence gap regarding the most appropriate approach to applying this methodology.

Specific gaps in the current literature addressed in this thesis are listed as follows:

- There is no agreed, standardized method for calculating sample size for meta-analyses.
- There is no agreed, standardized method for calculating the statistical parameters with dichotomous outcomes necessary to estimate OIS, namely, CER, ES and heterogeneity.
- The approach to OIS estimation is not well developed for continuous outcomes.
- Few existing studies evaluate published meta-analyses that include OIS estimates in terms of their characteristics and methods.
- Little is known regarding opinions and perceptions of OIS estimates by systematic review and meta-analysis guideline developers.

The above gaps in knowledge of OIS are addressed in this thesis as follows:

1. The lack of knowledge regarding OIS use is addressed by a literature search for Cochrane and non-Cochrane systematic reviews with meta-analyses (Chapter 3).
2. The lack of knowledge regarding opinions and perceptions of OIS estimates is addressed by means of surveys of developers of guidelines for systematic reviews with meta-analyses and of systematic review authors (Chapters 3 and 6).
3. The lack of information regarding the statistical parameters and values used to calculate OIS is addressed by analysing the impact of the 3 main parameters — CER, ES and heterogeneity — necessary to calculate OIS in published systematic reviews with meta-analyses (Chapters 4 and 5).

2.7. References

1. Medical Subject Heading-Mesh Medline. National Institutes of Health (NIH). National Library of Medicine, National Center for Biotechnology Information (NCBI). PubMed. Available from: www.pubmed.com (accessed 11 October 2017).
2. Sibbald B, Roland M. Understanding controlled trials. Why are randomised controlled trials important? BMJ. 1998; 316(7126):201.
3. Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ. 2010; 340:c332. doi: 10.1136/bmj.c332.
4. Chan AW, Altman DG. Epidemiology and reporting of randomised trials published in PubMed journals. Lancet. 2005; 365(9465):1159-62.
5. Freiman JA, Chalmers TC, Smith H Jr, Kuebler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 "negative" trials. N Engl J Med. 1978; 299(13):690-4.
6. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. JAMA. 1994; 272(2):122-4.
7. Schuit E, Roes KC, Mol BW, Kwee A, Moons KG, Groenwold RH. Meta-analyses triggered by previous (false-)significant findings: problems and solutions. Syst Rev. 2015; 4:57. doi: 10.1186/s13643-015-0048-9.
8. McKillup S. Statistics explained. An introductory guide for life scientists. 2nd Edition. United Kingdom: Cambridge University Press; 2012.

9. Kirkham JJ, Clarke M, Williamson PR. A methodological approach for assessing the uptake of core outcome sets using ClinicalTrials.gov: findings from a review of randomised controlled trials of rheumatoid arthritis. BMJ. 2017; 357:j2262. doi: 10.1136/bmj.j2262.
10. Gargon E, Williamson PR, Altman DG, Blazeby JM, Tunis S, Clarke M. The COMET Initiative database: progress and activities update (2015). Trials. 2017 Feb 3;18(1):54. doi: 10.1186/s13063-017-1788-8.
11. Williamson PR, Altman DG, Bagley H, Barnes KL, Blazeby JM, Brookes ST, Clarke M, Gargon E, Gorst S, Harman N, Kirkham JJ, McNair A, Prinsen CAC, Schmitt J, Terwee CB, Young B. The COMET Handbook: version 1.0. Trials. 2017; 18(Suppl 3):280. doi: 10.1186/s13063-017-1978-4.
12. Whitley E, Ball J. Statistics review 4: sample size calculations. Crit Care. 2002; 6(4):335-41.
13. Pocock SJ, Stone GW. The Primary Outcome Is Positive - Is That Good Enough? N Engl J Med. 2016; 375(10):971-9.
14. Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. J Clin Epidemiol. 2000; 53(11):1119-29.
15. Sackett DL. Why randomized controlled trials fail but needn't: 2. Failure to employ physiological statistics, or the only formula a clinician-trialist is ever likely to need (or understand!). CMAJ. 2001; 165(9):1226-37.
16. Krzywinski M, Altman N. Power and sample size. Nature Methods. 2013; 10(12):1139-1140.
17. Steve McKillup. Statistics Explained. An introductory guide for life scientists. Cambridge University Press. 2nd edition. 2012
18. Bacchetti P. Current sample size conventions: flaws, harms, and alternatives. BMC Med. 2010; 8:17. doi: 10.1186/1741-7015-8-17.
19. Stokes L. Sample size calculation for a hypothesis test. JAMA. 2014; 312(2):180-1.

20. Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, Altman DG. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol. 2016; 31(4):337-50.

21. Noordzij M, Dekker FW, Zoccali C, Jager KJ. Sample size calculations. Nephron Clin Pract. 2011;118(4):c319-23.

22. Chan AW, Hróbjartsson A, Jørgensen KJ, Gøtzsche PC, Altman DG. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. BMJ. 2008; 337:a2299. doi: 10.1136/bmj.a2299.

23. Sterne JA, Gavaghan D, Egger M. Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature. J Clin Epidemiol. 2000; 53(11):1119-29.

24. Charles P, Giraudeau B, Dechartres A, Baron G, Ravaud P. Reporting of sample size calculation in randomised controlled trials: review. BMJ. 2009; 338:b1732. doi: 10.1136/bmj.b1732.

25. Todd S. A 25-year review of sequential methodology in clinical studies. Stat Med. 2007; 26(2):237-52.

26. Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Eggert CH, Briel M, Lacchetti C, Leung TW, Darling E, Bryant DM, Bucher HC, Schünemann HJ, Meade MO, Cook DJ, Erwin PJ, Sood A, Sood R, Lo B, Thompson CA, Zhou Q, Mills E, Guyatt GH. Randomized trials stopped early for benefit: a systematic review. JAMA. 2005; 294(17):2203-9.

27. Bassler D, Briel M, Montori VM, Lane M, Glasziou P, Zhou Q, Heels-Ansdell D, Walter SD, Guyatt GH; STOPIT-2 Study Group, Flynn DN, Elamin MB, Murad MH, Abu Elnour NO, Lampropulos JF, Sood A, Mullan RJ, Erwin PJ, Bankhead CR, Perera R, Ruiz Culebro C, You JJ, Mulla SM, Kaur J, Nerenberg KA, Schünemann H, Cook DJ, Lutz K, Ribic CM, Vale N, Malaga G, Akl EA, Ferreira-Gonzalez I, Alonso-Coello P, Urrutia G, Kunz R, Bucher HC, Nordmann AJ, Raatz H, da Silva SA, Tuche F, Strahm B, Djulbegovic B, Adhikari NK, Mills EJ, Gwadry-Sridhar F, Kirpalani H, Soares

HP, Karanicolas PJ, Burns KE, Vandvik PO, Coto-Yglesias F, Chrispim PP, Ramsay T. Stopping randomized trials early for benefit and estimation of treatment effects: systematic review and meta-regression analysis. JAMA. 2010; 303(12):1180-7. doi: 10.1001/jama.2010.310.

28. Lachin JM. The role of measurement reliability in clinical trials. Clin Trials. 2004;1(6):553-66.

29. Glossary of statistical terms. Available from: http://stats.oecd.org/glossary/ (accessed 3 December 2017).

30. Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011.

31. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. Open Med. 2009;3(3):e123-30.

32. Glass GV. Meta-analysis at middle age: a personal history. Res Synth Methods. 2015; 6(3):221-31. doi: 10.1002/jrsm.1133.

33. Clarke M, Williamson PR. Core outcome sets and systematic reviews. Syst Rev. 2016; 5:11. doi: 10.1186/s13643-016-0188-6.

34. Lau J, Antman EM, Jimenez-Silva J, Kupelnick B, Mosteller F, Chalmers TC. Cumulative meta-analysis of therapeutic trials for myocardial infarction. N Engl J Med. 1992; 327(4):248-54.

35. Clarke M, Brice A, Chalmers I. Accumulating research: a systematic account of how cumulative meta-analyses would have provided knowledge, improved health, reduced harm and saved resources. PLoS One. 2014;9(7):e102670.

36. Imberger G, Thorlund K, Gluud C, Wetterslev J. False-positive findings in Cochrane meta-analyses with and without application of trial sequential analysis: an empirical review. BMJ Open. 2016; 6(8):e011890.

37. Schuit E, Roes KC, Mol BW, Kwee A, Moons KG, Groenwold RH. Meta-analyses triggered by previous (false-)significant findings: problems and solutions. Syst Rev. 2015; 4:57. doi: 10.1186/s13643-015-0048-9.

38. Imberger G, Gluud C, Boylan J, Wetterslev J. Systematic Reviews of Anesthesiologic Interventions Reported as Statistically Significant: Problems with Power, Precision, and Type 1 Error Protection. Anesth Analg. 2015; 121(6):1611-22.

39. Turner RM, Bird SM, Higgins JP. The impact of study size on meta-analyses: examination of underpowered studies in Cochrane reviews. PLoS One. 2013; 8(3):e59202. doi: 10.1371/journal.pone.0059202.

40. Thorlund K, Imberger G, Walsh M, Chu R, Gluud C, Wetterslev J, Guyatt G, Devereaux PJ, Thabane L. The number of patients and events required to limit the risk of overestimation of intervention effects in meta-analysis--a simulation study. PLoS One. 2011;6(10):e25491.

41. Brok J, Thorlund K, Gluud C, Wetterslev J. Trial sequential analysis reveals insufficient information size and potentially false positive results in many meta-analyses. J Clin Epidemiol. 2008; 61(8):763-9. doi: 10.1016/j.jclinepi.2007.10.007.

42. AlBalawi Z, McAlister FA, Thorlund K, Wong M, Wetterslev J. Random error in cardiovascular meta-analyses: how common are false positive and false negative results? Int J Cardiol. 2013; 168(2):1102-7. doi: 10.1016/j.ijcard.2012.11.048.

43. Dechartres A, Trinquart L, Boutron I, Ravaud P. Influence of trial sample size on treatment effect estimates: meta-epidemiological study. BMJ. 2013; 346:f2304. doi: 10.1136/bmj.f2304.

44. Bender R, Bunce C, Clarke M, Gates S, Lange S, Pace NL, Thorlund K. Attention should be given to multiplicity issues in systematic reviews. J Clin Epidemiol. 2008; 61(9):857-65.

45. Schulz KF, Grimes DA. Multiplicity in randomised trials II: subgroup and interim analyses. Lancet. 2005; 365(9471):1657-61.

46. Ranganathan P, Pramesh CS, Buyse M. Common pitfalls in statistical analysis: The perils of multiple testing. Perspect Clin Res. 2016; 7(2):106-7.

47. Imberger G, Vejlby AD, Hansen SB, Møller AM, Wetterslev J. Statistical multiplicity in systematic reviews of anaesthesia interventions: a quantification and comparison between Cochrane and non-Cochrane reviews. PLoS One. 2011;6(12):e28422.

48. Benda N, Brandt A. Regulatory issues with multiplicity in drug approval: Principles and controversies in a changing landscape. J Biopharm Stat. 2018;28(1):3-9.

49. Brok J, Thorlund K, Wetterslev J, Gluud C. Apparently conclusive meta-analyses may be inconclusive--Trial sequential analysis adjustment of random error risk due to repetitive testing of accumulating data in apparently conclusive neonatal meta-analyses. Int J Epidemiol. 2009; 38(1):287-98.

50. Polanin JR, Pigott TD. The use of meta-analytic statistical significance testing. Res Synth Methods. 2015; 6(1):63-73.

51. Tendal B, Nüesch E, Higgins JP, Jüni P, Gøtzsche PC. Multiplicity of data in trial reports and the reliability of meta-analyses: empirical study. BMJ. 2011; 343:d4829.

52. Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.handbook.cochrane.org.

53. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol. 2009; 62(10):1006-12.

54. Shea BJ, Bouter LM, Peterson J, Boers M, Andersson N, Ortiz Z, Ramsay T, Bai A, Shukla VK, Grimshaw JM. External validation of a measurement tool to assess systematic reviews (AMSTAR). PLoS One. 2007; 2(12):e1350.

55. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, Porter AC, Tugwell P, Moher D, Bouter LM. Development of AMSTAR: a measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol. 2007; 7:10.

56. Pollock M, Fernandes RM, Becker LA, Featherstone R, Hartling L. What guidance is available for researchers conducting overviews of reviews of healthcare interventions? A scoping review and qualitative metasummary. Syst Rev. 2016; 5(1):190.

57. Pollock M, Fernandes RM, Hartling L. Evaluation of AMSTAR to assess the methodological quality of systematic reviews in overviews of reviews of healthcare interventions. BMC Med Res Methodol. 2017; 17(1):48.

58. Whiting P, Savović J, Higgins JP, Caldwell DM, Reeves BC, Shea B, Davies P, Kleijnen J, Churchill R; ROBIS group. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. J Clin Epidemiol. 2016; 69:225-34.

59. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, DeBeer H, Jaeschke R, Rind D, Meerpohl J, Dahm P, Schünemann HJ. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011;64(4):383-94.

60. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, Norris S, Falck-Ytter Y, Glasziou P, DeBeer H, Jaeschke R, Rind D, Meerpohl J, Dahm P, Schünemann HJ. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol. 2011; 64(4):383-94.

61. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, Devereaux PJ, Montori VM, Freyschuss B, Vist G, Jaeschke R, Williams JW Jr, Murad MH, Sinclair D, Falck-Ytter Y, Meerpohl J, Whittington C, Thorlund K, Andrews J, Schünemann HJ. GRADE guidelines 6. Rating the quality of evidence--imprecision. J Clin Epidemiol. 2011; 64(12):1283-93.

62. Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Eggert CH, Briel M, Lacchetti C, Leung TW, Darling E, Bryant DM, Bucher HC, Schünemann HJ, Meade MO, Cook DJ, Erwin PJ, Sood A, Sood

R, Lo B, Thompson CA, Zhou Q, Mills E, Guyatt GH. Randomized trials stopped early for benefit: a systematic review. JAMA. 2005; 294(17):2203-9.

63. Bassler D, Montori VM, Briel M, Glasziou P, Walter SD, Ramsay T, Guyatt G. Reflections on meta-analyses involving trials stopped early for benefit: is there a problem and if so, what is it? Stat Methods Med Res. 2013; 22(2):159-68.
64. Wang H, Rosner GL, Goodman SN. Quantifying over-estimation in early stopped clinical trials and the "freezing effect" on subsequent research. Clin Trials. 2016;13(6):621-631.
65. Jitlal M, Khan I, Lee SM, Hackshaw A. Stopping clinical trials early for futility: retrospective analysis of several randomised clinical studies. Br J Cancer. 2012; 107(6):910-7.
66. Brunoni AR, Tadini L, Fregni F. Changes in Clinical Trials Methodology Over Time: A Systematic Review of Six Decades of Research in Psychopharmacology. Scherer RW, ed. PLoS ONE. 2010;5(3):e9479.
67. Goodman SN, Schneeweiss S, Baiocchi M. Using Design Thinking to Differentiate Useful From Misleading Evidence in Observational Research. JAMA. 2017; 317(7):705-707.
68. Lipworth W, Mason PH, Kerridge I, Ioannidis JP. Ethics and Epistemology in Big Data Research. J Bioeth Inq. 2017 Mar 20. doi: 10.1007/s11673-017-9771-3.
69. Sherman RE, Anderson SA, Dal Pan GJ, Gray GW, Gross T, Hunter NL, LaVange L, Marinac Dabic D, Marks PW, Robb MA, Shuren J, Temple R, Woodcock J, Yue LQ, Califf RM. Real-World Evidence - What Is It and What Can It Tell Us? N Engl J Med. 2016; 375(23):2293-2297.
70. Agoritsas T, Merglen A, Shah ND, O'Donnell M, Guyatt GH. Adjusted Analyses in Studies Addressing Therapy and Harm: Users' Guides to the Medical Literature. JAMA. 2017; 317(7):748-759.
71. Hudis CA. Big data: Are large prospective randomized trials obsolete in the future? Breast. 2015; 24 Suppl 2:S15-8. doi: 10.1016/j.breast.2015.07.005.

72. Wetterslev J, Jakobsen JC, Gluud C. Trial Sequential Analysis in systematic reviews with meta-analysis. BMC Med Res Methodol. 2017; 17: 39.

73. Pogue JM, Yusuf S. Cumulating evidence from randomized trials: utilizing sequential monitoring boundaries for cumulative meta-analysis. Controlled clinical trials 1997; 18: 580-593.

74. Pogue JM, Yusuf S. Overcoming the limitations of current meta-analysis of randomized controlled trials. Lancet 1998; 351: 47-52.

75. Guo X, Guo S, Miao Z, Li Z, Zhang H. Myo-inositol lowers the risk of developing gestational diabetic mellitus in pregnancies: A systematic review and meta-analysis of randomized controlled trials with trial sequential analysis. J Diabetes Complications. 2018; 32(3):342-348.

76. Hristovska AM, Duch P, Allingstrup M, Afshari A. The comparative efficacy and safety of sugammadex and neostigmine in reversing neuromuscular blockade in adults. A Cochrane systematic review with meta-analysis and trial sequential analysis. Anaesthesia. 2017 Dec 27.

77. Gu SC, Wang CD. Early Selective Serotonin Reuptake Inhibitors for Recovery after Stroke: A Meta-Analysis and Trial Sequential Analysis. J Stroke Cerebrovasc Dis. 2018;27(5):1178-1189.

78. Zhang ZQ, Zhong Y, Huang XM, Du LZ. Airway administration of corticosteroids for prevention of bronchopulmonary dysplasia in premature infants: a meta-analysis with trial sequential analysis. BMC Pulm Med. 2017; 17(1):207.

79. Wang B, Chen R, Guo X, Zhang W, Hu J, Gong Y, Cheng B. Effects of levosimendan on mortality in patients with septic shock: systematic review with meta-analysis and trial sequential analysis. Oncotarget. 2017 Aug 10;8(59):100524-100532.

80. Wald A. *Sequential Analysis*. New York: Dover Publications, Inc.; 1947.

81. Mukhopadhyay N. Celebrating Abraham Wald's Birth Centenary. Sequential Analysis 2004; 23 (1): 1-9

82. Todd S. A 25-year review of sequential methodology in clinical studies. Stat Med. 2007; 26(2):237-52.

83. Armitage P. Sequential medical trials. Oxford: Blackwell Scientific Publications; 1975.

84. Strube MJ. SNOOP: a program for demonstrating the consequences of premature and repeated null hypothesis testing. Behav Res Methods. 2006; 38(1):24-7.

85. Pocock SJ. Group sequential methods in the design and analysis of clinical trials. Biometrika 1977; 64(2): 191-199.

86. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics. 1979; 35(3):549-56.

87. Kim K, DeMets DL. Confidence intervals following group sequential tests in clinical trials. Biometrics. 1987; 43(4):857-64.

88. Lan G, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika 1983; 70(3): 659-63.

89. Whitehead A. A prospectively planned cumulative meta-analysis applied to a series of concurrent clinical trials. Stat Med. 1997; 16(24):2901-13.

90. Higgins JP, Whitehead A, Simmonds M. Sequential methods for random-effects meta-analysis. Stat Med. 2011; 30(9):903-21.

91. Whitehead A. A prospectively planned cumulative meta-analysis applied to a series of concurrent clinical trials. Stat Med. 1997 Dec 30;16(24):2901-13.

92. Wetterslev J, Thorlund K, Brok J, Gluud C. Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. J Clin Epidemiol. 2008; 61(1): 64-75.

93. Wetterslev J, Thorlund K, Brok J, Gluud C. Estimating required information size by quantifying diversity in random-effects model meta-analyses. BMC Med Res Methodol. 2009; 9:86.

94. Kulinskaya E, Wood J. Trial sequential methods for meta-analysis. Res Synth Methods. 2014; 5(3):212-20.

95. Brok J, Thorlund K, Wetterslev J, Gluud C. Apparently conclusive meta-analyses may be inconclusive--Trial sequential analysis adjustment of random error risk due to repetitive testing of accumulating data in apparently conclusive neonatal meta-analyses. Int J Epidemiol. 2009; 38(1):287-98.

96. Thorlund K, Imberger G, Walsh M, Chu R, Gluud C, Wetterslev J, Guyatt G, Devereaux PJ, Thabane L. The number of patients and events required to limit the risk of overestimation of intervention effects in meta-analysis--a simulation study. PLoS One. 2011;6(10):e25491.

97. Van der Tweel I, Bollen C. Sequential meta-analysis: an efficient decision-making tool. Clin Trials. 2010; 7(2):136-46.

98. Shuster JJ, Neu J. A Pocock approach to sequential meta-analysis of clinical trials. Res Synth Methods. 2013; 4(3):269-79.

99. Spence GT, Steinsaltz D, Fanshawe TR. A Bayesian approach to sequential meta-analysis. Stat Med. 2016; 35(29):5356-5375.

100. Miladinovic B, Hozo I, Djulbegovic B. Trial sequential boundaries for cumulative meta-analyses. The Stata Journal 2013; 13(1): 77-91.

Chapter 3. Optimal information size in published meta-analyses

Overview

Objectives: To determine the frequency of optimal information size (OIS) estimates in published systematic reviews with meta-analyses and Cochrane policy regarding OIS and how this has changed over time, to assess awareness and perceptions of Cochrane Review Groups (CRGs) regarding OIS estimates and to describe systematic reviews with meta-analyses that include OIS estimates.

Background: No data have been published to date on the number of meta-analyses incorporating OIS estimates nor is it known what Cochrane policy is regarding the use of OIS estimates.

Methods: Medline was systematically searched for systematic reviews with meta-analyses published between 2006 and 2015 that included OIS estimates and the main statistical assumptions used were extracted and analysed. All 53 CRGs were surveyed in 2010 and again in 2016 and responses were analysed to determine awareness and perceptions of OIS estimation in meta-analyses.

Results: A total of 117 systematic reviews with meta-analyses that included OIS estimates were analysed, of which almost half originated in Denmark. Most reported values for type I and II errors, power and effect size but fewer than half reported values for the control event rate and heterogeneity. Overall, only a small proportion of meta-analyses published annually included OIS estimates (0.4%), although recent years have seen a clear upward trend. Regarding the survey of CRGs, 53% and 87% of respondents in 2010 and 2016, respectively, stated having heard of OIS in a meta-analysis context; however, no increase was observed in absolute terms. Only 11% and 13% of respondents in 2010 and 2016, respectively, stated that their CRG had a policy regarding the inclusion of OIS estimation in their meta-analyses.

Conclusion: In meta-analyses that estimate OIS, statistical assumptions used vary greatly, which would suggest that consensus or guidelines are necessary regarding these assumptions. Only a small number of CRGs have a specific policy regarding OIS estimates and this number has largely remained constant over the period of the study. In absolute terms, from the findings of this research it cannot be confirmed conclusively that there is a growing interest in the inclusion of OIS estimates in meta-analyses.

3.1. Introduction

In 1997, Pogue and Yusuf (1) proposed adapting classical monitoring boundaries for use in cumulative meta-analysis as "guidelines for deciding when accumulating evidence is statistically significant and medically convincing". Two further articles from 2008 — by Wetterslev J et al. (2). and Brok J et al. (3) — reinforced that proposal to incorporate OIS estimates in meta-analyses. In 2011, the GRADE (Grading of Recommendations Assessment, Development and Evaluation) Working Group (www.gradeworkinggroup.org) recommended estimating OIS for meta-analyses as an additional precision measure to the use of confidence intervals (4).

The parameters necessary to calculate OIS are similar to those used to calculate sample size for a randomized controlled trial (RCT), namely, the alpha, beta and power values, effect size (ES) — typically expressed as relative risk reduction (RRR) — and the control event rate (CER). However, as documented in Chapter 2, given likely differences in RCTs included in a meta-analysis, inclusion of the heterogeneity parameter, usually expressed as the inconsistency measure I^2 or as the diversity measure D^2, is also recommended by Wetterslev et al. (5).

There appears to be an implicit agreement to use the values 5%, 20% and 80% for type I error, type II error and power, respectively, for OIS estimates (4,5). However, to my knowledge, no consensus document or explicit recommendations have been published regarding numeric values for CER, ES and heterogeneity. If meta-analyses include OIS calculations based on different sets of, and different values for, statistical assumptions, their reliability and

comparability will be undermined. For this reason, it was necessary to explore the literature to determine what statistical parameters and numeric values were used for OIS estimates in published meta-analyses.

Cochrane (http://www.cochrane.org), which aims to ensure that existing evidence is reliable and up-to-date, has published guidelines regarding the minimization of bias (6). Established in 1993, Cochrane is the world's largest organization producing systematic reviews of healthcare interventions, which are published in the *Cochrane Database of Systematic Reviews*, part of the Cochrane Library online platform. Cochrane reviews, as they are commonly known, are indexed in Medline (7). Cochrane's benchmark publication, *The Cochrane Handbook for Systematic Reviews of Interventions* (6), disseminates and highlights the importance of systematic reviews and meta-analyses in healthcare by including guidelines and methodologies for their preparation and performance (8,9).

Such is the Cochrane reputation — at the forefront as it is of fostering the use of suitable methods for conducting systematic reviews — that its opinions and recommendations are widely accepted within the systematic reviewer community. However, Cochrane policy regarding the use of OIS estimates is unknown. Therefore, identifying trends in Cochrane approaches to OIS might help understand how OIS is used at the worldwide level.

Since no data have been published to date on the number or origin of meta-analyses incorporating OIS estimates, the first objective of this chapter is to analyse the frequency of OIS estimates in published Cochrane and non-Cochrane reviews. The second objective is to assess awareness and perceptions of Cochrane Review Groups (CRGs) regarding OIS estimates in Cochrane reviews. A final objective is to describe the published meta-analyses that applied OIS estimates.

3.2. Methods

3.2.1. Literature search

Identification

Since Cochrane reviews represent but a small proportion of all systematic reviews published (9), it was necessary to include other systematic reviews in this analysis in order to accurately determine frequency of OIS use in all types of reviews, regardless of origin.

Medline — a biomedical database of some 24 million journal citations from more than 5,600 scholarly journals published around the world (10) — was systematically searched electronically via PubMed for indexed systematic reviews with meta-analyses that included OIS estimates (referred to, in the interest of brevity, as meta-analyses with OIS estimates in what remains of this chapter) published in the 10-year period 1 January 2006 to 31 December 2015. The year 2006 was selected because an article from that year by Bollen et al. (11) seemed to represent a reliable reference and starting point, although admittedly, any categorical claim that 2006 was the year of publication of the first such meta-analysis is risky, as the fact that previously published such articles have not been identified does not mean that they do not exist.

Used for the search was the Medical Subject Heading (MeSH) 'meta-analysis [pt]' and, as free-text terms, 'systematic review' and — reflecting the different terms used for the concept of calculating the sample size required for an adequately powered study — 'trial sequential analysis', 'optimal information size', 'optimum information size' and 'required information size'.

The reference lists of the retrieved articles were further searched manually to identify other relevant studies for inclusion in the initial sample.

In order to be able to calculate frequency of use of OIS estimates, Medline was again systematically searched electronically via PubMed for all indexed systematic reviews with meta-

analyses (with and without OIS estimates) published in the same 10-year period. Two search strategies were used:

(1) A simple low-sensitivity search strategy using the MESH terms 'meta-analysis [pt]' and the free-text term 'systematic review'.

(2) An advanced search strategy in the form of a validated high-sensitivity strategy designed for PubMed, as reproduced below:

```
(systematic review [ti] OR meta-analysis [pt] OR meta-analysis [ti] OR
systematic literature review [ti] OR this systematic review [tw] OR
pooling project [tw] OR (systematic review [tiab] AND review [pt]) OR meta
synthesis [ti] OR meta-analy*[ti] OR integrative review [tw] OR
integrative research review [tw] OR rapid review [tw] OR umbrella review
[tw] OR consensus development conference [pt] OR practice guideline [pt]
OR drug class reviews [ti] OR cochrane database syst rev [ta] OR acp
journal club [ta] OR health technol assess [ta] OR evid rep technol assess
summ [ta] OR jbi database system rev implement rep [ta]) OR (clinical
guideline [tw] AND management [tw]) OR ((evidence-based[ti] OR evidence-
based medicine [mh] OR best practice* [ti] OR evidence synthesis [tiab]
AND
(review [pt] OR diseases category[mh] OR behaviour and behavior mechanisms
[mh] OR therapeutics [mh] OR evaluation studies[pt] OR validation
studies[pt] OR guideline [pt] OR pmcbook))
OR
(systematic [tw] OR systematically [tw] OR critical [tiab] OR (study
selection [tw]) OR (predetermined [tw] OR inclusion [tw] AND criteri*
[tw]) OR exclusion criteri* [tw] OR main outcome measures [tw] OR standard
of care [tw] OR standards of care [tw])
AND
(survey [tiab] OR surveys [tiab] OR overview* [tw] OR review [tiab] OR
reviews [tiab] OR search* [tw] OR handsearch [tw] OR analysis [ti] OR
critique [tiab] OR appraisal [tw] OR (reduction [tw]AND (risk [mh] OR risk
[tw]) AND (death OR recurrence)))
AND
(literature [tiab] OR articles [tiab] OR publications [tiab] OR
publication [tiab] OR bibliography [tiab] OR bibliographies [tiab] OR
published [tiab] OR pooled data [tw] OR unpublished [tw] OR citation [tw]
OR citations [tw] OR database [tiab] OR internet [tiab] OR textbooks[tiab]
OR references [tw] OR scales [tw] OR papers [tw] OR datasets [tw] OR
trials [tiab] OR meta-analy* [tw] OR(clinical [tiab] AND studies [tiab])
OR treatment outcome [mh] OR treatment outcome [tw] OR pmcbook))
NOT (letter [pt] OR newspaper article [pt])
```

(https://www.nlm.nih.gov/bsd/pubmed_subsets/sysreviews_strategy.html; updated February 2017).

To identify all exclusively Cochrane reviews (with and without OIS estimates), Medline was again systematically searched electronically via PubMed for articles published in the same 10-year period, using the journal name, i.e. 'Cochrane Database Syst Rev [Jour]'.

The automated online PubMed indexing tool was used to analyse publication trends in terms of the number of systematic reviews and meta-analyses published by year (12).

Inclusion/exclusion criteria

Titles and abstracts of the articles in the initial sample were screened according to 3 inclusion criteria: meta-analyses or systematic reviews, in any language, that evaluated any medical or surgical intervention used for treatment or prevention. Meta-analyses of diagnostic techniques were excluded. For articles potentially meeting the inclusion criteria, full texts were retrieved and assessed for their eligibility.

Data extraction

Apart from the meta-data that identified articles, data were identified by screening the methods section and, if necessary, the results section. Extracted data (12 variables) were as follows: author(s), country of origin, year of publication, journal title and primary outcome (5 variables); statistical assumptions regarding type I error, type II error, power, CER and ES for dichotomous outcomes, heterogeneity (6 variables); and, finally, software used to estimate OIS (1 variable). The country of origin was defined as the country of residence of the first author. The primary outcome (dichotomous or continuous or both) was considered as the outcome described as such by the author or, if not stated explicitly, the first-listed outcome in the methods. As for the statistical assumptions, identified were whatever parameters were reported and their values, if provided; identified specifically in relation to heterogeneity was whether this was reported and the measure used (I^2 or D^2 or another measure). In the case that more than one value was used for a variable, the most restrictive value was selected for the analysis (e.g. if alpha values of 1% and 5% were used, then 1% was included in the analysis and 5% was ignored).

Data analysis

Data for the 12 variables were saved in a customized Microsoft Excel spreadsheet and analysed and graphed using Excel and R software. Data were tabulated and frequencies, means and standard deviations were calculated. The results were also represented in graphs, mainly bar charts. The frequency of OIS estimates was represented in 3 ways: (a) as high-sensitivity

frequency, obtained as the number of meta-analyses with OIS estimates divided by the total number of candidate reviews identified in the high-sensitivity search; (b) as low-sensitivity frequency, obtained as the number of meta-analyses with OIS estimates divided by the total number of candidate reviews identified in the low-sensitivity search; and (c) as Cochrane frequency, obtained as the number of Cochrane reviews with OIS estimates divided by the total number of candidate Cochrane reviews identified.

3.2.2. Survey

Study population

Since the Cochrane Collaboration has led the methodological development of systematic reviews, it was considered to be particularly important to obtain the opinions of different CRGs, in particular, regarding policy — if any — on the use of OIS estimates for their meta-analyses. CRGs support Cochrane's primary organizational function — the preparation and maintenance of systematic reviews — by providing editorial support to authors of Cochrane reviews. Cochrane Methods Groups (CMGs) were not consulted in this survey because they do not directly provide support to review authors.

The survey questionnaire, consisting of just 4 questions, was designed to be brief and simple in an endeavour to obtain a high response rate. The questions essentially aimed to determine whether CRGs were aware of or familiar with the OIS methodology and whether they had any specific policy regarding OIS. The survey was designed, tested, piloted, modified and retested using a random sample of 5 CRGs. The final questionnaire was sent by email to all CRGs (n=53), addressed to the contact person mentioned in the webpage of each CRG (www.cochrane.org/contact/review-groups). To compare awareness and perceptions of OIS estimates over time, the questionnaire was sent twice, in December 2010 and again in December 2016 (the second time using Google Forms, in compliance with new Cochrane

Collaboration rules for addressed surveys). An email reminder was sent in March 2011 and 2017 and again in April 2011 and 2017 to all CRGs that had not responded.

Ethical approval was not required for an email/online survey limited to CRGs.

Questionnaire

The questionnaire consisted of the following 4 questions on OIS (3 dichotomous questions and 1 multiple-choice question allowing more than one answer):

Q1. Have you heard or used the terms "trial sequential analysis" or "optimal information size" within the context of a systematic review?	Yes/No
Q2. Does your CRG consider that this or a similar methodology for determining the required number of participants in a systematic review is appropriate?	Yes/No
Q3. Does your CRG have a policy regarding the calculation of a minimal level of information (number of trials/participants) for your systematic reviews?	Yes/No
Q4. If your CRG does not use/recommend this methodology to reviewers, what is the reason?	a. The methodology is poorly developed. b. The reviewers are not familiar with the methodology. c. We are considering using the methodology. d. Other (specify).

Data extraction

Data on responses to the survey questionnaire were entered in a Microsoft Excel spreadsheet. All responses were coded to ensure confidentiality and that no data could be linked to any particular CRG.

To identify reviews associated with the 53 CRGs, Medline was systematically searched electronically via PubMed for Cochrane reviews published in 2010 and in 2016 (the 2 years corresponding to the 2 phases of the survey). The following search terms were used: 'Cochrane Database Syst Rev [Jour]', 'optimal information size', 'trial sequential analysis' and 'required information size'.

Data analysis

The responses from the survey were analysed descriptively. Frequencies and percentages were calculated for the quantitative data and results were tabulated and depicted as graphs (mainly bar charts) using Microsoft Excel software.

3.3. Results

3.3.1. Literature search

Meta-analyses with OIS estimates

A total of 171 potentially relevant meta-analyses were identified in the initial search, of which a final set of 117 was selected as meeting the inclusion criteria. Figure 3.1 depicts the identification-to-inclusion flowchart. A descriptive list of the 117 included meta-analyses is provided in the Appendix (Table A3.1).

Year of publication and country of origin

Figure 3.2 depicts year of publication of meta-analyses with OIS estimates, showing a clear differentiation between the subperiods 2006-2009 and 2010-2015 within the study period. A clear rising trend from 2010 is evident and also notable is the fact that 2015 alone accounts for almost a third (31%) of all meta-analyses with OIS estimates published in the 10-year period 2006-2015.

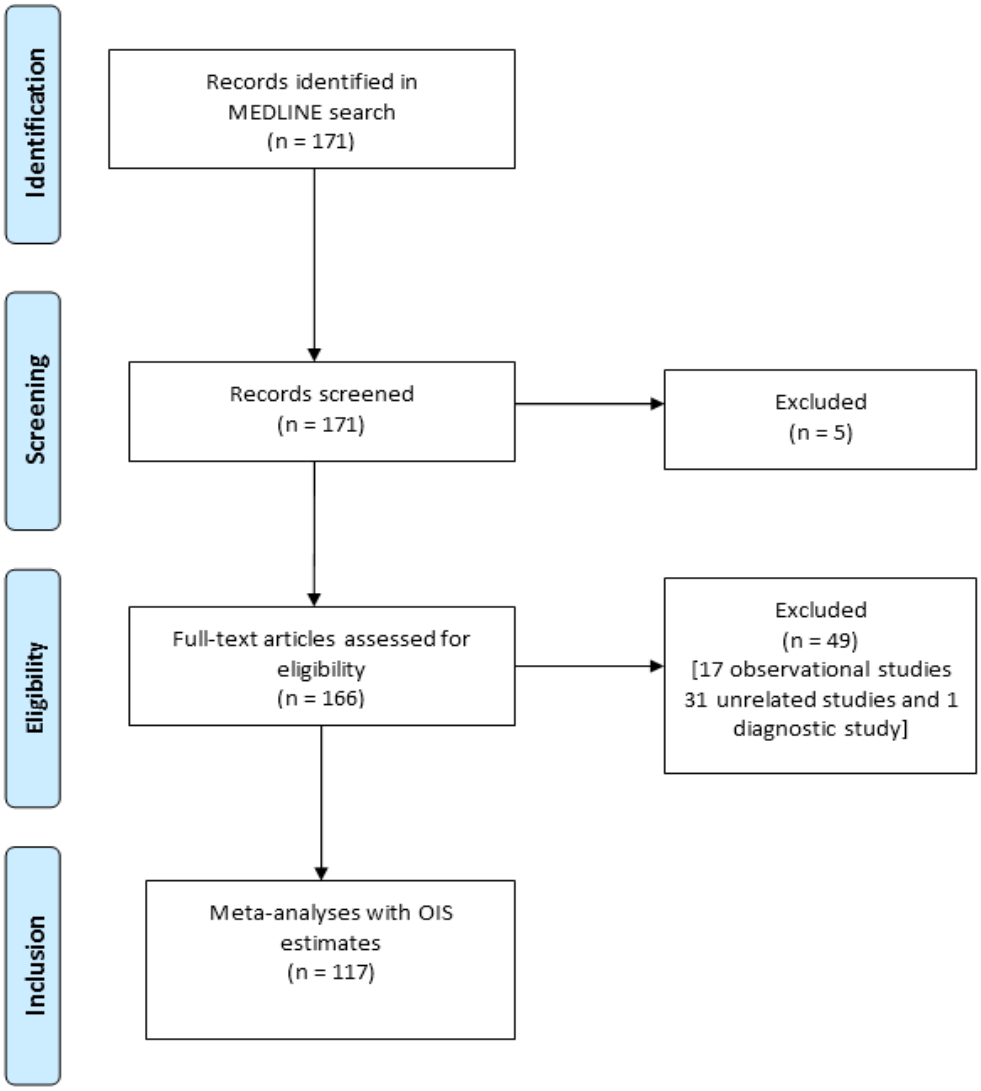


Figure 3.1. Flowchart depicting the literature search.

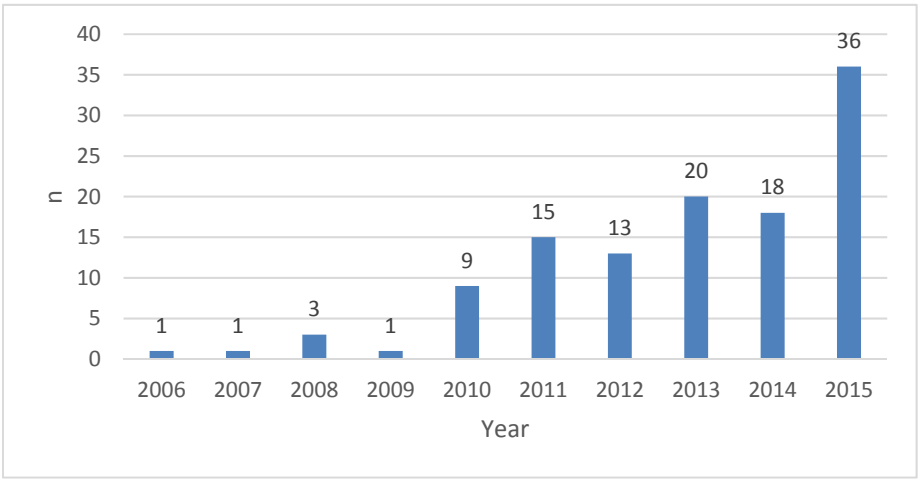


Figure 3.2. Year of publication of meta-analyses with OIS estimates.

Table 3.1 shows distribution by country of origin and by year between 2006 and 2015. Denmark accounted for almost all meta-analyses with OIS estimates published in the first subperiod (2006-2009) and for 43.6% published in the second subperiod (2010-2015). Again, notable is the fact that 2015 witnessed a considerable increase in both the publication and internationalization of meta-analyses with OIS estimates. In terms of worldwide distribution by country, Denmark accounted for nearly half of all meta-analyses with OIS estimates published in the study period (n=51; 43.6%), followed by China (n=17; 14.5%) and then by the UK and the USA (each n=6; 5.1%).

Table 3.1. Meta-analyses with OIS estimates by country of origin and year.

Country	2006-09	2010	2011	2012	2013	2014	2015	2010-15	Total
Denmark	5	7	9	6	10	7	7	46	51
China	-	-	1	1	1	4	10	17	17
UK	-	-	-	1	1	2	2	6	6
USA	-	-	2	-	3	-	1	6	6
Canada	-	1	2	-	1	-	1	5	5
Italy	-	-	-	-	2	2	-	4	4
Netherlands	1	-	-	-	1	-	2	3	4
Serbia	-	-	1	3	-	-	-	4	4
Brazil	-	-	-	-	-	1	2	3	3
Germany	-	-	-	-	1	-	2	3	3
Venezuela	-	-	-	1	-	-	2	3	3
Australia	-	-	-	-	-	-	2	2	2
Croatia	-	-	-	-	-	2	-	2	2
Ecuador	-	-	-	1	-	-	1	2	2
Chile	-	-	-	-	-	-	1	1	1
Japan	-	-	-	-	-	-	1	1	1
Lebanon	-	1	-	-	-	-	-	1	1
Mexico	-	-	-	-	-	-	1	1	1
Norway	-	-	-	-	-	-	1	1	1
Totals:									
- meta-analyses	**6**	**9**	**15**	**13**	**20**	**18**	**36**	**111**	**117**
- countries	**2**	**3**	**5**	**5**	**8**	**6**	**15**	**19**	**19**

Outcome types

Primary outcomes were mostly dichotomous outcomes (n=102; 87.2%), followed by continuous outcomes (n=14; 12%). Studies that reported both continuous and dichotomous outcomes as co-primary outcomes (n=1; 0.8%) were minimally represented.

Statistical assumptions

Table 3.2 summarizes the 6 possible statistical assumptions used in the published meta-analyses with OIS estimates. Type I and II error, power and ES were reported in most meta-analyses, but heterogeneity was only reported in around half (57.3%) and CER in under a quarter (23.5%). The value ranges were narrow for the type I and II errors and power assumptions but broad for the ES, CER and heterogeneity assumptions. Table 3.3 summarizes the exact values used for type I and II errors and power, with most meta-analyses using a type I error of 5% and a power of 80%. A small number of meta-analyses (fewer than 10%) did not report the type I error and power values.

Table 3.2. Meta-analyses with OIS estimates reporting specific parameters and values.

	Reported	Value range	Mean (SD)
Type I error	110 (94%)	1-5%	-
Type II error	107 (91.4%)	10-25%	-
Power	107 (91.4%)	75-90%	-
ES*	93 (79.5%)	2.5-90	20 (12.2)
CER‡	24 (23.5%)	1-85	29 (23.7)
Heterogeneity	67 (57.3%)	-	-

*ES was recorded only for systematic review with dichotomous outcomes.

‡CER can only be calculated when a meta-analysis reports a dichotomous outcome, hence n=102.

Table 3.3. Statistical assumptions reported in meta-analysis with OIS estimates.

Type I error			Power			
1%	**5%**	**NR**	**75%**	**80%**	**90%**	**NR**
5 (4.3%)	105 (89.7%)	7 (6%)	1 (0.8%)	92 (78.6%)	14 (12%)	6 (5.1%)

NR: not reported.

Journals

A total of 37 different scientific journals published the 117 meta-analyses with OIS estimates included in the analysis (see Table A3.2 in the Appendix for the full list), with nearly three quarters publishing just a single such study (n=25; 67.6%). Cochrane reviews published in the *Cochrane Database of Systematic Reviews* accounted for nearly 40% of these meta-analyses (n=43; 36.7%), followed by the *BMJ* (n=10; 8.5%), *PLoS One* (n=10; 8.5%) and *Intensive Care Medicine* (n=8; 6.8%).

A total of 7 different CRGs published the 43 Cochrane reviews (see Table A3.3 in the Appendix for the full list). Just 2 CRGs, namely, the Hepato-Biliary Group and Anaesthesia, Critical and Emergency Care Group, accounted for three quarters (n=32; 74.5%) of the Cochrane reviews; interestingly, both these CRGs are based in Denmark. The remaining 5 CRGs are based in the United Kingdom (Heart Group and Injuries Group), Germany (Metabolic and Endocrine Disorders Group), Australia (Kidney and Transplant Group) and Canada (Upper Gastrointestinal and Pancreatic Diseases Group).

Software

A total of 92 (78.6%) meta-analyses reported the software used to estimate OIS, with all those reporting this information using TSA, a software developed by the Copenhagen Trial Unit (CTU; www.ctu.dk).

OIS frequency

Frequency of use of OIS estimates depended on the search strategy — high-sensitivity or low-sensitivity — used to retrieve systematic reviews with meta-analyses published annually. Thus, annual frequency was below 0.1% and below 0.3% according to the high-sensitivity and low-sensitivity strategies, respectively, except for 2011. As for frequency of OIS use in Cochrane reviews, this was below 0.8% except for 2009 (Table 3.4).

Table 3.4. OIS frequency in meta-analyses by search strategy 2006-2015.

	High-sensitivity search			Low-sensitivity search			Cochrane review search		
Year	n	With OIS	Freq (%)	n	With OIS	Freq (%)	n	With OIS	Freq (%)
2006	11552	1	0.008656	2338	1	0.04277	578	0	0
2007	12882	1	0.007762	2596	1	0.03852	750	0	0
2008	14151	3	0.0211999	2778	3	0.10799	572	1	0.17482
2009	15263	1	0.0065517	3244	1	0.03082	650	1	1.53846
2010	17554	9	0.0512703	3934	9	0.22877	779	3	0.38510
2011	19625	15	0.0764331	486	15	3.08642	757	4	0.52840
2012	22834	13	0.0569326	6567	13	0.19795	1022	9	0.88062
2013	25786	20	0.0775614	8207	20	0.24369	1088	9	0.82720
2014	29456	18	0.0611080	9701	18	0.18554	955	7	0.73298
2015	32367	36	0.1112243	10717	36	0.33591	1039	9	0.86621
2006-2015*	**20147**	**11.7**	**0.04786**	**5056.8**	**11.7**	**0.44983**	**693**	**4.3**	**0.59338**

*Mean values.

Figures 3.3 and 3.4 show publication trends for the study period (2006-2015) for all meta-analyses and for Cochrane reviews, respectively, with and without OIS estimates. In both cases, the number of published meta-analyses has grown steadily, whereas the number of meta-analyses with OIS estimates continues to be but a small proportion of all meta-analyses.

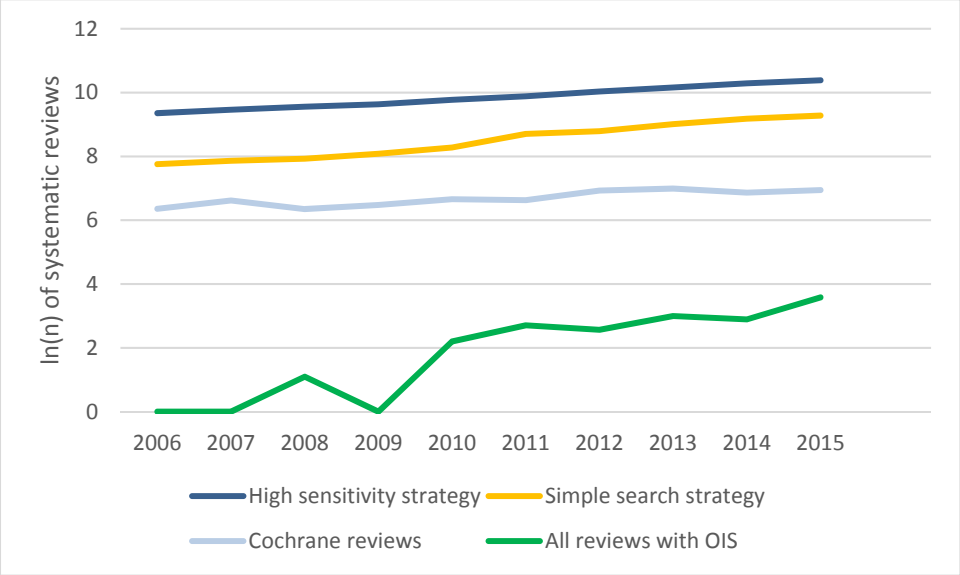


Figure 3.3. Meta-analyses with OIS estimates published 2006-2015 by search strategy (ln values).

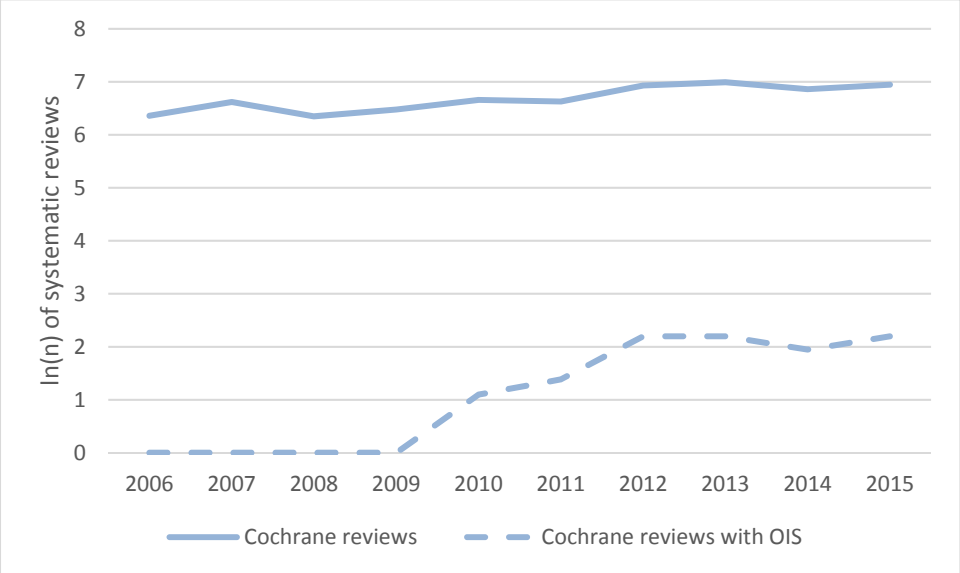


Figure 3.4. Cochrane reviews with OIS estimates published 2006-2015.
n=6930 Cochrane reviews; n=43 Cochrane reviews with OIS estimates (ln values).

3.3.2. Survey

Response rates

Table 3.5 summarizes response rates for the 2 rounds of the survey (2010 and 2016). For 2010, 36 of 53 (68%) CRGs completed the questionnaire after 3 contacts. Of the non-responders (n=17; 32%), 9 (17%) did not respond to any emails, 6 (11.3%) responded but did not return the questionnaire and 2 (3.7%) indicated having a non-specific non-response policy regarding non-Cochrane surveys. For 2016, 15 of 53 (28.3%) CRGs completed the questionnaire after 3 contacts and over half (n=38; 67.8%) did not respond to the survey.

Table 3.5. Response rates from CRGs (n=53) for the 2010 and 2016 surveys.

Year	Contact			Total response rate	
	1st	2nd	3rd	Yes	No
2010	12	16	8	36 (68%)	17 (32%)
2016	9	6	0	15 (28.3%)	38 (67.8%)

Responses to Q1-Q3

Response rates for the dichotomous (yes/no) questions (Q1-Q3) are summarized in Figure 3.5.

Regarding Q1 — "Have you heard or used the terms "trial sequential analysis" or "optimal information size" within the context of a systematic review?" — just over half (n=19; 52.8%) of CRG respondents for the 2010 survey and over three quarters (n=13; 86.7%) of CRG respondents for the 2016 survey responded in the affirmative.

Regarding Q2 — "Does your CRG consider that this or a similar methodology for determining the required number of participants in a systematic review is appropriate?" — 14 (38.9%) CRGs for the 2010 survey and 8 (53.3%) CRGs for the 2016 survey answered in the affirmative. Four anomalous answers for 2010 that were incorrectly completed, despite a request for clarification, were excluded.

Finally, Regarding Q3 — "Does your CRG have a policy regarding the calculation of a minimal level of information (number of trials/participants) for your systematic reviews?" — the proportion of affirmative responses was low: 4 (11.1%) CRGs for the 2010 survey and 2 (13.3·%) CRGs for the 2016 survey. No response was received to follow-up email requests for documentation on this policy. One anomalous answer for 2010 was excluded.

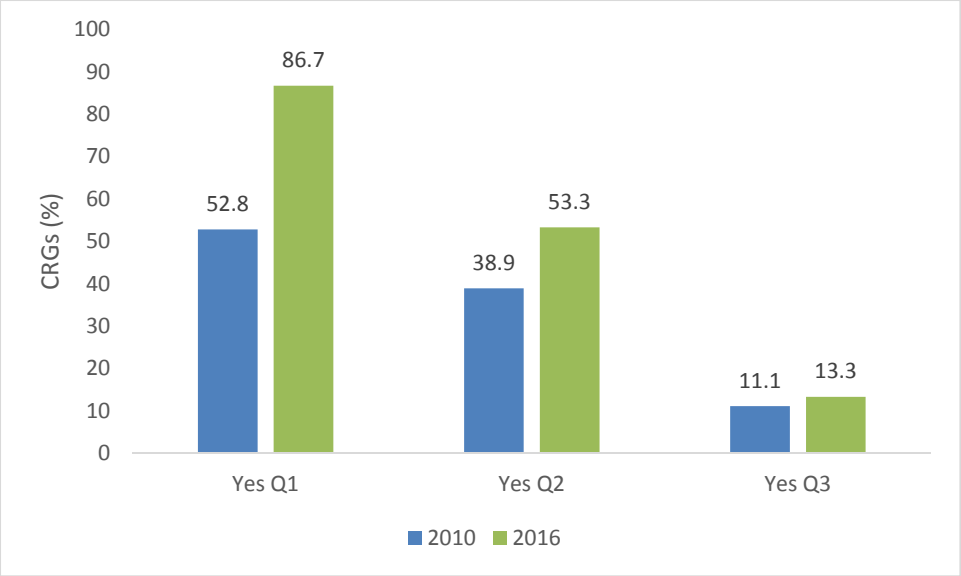


Figure 3.5. CRG response rates for Q1-Q3 in the 2010 and 2016 surveys.
Q1: "Have you heard or used the terms "trial sequential analysis" or "optimal information size" within the context of a systematic review?"
Q2: "Does your CRG consider that this or a similar methodology for determining the required number of participants in a systematic review is appropriate?"
Q3: "Does your CRG have a policy regarding the calculation of a minimal level of information (number of trials/participants) for your systematic reviews?"

Responses to Q4

In response to the question (which allowed multiple answers), "If your CRG does not use/recommend this methodology to reviewers, what is the reason?" (Table 3.6), for the 2010 survey there was a total of 28 responses. Single answers (n=28) by CRGs were as follows: 4 (11.1%) that the methodology was poorly developed (option a), 9 (25%) that reviewers were not familiar with the methodology (option b), 1 (2.8%) that the methodology was under consideration (option c) and 14 (38.9%) 'other' (option d). Multiple answers (n=4) by CRGs were as follows: a+b, 2 (11.1%); and a+d and b+d, 1 (2.8%) each. Additionally, 4 CRGs that responded to Q1-3 did not respond at all to Q4.

For the 2016 survey there was a total of 19 responses. Answers were as follows: 2 (15.4%) that the methodology was poorly developed, 4 (30.8%) that reviewers were not familiar with the methodology, 3 (23.1%) that the methodology was under consideration and, finally, 10 (76.9%) 'other', but without including any comment. As can be observed in Table 3.6, a significantly larger proportion of CRGs that responded to the questionnaire were considering using the OIS methodology in 2016 compared to 2010 (Q4c). Some pertinent comments in response to why CRGs did not use or recommend OIS are reproduced in Table 3.7 (and are discussed in Section 3.4.2).

Table 3.6. CRG response rates for the different response options for Q4.

Response options Q4	2010	2016
a: the methodology is poorly developed	4 (11%)	2 (15%)
b: not familiar with the methodology	9 (25%)	4 (31%)
c: the methodology is under consideration	1 (2.8%)	3 (23.1%)
d: other	14 (38.9%)	10 (77%)
Total	28	19

Table 3.7. Comments in response to Q4d.

Reasons why CRGs do not use/recommend OIS
"Some of the research in our area involves small samples."
"We try not to restrict the review."
"We do not require a minimal number of trials/participants; however, reviewers should discuss the effect of a possible lack of power."
"We are not prescriptive about methodology."
"We take what evidence we can find."

Referring only to the subset of Cochrane reviews, of 425 and 912 such reviews published in 2010 and 2016, respectively, just 3 (1%) and 14 (1.5%) reported OIS estimates in the corresponding years.

3.4. Discussion

3.4.1. Literature search results

Although the first article proposing the use of OIS in meta-analyses was published by Pogue and Yusuf in 1997 (1), uptake of this methodology was initially slow, as indicated by the fact that very few meta-analyses with OIS estimates were published before 2010; there were just 6, for instance, in the 4-year period 2006-2009. However, 2010 marked a turning point, as 9 meta-analyses with OIS estimates were published that year. The period 2011-2015, moreover, was marked by a clear upward trend, with a total of 102 such meta-analyses published (ranging from 13 in 2012 to 36 in 2015). Of the 117 meta-analyses with OIS estimates published between 2006 and 2015 that were included in this 10-year study, 95% were published in the second half of that period and more than a quarter (30.8%) were published in 2015 alone.

The country of origin was Denmark for nearly half (43.6%) of the 117 meta-analyses with OIS estimates included in this study, followed by China (14.5%). In the initial years, in fact, such meta-analyses were almost exclusively accounted for by a single research group — CTU — based in Denmark; in the period 2006-2009 and in the year 2010, this group accounted for 83.3% and almost 87.5% of these meta-analyses, respectively. Thereafter, on average over the 5-year period 2011-2014, Denmark accounted for between 40% and 60% of all meta-analyses with OIS estimates. By 2015, although Denmark's share had dropped to 20% (7 studies), it still was a predominant publisher of meta-analyses with OIS estimates, but now sharing leadership with China (10 studies). Of the remaining 19 studies in 2015, countries accounting for 5 or more studies each were the UK, USA and Canada.

It would appear, therefore, that use of OIS estimates in meta-analyses is becoming more internationalized: whereas only 2 countries accounted for such studies in 2006-2009, by 2015, 15 different countries had produced meta-analyses with OIS estimates, 3 times more than the number corresponding to 2011 (5 countries).

In a breakdown by CRGs, almost three quarters of Cochrane meta-analyses were published by just 2 groups, both based in Denmark: the Hepato-Biliary Group and the Anaesthesia, Critical and Emergency Care Group. Particularly prolific and an early adopter of OIS estimation was the Hepato-Biliary Group, associated with the researcher Christian Gluud, who, as a member of the CTU, actively promotes uses of the TSA software, as does Arash Afshari of the Anaesthesia, Critical and Emergency Care Group. Therefore, this suggests that OIS estimation is not widely or evenly used by CRGs in their systematic reviews. Furthermore, when we consider the overall number of meta-analyses published annually (both Cochrane and non-Cochrane), use of the OIS estimation methodology is anecdotal.

Overall publication frequency of meta-analyses with OIS estimates ranged between just under 0.1% and 3%, depending on the search strategy used (high-sensitivity and low-sensitivity strategies, respectively). As for the subgroup of Cochrane reviews, the estimated frequency was under 0.8% except for a single year, 2009, when frequency was 1.5%.

Regarding the statistical assumptions used to estimate OIS in the meta-analyses included in this study, type I and type II error, power and ES (mainly reported as RRR) were reported in about 90% of the studies, whereas heterogeneity and CER were only reported in around half and under a quarter of the studies, respectively. Thus, for a significant proportion of the studies, no explicit information has been provided regarding the methods used to calculate CER and heterogeneity, and, in some cases, it is not known if heterogeneity was even included as an assumption in estimating OIS. However, since some meta-analyses reported having used TSA software, we could assume — without being absolutely certain — that heterogeneity was included, given that this software, by default, applies a value for I^2 or D^2. Poor reporting of systematic reviews — even when the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; www.prisma-statement.org) checklist is used — has been discussed in a recent study (13) and in several older studies (14-16), pointing to a situation that obviously limits or renders impossible any attempt to replicate the methodology.

The range of values used for type I and II errors and power was narrow, with around 90% of the studies using a type I error value of 5% and around 80% using a power value of 80%. For CER, ES and heterogeneity, the range of values — when reported — was understandably much broader, because such parameters are determined for each meta-analysis and not according to a fixed and commonly accepted parameter (as happens with type I/II errors and power).

Nonetheless, greater debate is necessary on how to select appropriate values for CER, ES and heterogeneity assumptions. One option is to use mean or median values as reported in the RCTs included in the meta-analysis; another option is to use mean values corresponding to only high-quality RCTs (2,3). Heterogeneity assumptions are especially controversial, first in relation to whether this parameter should be included at all in OIS estimates (5,17), but also regarding how to determine the heterogeneity level (18) and specific heterogeneity value (19,20). Li et al. have reported that the inconsistency measure (I^2) may be unstable (21). Other authors propose that heterogeneity should be calculated on the basis of only large RCTs with a greater weight in the meta-analysis (22). Given such issues, it would seem that a broad scientific discussion about the implications and selection of these statistical assumptions is necessary.

It is interesting to note that just over 50% of the meta-analyses with OIS estimates were published in only 3 journals, one of which is rated as among the top 5 medical journals in the world, namely, the *BMJ*. Also noteworthy is the fact that, even though *Cochrane Database Systematic Reviews* accounts for a high proportion of meta-analyses with OIS estimates, no explicit guidelines are provided for reviewers regarding OIS calculation.

3.4.2. Survey results

As can be observed in the analysis of published meta-analyses with OIS estimates, most (40%) were published in *Cochrane Database Systematic Reviews*. As indicated in Chapter 2, the Cochrane Collaboration is an international benchmark in promoting high-quality systematic reviews of healthcare interventions and in developing methodologies for inclusion in its

reference guide, namely, *The Cochrane Handbook for Systematic Reviews of Interventions*. For this reason, it was considered necessary to consult all 53 CRGs regarding their policy, if any, on the use of OIS estimates in their reviews.

The response rate for the first survey of CRGs, conducted in 2010, was close to 70%, quite a high rate for a survey conducted via email (23). The non-response bias was, correspondingly, 30%. For the survey conducted in 2016, the response rate, at just over 28%, was much lower, even though strategies for improving responses from electronic surveys were used, such as sending up to 3 reminders, personalizing emails and stating the average survey completion time (24). The reasons for this poor response are possibly related to a new Cochrane Collaboration policy that limits direct contact with CRGs.

Based on the responses to the 4 questions in the survey, the percentage of groups that had heard of OIS estimation (Q1) was substantially higher in 2016 than in 2010 (86.7% vs. 52.8%); likewise, the percentage of CRGs that were considering using OIS estimates (Q2) was also higher in 2016 than in 2010 (53.3% vs. 38.9%). One possible explanation for the increased proportion of affirmative responses over time is that the response rate to the second survey was much lower, as can be observed in an analysis of absolute numbers rather than percentages.

Regarding the availability of a policy on OIS estimates (Q3), responses were very similar for 2010 and 2016 (around 14%). Nonetheless, no CRG has, to date, officially published its policy on OIS estimation.

Finally, as for the reasons why OIS estimates were not used or recommended (Q4), response rates were similar for 2010 and 2016.

Some CRG comments (see Table 3.7 above) clarifying why OIS estimates were not used are addressed as follows. The comment that “Some of the research in our area involves small samples” may reflect RCTs involving rare diseases. Nonetheless, it may still be of use to have information on OIS to contrast with what information is available, even though it should be

borne in mind that CER, for instance, could be small. Regarding the comments “We try not to restrict the review” and “We take what evidence we can find”, these are more difficult to interpret, as having a greater information size is likely to lead to greater precision in estimates. A similar, more amplified comment was “We do not require a minimal number of trials/participants; however, reviewers should discuss the effect of a possible lack of power”. Yet frequently encountered as a conclusion to a review is that there is insufficient evidence to be able to make a firm recommendation. The use of OIS, however, by contributing information on statistical power could support more precise recommendations regarding the evidence. It could also provide information on whether a meta-analysis requires a larger sample size to ensure adequate power or could support a conclusion regarding inefficacy for a meta-analysis that surpassed OIS. Another comment — “We are not prescriptive about methodology” — reflects the fact that CRG recommendations are not mandatory, yet Cochrane Collaboration recommendations are adhered to by many review authors.

3.4.3. Limitations

Regarding limitations to the literature search, one issue is how meta-analyses that included estimates of OIS were identified, as the nature of electronic searching is such that some such reviews may not have been detected. It is possible that the various information size search terms used may not have identified all studies that included this estimate. That said, the impact is likely to be minimal, as the terminology used is probably quite stable within the relatively small OIS community. Another possible limitation is that only the Medline database was searched and not any other databases, e.g. Embase. Nonetheless, this limitation is not absolute as the evidence would suggest that there is a high degree of duplication between these major databases (25).

Another limitation is in relation to the estimated frequency of use of OIS estimates in meta-analyses — a limitation, in fact, that is inherent to the sensitivity and precision of electronically conducted searches. However, the fact of having conducted a high-sensitivity search and a low-

sensitivity search may indicate a more precise range for the exact number of meta-analyses indexed by Medline in the 10-year study period.

A limitation regarding the survey was that the questionnaire was not validated but was merely tested using a pilot group representing under 10% of CRGs. Another limitation was the poor response rate in 2016 (around 28%) compared to 2010 (close to 70%). This is probably explained by the introduction, in the intervening years, of a more restrictive policy regarding CRG participation in non-Cochrane Collaboration surveys. Thus, in 2010, CRGs were contacted directly and responded directly, whereas in 2016, due to the new policy, all contact was through a central office, so it was not possible to ascertain whether reminders had been sent to individual CRGs. A final limitation regarding the survey was that the fact that responses in 2010 were not anonymous and so may have induced bias on the part of the CRGs.

3.4.4. Implications for research and unanswered questions

Given the variability that exists in how OIS estimates are arrived at, to ultimately enhance the reproducibility and comparability of meta-analysis findings, further research is required into the statistical assumptions underlying OIS estimates and into how to determine CER, ES and heterogeneity values.

Regarding the authorship of reviews with OIS estimates (both Cochrane and non-Cochrane), it would be useful to conduct an analysis based on the mining of author relationship networks (26), so as to verify whether a relevant number of authors belong to a set of closely linked research groups.

Although the use of OIS in relation to the total number of published reviews still remains anecdotal, there has been an evident rising trend since 2010, with 2015 alone accounting for almost a third (31%) of meta-analyses with OIS estimates published in the 10-year period 2006-2015. Given the growing interest in OIS, it would seem necessary for a benchmark body like Cochrane or PRISMA to provide explicit recommendations on OIS estimates, and more

specifically, regarding what statistical parameters to include and how to calculate them. A complementary initiative could be to include a description of OIS calculation in PROSPERO (International Prospective Register of Systematic Reviews; www.crd.york.ac.uk/prospero), which records features from review protocols to be maintained as a permanent record, thereby enabling a comparison of completed reviews with what was planned in the corresponding protocols. Also useful would be the inclusion of these recommendations in journal editorial policies and in instructions for authors.

Regarding the software used, OIS estimation represents an area to be further developed, given that the included meta-analyses mainly used TSA software, developed by the CTU in Denmark, which, as the source of many early meta-analyses with OIS estimates, has made an important contribution to increasing knowledge and improving understanding of the OIS concept. However, TSA adopts a frequentist approach, so it would be of interest to see the development of software that could be used for Bayesian and semi-Bayesian approaches. Another important consideration is the quality of graphic output, which is a drawback of the TSA software. R, for instance, produces excellent graphs and could be adapted to OIS estimation (27).

In relation to the CRG survey for 2016, it would be useful to re-run it — provided a higher response rate could be ensured — for comparison with the results from 2010, for which the response rate was close to 70%. This could be done in partnership with the Cochrane Collaboration or, preferably, could be promoted to CRGs and run directly by the Cochrane Collaboration. It would also be useful to include the CMGs in the survey so as to be able compare and analyse differences in responses from CRGs and CMGs.

3.4.5. Implications for practice

Despite limitations — such as the 'moving target' phenomenon, whereby as a meta-analysis is updated with new studies, the number of patients required to obtain a reliable response varies — OIS estimation has attracted growing interest in the last decade, as indicated by the growth

in the absolute number of published meta-analyses that include OIS estimates. This interest is also reflected in the inclusion of OIS estimation as a GRADE recommendation.

Although OIS will vary depending on the statistical assumptions used, most widely used assumptions from published RCTs and reviews can be used, namely, alpha, beta and power values of 5%, 20% and 80%, respectively — although there is a case for using more conservative values (1%, 10% and 90%, respectively). Regarding CER, ES and heterogeneity values, however, it is not possible to make any definitive recommendation. However, it is proposed that distinguishing between different scenarios, e.g. between high-quality and low-quality RCTs, could be a useful starting point.

3.5. Conclusions

Although OIS estimates are increasingly being used in meta-analyses, the number of such meta-analyses is just a fraction of all meta-analyses published annually (less than 1%). Use of certain statistical assumptions, particularly regarding CER, ES and heterogeneity, is arbitrary, for which reason, further debate and research in this area is necessary.

3.6. Chapter summary

- A growing number of meta-analyses that include OIS estimates have been published in the last decade, with 2010, in particular, marking the beginning of a very notable rising trend.
- Despite the increase in the absolute number of meta-analyses with OIS estimates, frequency of OIS use is still very low in both Cochrane reviews and non-Cochrane reviews (less than 1%).
- No consensus exists in the systematic review community regarding the statistical assumptions and values to be used in OIS estimates.
- OIS is a practical approach to sample size calculation that would potentially enhance the reliability and comparability of meta-analysis findings.

- The proportion of CRGs that have heard of the term OIS is high and has grown between 2010 and 2016.

3.7. Acknowledgements

I would like to thank the 53 CRGs that responded to the surveys conducted in 2010 and 2016. Special thanks to Sheila Wallace, from the Cochrane Incontinence Group, and to Muriah Umoquit, the Cochrane Collaboration Internal Communications and Content Officer, for their help in transmitting the survey to the CRGs.

3.8. References

1. Pogue JM, Yusuf S. Cumulating evidence from randomized trials: utilizing sequential monitoring boundaries for cumulative meta-analysis. Control Clin Trials. 1997; 18(6):580-93.
2. Wetterslev J, Thorlund K, Brok J, Gluud C. Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. J Clin Epidemiol. 2008; 61(1):64-75.
3. Brok J, Thorlund K, Gluud C, Wetterslev J. Trial sequential analysis reveals insufficient information size and potentially false positive results in many meta-analyses. J Clin Epidemiol. 2008; 61(8):763-9.
4. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, Devereaux PJ, Montori VM, Freyschuss B, Vist G, Jaeschke R, Williams JW Jr, Murad MH, Sinclair D, Falck-Ytter Y, Meerpohl J, Whittington C, Thorlund K, Andrews J, Schünemann HJ. GRADE guidelines 6. Rating the quality of evidence--imprecision. J Clin Epidemiol. 2011; 64(12):1283-93.
5. Wetterslev J, Thorlund K, Brok J, Gluud C. Estimating required information size by quantifying diversity in random-effects model meta-analyses. BMC Med Res Methodol. 2009; 9:86.
6. Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011.
7. Clarke M. The Cochrane Collaboration and systematic reviews. Br J Surg. 2007; 94(4):391-2.
8. Kenzie JE, Salanti G, Lewis SC, Altman DG. Meta-analysis and The Cochrane Collaboration: 20 years of the Cochrane Statistical Methods Group. Syst Rev. 2013;2:80.
9. Lefebvre C, Glanville J, Wieland LS, Coles B, Weightman AL. Methodological developments in searching for studies for systematic reviews: past, present and future? Syst Rev. 2013; 2:78.
10. National Institutes of Health (NIH). National Library of Medicine, National Center for Biotechnology Information (NCBI). PubMed. Available from: www.pubmed.com (accessed 15 December 2017).

11. Bollen CW, Uiterwaal CS, van Vught AJ, van der Tweel I. Sequential meta-analysis of past clinical trials to determine the use of a new trial. Epidemiology. 2006; 17(6):644-9.
12. Corlan, A.D (2004). Medline trend: automated yearly statistics of PubMed results for any query. Available from: http://dan.corlan.net/medline-trend.html (accessed 21 June 2016).
13. Faggion CM Jr, Bakas NP, Wasiak J. A survey of prevalence of narrative and systematic reviews in five major medical journals. MC Med Res Methodol. 2017; 17(1):176.
14. Peters JP, Hooft L, Grolman W, Stegeman I. Reporting quality of systematic reviews and meta-analyses of Otorhinolaryngologic articles based on the PRISMA statement. PLoS One. 2015;10:e0136540.
15. Gagnier JJ, Kellam PJ. Reporting and methodological quality of systematic reviews in the orthopaedic literature. J Bone Joint Surg Am. 2013;95:e771–7.
16. Wasiak J, Tyack Z, Ware R, Goodwin N, Faggion CM Jr. Poor methodological quality and reporting standards of systematic reviews in burn care management. Int Wound J. 2017;14:754–63.
17. Kulinskaya E, Wood J. Trial sequential methods for meta-analysis. Res Synth Methods. 2014; 5(3):212-20.
18. Borenstein M, Higgins JP, Hedges LV, Rothstein HR. Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity. Res Synth Methods. 2017; 8(1):5-18.
19. Sangnawakij P, Böhning D, Niwitpong SA, Adams S, Stanton M, Holling H. Meta-analysis without study-specific variance information: Heterogeneity case. Stat Methods Med Res. 2017 Jan 1:962280217718867.
20. Melsen WG, Bootsma MC, Rovers MM, Bonten MJ. The effects of clinical and statistical heterogeneity on the predictive values of results from meta-analyses. Clin Microbiol Infect. 2014; 20(2):123-9.

21. Li SJ, Jiang H, Yang H, Chen W, Peng J, Sun MW, Lu CD, Peng X, Zeng J. The dilemma of heterogeneity tests in meta-analysis: a challenge from a simulation study. PLoS One. 2015; 10(5):e0127538.

22. Wetterslev J, Jakobsen JC, Gluud C. Trial Sequential Analysis in systematic reviews with meta-analysis. BMC Med Res Methodol. 2017; 17: 39.

23. Hardigan PC, Popovici I, Carvajal MJ. Response rate, response time, and economic costs of survey research: A randomized trial of practicing pharmacists. Res Social Adm Pharm. 2016; 12(1):141-8.

24. McPeake J, Bateson M, O'Neill A. Electronic surveys: how to maximise success. Nurse Res. 2014; 21(3):24-6.

25. Qi X, Yang M, Ren W, Jia J, Wang J, Han G, Fan D. Find duplicates among the PubMed, EMBASE, and Cochrane Library Databases in systematic review. PLoS One. 2013; 8(8):e71838.

26. Wang F, Wang X, Yang S. Mining author relationship in scholarly networks based on tripartite citation analysis. PLoS One. 2017;12(11):e0187653.

27. Spence GT, Steinsaltz D, Fanshawe TR. A Bayesian approach to sequential meta-analysis. Stat Med. 2016; 35(29):5356-5375.

Chapter 3. Appendix

Table A3.1. Meta-analyses with OIS estimates published 2006-2015 (n=117).

Note: For ES, the percentages refer to RRR unless otherwise indicated.

Author, year, journal, country	Outcome	Type I error	Type II error	Power	CER	ES	Heterogeneity	Software
Phan, 2015, International Journal of Cardiology, Australia.	Dichotomous	5%	20%	80%	-	50.5%	-	-
Schwendicke, 2015, Journal of Dentistry, Denmark.	Dichotomous	5%	20%	80%	-	20%	D^2	TSA v0.9
Serpa, 2015, Best Practice & Research Clinical Anaesthesiology, Brazil.	Dichotomous	1%, 5%	20%	80%	-	16.5%, 29.0%	-	TSA v0.9
Phan, 2015, International Journal of Cardiology, Australia.	Dichotomous	5%	20%	80%	17.1%	20%	-	TSA v0.9

Claudius, 2015, Systematic reviews, Denmark.	Dichotomous	5%	20%	80%	-	20%	D^2	TSA v0.9
Manzano-Robleda, 2015, Annals of hepatology, Mexico.	Dichotomous	5%	20%	80%	-	32%	-	TSA
Allingstrup, 2015, Cochrane Database Syst Rev, Denmark.	Dichotomous	5%	20%	80%	32%	18%	-	TSA v0.9
Kuriyama, 2015, Intensive Care Med, Japan.	Dichotomous	5%	20%	80%	-	20%	-	TSA v0.9
Koster, 2015, Intensive Care Med, Netherlands.	Dichotomous	5%	10%	90%	-	10%, 20%	D^2	TSA v0.9
Beitland, 2015, Intensive Care Med, Norway.	Dichotomous	5%	20%	80%	13%	20%	-	TSA v0.9
Volbeda, 2015, Intensive Care Med, Netherlands.	Dichotomous	5%	10%	90%, 80%	-	10%, 20%	D^2	TSA v0.9
Gu, 2015, Intensive Care Med, China.	Dichotomous	5%	20%	80%	-	20%	D^2	TSA v0.9
Zampieri, 2015, Critical care, Brazil.	Dichotomous	5%	20%	80%	37.5%	20%	D^2	TSA v0.9

Zhang, 2015, Critical care, China.	Dichotomous	-	-	-	-	-	-	-
Holst, 2015, BMJ, Denmark.	Dichotomous	5%	20%	80%	-	15%, 50%	D^2	TSA v0.9
Zhang, 2015, BMJ, China.	Dichotomous	5%	20%	80%	30%	15%	-	TSA v0.9
Ker, 2015, BMJ, United Kingdom.	Dichotomous	5%	10%	90%	40%	15%	I^2 = 75%	TSA v0.9
Rios-Castellanos, 2015, Cochrane Database Syst Rev, Chile.	Dichotomous	5%	20%	80%	-	20%	D^2	TSA v0.9
Bangalore, 2015, Circ Cardiovasc Interv, USA.	Dichotomous	5%	20%	80%	-	25%	-	TSA v0.9
Fairfield, 2015, Cochrane Database Syst Rev, United Kingdom.	Dichotomous	5%	20%	80%	-	20%	D^2	TSA v0.9
Gluud, 2015, Cochrane Database Syst Rev, Denmark.	Dichotomous	5%	20%	80%	25%	5%	I^2 = 17%	TSA v0.9
Gu, 2015, Journal of Thoracic Disease, China.	Dichotomous	5%	20%	80%	7.9%	20%	-	TSA v0.9

Wang, 2015, J Neurol Neurosurg Psychiatry, China.	Dichotomous	5%	20%	80%	37.7%	-	I^2 = 23%	TSA v0.9
Zhang, 2015, PLoS One, China.	Dichotomous	-	25%	75%	30%	20%	-	TSA v0.9
Schwendicke, 2015, PLoS One, Germany.	Dichotomous	5%	20%	80%	-	20%	-	TSA v0.9
Egerup, 2015, PLoS One, Denmark.	Dichotomous	5%	20%	80%	-	20%	D^2	TSA v0.9
Marti-Carvajal, 2015, Cochrane Database Syst Rev, Venezuela.	Dichotomous	5%	20%	80%	-	20%	D^2	TSA v0.9
Marti-Carvajal, 2015, Cochrane Database Syst Rev, Venezuela.	Dichotomous	5%	20%	80%	-	20%	D^2	TSA v0.9
Wu, 2015, Oncotarget, China.	Dichotomous	5%	20%	80%	-	-	-	TSA v0.9
Tian, 2015, Medicine, China.	Continuous	5%	20%	80%	-	-	-	TSA v0.9
Isayama, 2015, JAMA Pediatr, Canada.	Dichotomous	-	-	-	-	20%	-	-
Simancas-Racines, 2015, Cochrane Database Syst Rev, Ecuador.	Dichotomous	5%	20%	80%	9.3	25%	-	TSA v0.9

Tang, 2015, The Pharmacogenomics Journal, China.	Continuous	5%	20%	80%	-	-	-	TSA v0.9
Weis, 2015, Cochrane Database Syst Rev, Germany.	Dichotomous	5%	20%	80%	-	20	D^2	TSA v0.9
Wetterslev, 2015, Cochrane Database Syst Rev, Denmark.	Dichotomous	5%	20%	80%	15.7%	20%	D^2	TSA v0.9
Wang, 2015, J Neurol Neurosurg Psychiatry, China.	Continuous	5%	20%	80%	-	10%, 30%	-	TSAv0.9
Aranha, 2014, Intensive Care Med, Brazil.	Dichotomous	1%, 5%	10%	90%	36%	20%	-	TSA
Bennett, 2014, Cochrane Database Syst Rev, United Kingdom.	Dichotomous	5%	20%	80%	10%	30%	-	TSA
Bjelakovic, 2014, Cochrane Database Syst Rev, Denmark.	Dichotomous	5%	20%	80%	-	5%	D^2	TSA
Hauser, 2014, Cochrane Database Syst Rev, Denmark.	Dichotomous	5%	20%	80%	-	20%	D^2	TSA

Hauser, 2014, Cochrane Database Syst Rev, Croatia.	Dichotomous	5%	10%	90%	50%	10%	D^2	TSA
Imberger, 2014, British Journal of Anaesthesia, Denmark.	Continuous	5%	10%	90%	-	10%, 20%	D^2	TSAv0.9
Jiang, 2014, Cancer investigation, China.	Dichotomous	5%	20%	80%	-	20%	-	-
Kimer, 2014, Aliment Pharmacol Ther, Denmark.	Dichotomous	5%	20%	80%	-	-	D^2	TSA
Krag, 2014, Intensive Care Med, Denmark.	-	-	-	-	-	-	I^2 and D^2	TSA
Li, 2014, PLoS One, China.	Dichotomous Continuous	5%	20%	80%	-	20%	I^2	TSAv0.9
Maratea, 2014, Clinical Nutrition, Italy.	Dichotomous	5%	20%	80%	-	25%	-	TSA
Nikolova, 2014, Cochrane Database Syst Rev, Denmark.	Dichotomous	5%	20%	80%	-	20%	D^2	TSA
Patel, 2014, BMJ, United Kingdom.	Continuous	5%	20%	80%	-	10%	D^2	TSAv0.9

Penninga, 2014, Cochrane Database Syst Rev, Denmark.	Dichotomous	5%	20%	80%	-	20%	D^2	TSA
Penninga, 2014, Cochrane Database Syst Rev, Denmark.	Dichotomous	5%	20%	80%	-	20%	D^2	TSA
Piccolo, 2014, PLoS One, Italy.	Dichotomous	5%	20%	80%	-	20%, 35% and 50%	-	TSAv0.9
Pin Chen, 2014, J Clin Gastroenterol, China.	Dichotomous	5%	20%	80%	-	-	-	TSA
Zhu, 2014, Intensive Care Med, China.	Dichotomous	5%	20%	80%	-	20%	I^2 and D^2	TSA
Albalawi, 2013, International Journal of Cardiology, Canada.	Dichotomous	5%	20%	80%	-	25%	HAIS, multiplying the previously calculated required information size by 1/(1-I^2), I^2 represent the observed heterogeneity in the meta-analysis	TSA v0.9
Bjelakovic, 2013, PLoS One, Denmark.	Dichotomous	5%	20%	80%	10%	5%	D^2	TSA

Chiarotti, 2013, Int J Cardiol, Italy.	Dichotomous	5%	20%	80%	-	33%, 66%	-	None explained
Gioia, 2013, Int J Cardiol, Italy.	Dichotomous	5%	20%	80%	-	20%	0%	TSA v0.9
Gluud, 2013, The Journal of Nutrition, Denmark.	Dichotomous	5%	20%	80%	27%	65%	51%	TSA
Haase, 2013, BMJ, Denmark.	Dichotomous	5%	20%	80%	-	11%, 20%, 35%	D^2	TSA
Hemmingsen, 2013, Cochrane Database Syst Rev, Denmark.	Dichotomous	5%	20%	80%	-	10% and 30%	D^2	TSA
Hemmingsen, 2013, Cochrane Database Syst Rev, Denmark.	Dichotomous	5%	20%	80%	-	10%	D^2	TSA
Koretz, 2013, Cochrane Database Syst Rev, USA.	Dichotomous	5%	20%	80%	1%	-	-	-
Koning, 2013, PLoS One, Netherlands.	Dichotomous	5%	20%	80%	-	20%	-	TSA
Krag, 2013, Acta Anaesthesiol Scand, Denmark.	Dichotomous	5%	20%	80%	16%	30%	16% D^2	TSA

Miladinovic, 2013, Contemporary Clinical Trials, USA.	Continuous	5%	-	-	-	10%, 15%, 20%	D^2	None explained
Miladinovic, 2013, Journal of Clinical Epidemiology, USA.	Dichotomous	5%	20%	80%	-	15%	$1/1-D^2$	None explained
Penninga, 2013, Cochrane Database Syst Rev, Denmark.	Dichotomous	5%	20%	80%	-	20% included studies with low risk of bias	I^2	TSA
Penninga, 2013, Cochrane Database Syst Rev, Denmark.	Dichotomous	5%	20%	80%	-	20% or the observed in the included studies with low risk of bias	D^2	TSA
Penninga, 2013, Cochrane Database Syst Rev, Denmark.	Dichotomous	5%	20%	80%	-	20% or the observed in the included studies with low risk of bias	I^2	TSA
Uthman, 2013, BMJ, United Kingdom.	Continuous	1%	10%	90%	-	-	D^2	TSAv0.9
Weis, 2013, Cochrane Database Syst Rev, Germany.	Dichotomous	5%	20%	80%	-	20% or the observed in the included trials with low risk of bias	D^2	TSA

Wikkelso, 2013, Cochrane Database Syst Rev, Denmark.	Dichotomous	5%	-	-	-	-	D^2	TSA
Xia, 2013, Cochrane Database Syst Rev, China.	Dichotomous	5%	20%	80%	-	10% RRR or the observed in the included studies with low risk of bias	D^2	TSA
Bjelakovic, 2012, Cochrane Database Syst Rev, Serbia.	Dichotomous	5%	20%	80%	-	5%	D^2	TSA
Fang, 2012, Cochrane Database Syst Rev, China.	Dichotomous	5%	20%	80%	-	20%	-	TSAv0.9
Gluud, 2012, Cochrane Database Syst Rev, Denmark.	Dichotomous	5%	20%	80%	-	25%	$I^2 = 47\%$	-
Hemmingsen, 2012, BMJ, Denmark.	Dichotomous	5%	20%	80%	-	30%	D^2	TSAv0.8
Irving, 2012, Cochrane Database Syst Rev, United Kingdom.	Dichotomous	5%	20%	80%	-	90%	I^2	TSA

Jakobsen, 2012, Journal of Affective Disorders, Denmark.	Continuous	5%	10%	90%	-	Minimal relevant difference of 2 points on the HDRS	-	TSA v0.9
Jakobsen, 2012, Psychological Medicine, Denmark.	Continuous	5%	20%	80%	-	Minimal relevant difference of 2 points on the HDRS or 4 points on the BDI	-	-
Martí-Carvajal, 2012, Cochrane Database Syst Rev, Ecuador.	Dichotomous	5%	20%	80%	-	20%	D^2	TSA
Martí-Carvajal, 2012, Cochrane Database Syst Rev, Venezuela.	Dichotomous	5%	20%	80%	-	60%	D^2	TSA
Moller, 2012, Cochrane Database Syst Rev, Denmark.	Dichotomous	5%	20%	80%	Assumed proportion of events	30%	-	TSA
Rudic, 2012, Cochrane Database Syst Rev, Serbia.	Dichotomus	5%	20%	80%	-	20%	D^2	TSAv0.9

Rudic, 2012, Cochrane Database Syst Rev, Serbia.	Continuous	5%	20%	80%	-	Minimal relevant difference of half a standard deviation	D^2	TSA
Thiele, 2012, Aliment Pharmacol Ther, Denmark.	Dichotomous	5%	20%	80%	-	-	-	-
Afshari, 2011, Anesthesia-Analgesia, Denmark.	Dichotomous	5%	20%	80%	-	10%	A priori heterogeneity adjusted	TSA
Afshari, 2011, Cochrane Database Syst Rev, Denmark.	Dichotomous	5%	20%	80%	-	20%	D^2	TSAv0.8
Bangalore, 2011, Arch Neurol, USA.	Dichotomous	5%	20%	80%	-	20%	D^2	TSA
Bangalore, 2011, BMJ, USA.	Dichotomous	5%	10%-20%	80% - 90%	-	2.5%, 7.5%, 10%, 20%	D^2	TSA
Hemmingsen, 2011, BMJ, Denmark.	Dichotomous	5%	20%	80%	-	10%	D^2	TSA
Jakobsen, 2011, PLoS One, Denmark.	Continuous	5%	10%	90%	-	Minimal relevant	-	-

						difference of 2 points on the HDRS		
Jakobsen, 2011, PLoS One, Denmark.	Continuous	5%	20%	80%	-	Minimal relevant difference of 2 points on the HDRS or 4 points on the BDI	-	-
Jakobsen, 2011, PLoS One, Denmark.	Continuous Dichotomous	5%	10%	90%	-	Continuous: Minimal relevant difference of 2 points on the HDRS Dichotomous : 30%	-	-
Mills, 2011, European Heart Journal, Canada.	Dichotomous	1%	20%	80%	5.5%	20%	-	-
Nielsen, 2011, International Journal of Cardiology, Denmark.	Dichotomous	-	-	-	-	-	D^2	TSA v0.9

Oliveri, 2011, Cochrane Database Syst Rev, Denmark.	Dichotomous	5%	20%	80%	80%	10%	D^2	TSAv8.03
Rudic, 2011, Cochrane Database Syst Rev, Serbia.	Dichotomous	5%	20%	80%	-	-	D^2	TSA
Wells, 2011, CMAJ, Canada.	Dichotomous	-	-	-	-	-	-	Lan-DeMets
Wikkelsoe, 2011, Acta Anaesthesiol Scand, Denmark.	Continuous	5%	20%	80%	-	-	D^2	-
Xia, 2011, Cochrane Database Syst Rev, China.	Dichotomous	5%	20%	80%	-	20% or the observed in the included studies with low risk of bias	D^2	TSA
Afshari, 2010, Cochrane Database Syst Rev, Denmark.	Dichotomous	5%	20%	80%	-	10%	-	TSAv0.8
Afshari, 2010, Cochrane Database Syst Rev, Denmark.	Dichotomous	5%	-	-	-	10%	-	TSAv0.8

Brok, 2010, Aliment Pharmacol Ther, Denmark.	Dichotomous	5%	20%	80%	85%	10%	I^2	TSAv0.8
Brok, 2010, Cochrane Database Syst Rev, Lebanon.	Dichotomous	5%	10%	90%	-	10%	-	-
Gluud, 2010, Aliment Pharmacol Ther, Denmark.	Dichotomous	5%	20%	80%	-	-	-	-
Huusom, 2010, International Journal of Obstetrics and Gynaecology, Denmark.	Dichotomous	1%, 5%	10%	90%	5%	25%	-	-
Keus, 2010, Journal of Clinical Epidemiology, Denmark.	Dichotomous	5%	20%	80%	-	20%	-	TSA
Penninga, 2010, Eur J Clin Pharmacol, Denmark.	Dichotomous	5%	20%	80%	-	20%	-	-
Whitfield, 2009, Cochrane Database Syst Rev, Denmark.	Dichotomous	5%	20%	80%	-	20%	-	TSA

Afshari, 2008, Cochrane Database Syst Rev, Denmark.	Dichotomous	5%	20%	80%	-	5%	-	-
Moller, 2008, European Heart Journal, Denmark.	Dichotomous	5%	20%	80%	-	20%	-	TSA
Rambaldi, 2008, Aliment Pharmacol Ther, Denmark.	Dichotomous	5%	-	-	-	20% or the observed in trials with low risk of bias	-	-
Afshari, 2007, BMJ, Denmark.	Dichotomous	5%	20%	80%	-	10%	-	-
Bollen, 2006, Epidemiology, Netherlands.	Dichotomous	-	-	-	-	-	-	-

Table A3.2. Meta-analyses with OIS estimates published in scientific journals (n=37) 2006-2015.

Journal	n (%)
1. Cochrane Database Systematic Reviews	43 (36.7)
2. British Medical Journal (BMJ)	10 (8.5)
3. PLoS One	10 (8.5)
4. Intensive Care Medicine	8 (6.8)
5. Alimentary Pharmacology & Therapeutics	5 (4.2)
6. International Journal of Cardiology	4 (3.4)
7. Acta Anaesthesiologica Scandinavica	2 (1.7)
8. Critical Care	2 (1.7)
9. European Heart Journal	2 (1.7)
10. International Journal of Cardiology	2 (1.7)
11. Journal of Clinical Epidemiology	2 (1.7)
12. Journal of Neurology, Neurosurgery, and Psychiatry |	2 (1.7)
13. Anesthesia & Analgesia	1 (0.85)
14. Annals of Hepatology	1 (0.85)
15. Archives of Neurology	1 (0.85)
16. Best Practice and Research Clinical Anaesthesiology	1 (0.85)
17. British Journal of Anaesthesia	1 (0.85)
18. Cancer Investigation	1 (0.85)
19. Circulation: Cardiovascular Interventions	1 (0.85)
20. Clinical Epidemiology	1 (0.85)
21. Clinical Nutrition	1 (0.85)
22. Canadian Medical Association Journal	1 (0.85)
23. Contemporary Clinical Trials	1 (0.85)
24. Epidemiology	1 (0.85)
25. European Journal of Clinical Pharmacology	1 (0.85)
26. International Journal of Obstetrics and Gynaecology	1 (0.85)
27. Journal of Clinical Gastroenterology	1 (0.85)
28. Journal of the American Medical Association Pediatrics	1 (0.85)
29. Journal of Affective Disorders	1 (0.85)
30. Journal of Dentistry	1 (0.85)
31. Journal of Thoracic Disease	1 (0.85)
32. Medicine	1 (0.85)
33. Oncotarget	1 (0.85)
34. Psychological Medicine	1 (0.85)
35. Systematic Reviews	1 (0.85)
36. Journal of Nutrition	1 (0.85)
37. Pharmacogenomics Journal	1 (0.85)

Table A3.3. CRGs that published systematic reviews with OIS (n=43) 2006-2015.

Cochrane Review Group	n (%)
Hepato-Biliary Group	24 (55.8)
Anaesthesia, Critical and Emergency Care Group	8 (18.7)
Heart Group	3 (7)
Metabolic and Endocrine Disorders Group	3 (7)
Injuries Group	2 (4.6)
Kidney and Transplant Group	2 (4.6)
Upper Gastrointestinal and Pancreatic Diseases Group	1 (2.3)

Chapter 4. Impact of heterogeneity and effect size on optimal information size estimates

Overview

Objectives: To estimate the proportion of systematic reviews with meta-analyses that meet the optimal information size (OIS) and assess the impact of heterogeneity and effect size (ES) on the OIS estimate by outcome type (all-cause mortality/semi-objective/subjective).

Background: OIS estimation is influenced by variables such as the control event rate, ES and heterogeneity. However, because there is no consensus regarding these parameters, the assumptions are difficult to define in advance.

Methods: Medline and the Cochrane Collaboration Archie Database were searched to retrieve meta-analyses published in systematic reviews from 2010 to 2012. The OIS, estimated using TSA v0.9 software, was stratified by Cochrane and non-Cochrane reviews and outcome type for 3 different heterogeneity scenarios.

Results: Included were 137 (63%) meta-analyses from 218 potential systematic reviews (one meta-analysis per systematic review): 83 (61%) Cochrane and 54 (39%) non-Cochrane. The Cochrane and non-Cochrane reviews included a mean of 6.5 (SD 6.1) studies and 13.2 (SD 10.2) studies, respectively. The mean (SD; median) number of patients was 2619.1 (6245.8; 586.0) in Cochrane reviews and 19888.5 (32925.7; 6566.5) in non-Cochrane reviews. The percentage of Cochrane and non-Cochrane reviews (respectively) achieving OIS were as follows: 0% and 25% for all-cause

mortality; 17% and 46% for semi-objective outcomes; and 45% and 72% for subjective outcomes. OIS estimates incorporating heterogeneity could be calculated from the formula for the regression line, bearing in mind the outcome type evaluated in the review.

Conclusions: The results point to wide variability in the range of values for ES and heterogeneity that impact on OIS calculation, irrespective of meta-analysis source (Cochrane or non-Cochrane). The number of meta-analyses that achieved OIS was low and varied depending on outcome type and review type (Cochrane vs non-Cochrane). Fewer than half of the primary outcomes synthesized in systematic reviews achieved OIS, which could indicate that their conclusions are subject to substantial uncertainty.

Note. An earlier version of this chapter has been published as: Garcia-Alamino JM, Bankhead C, Heneghan C, Pidduck N, Perera R. Impact of heterogeneity and effect size on the estimation of the optimal information size: analysis of recently published meta-analyses. BMJ Open. 2017 Nov 8;7(11): e015888. doi: 10.1136/bmjopen-2017-015888.

4.1. Introduction

The concept of optimum information size (OIS) was first proposed in 1997 by Pogue et al. (1,2) as "the minimum amount of information required in the collective literature for reliable conclusions about an intervention to be reached". OIS estimates are based on standard sample size calculations. Thus, the required number of participants (information size) for a meta-analysis should match the number required for any individual adequately powered randomized controlled trial (RCT) (3). Although other measures of information size have been proposed (4,5), estimating OIS is relatively simple, but, in some scenarios, will underestimate the information required to determine whether or not sufficiently firm evidence has been obtained that enable robust conclusions (6). Brok et al. demonstrated, for a subset of Cochrane reviews, that many meta-analyses have false positive

results due to insufficient information (3), while Turner et al. showed that most meta-analyses do not have sufficient power to identify even moderate effects (7,8).

Sample size and OIS calculation are influenced by variables such as the alpha value and statistical power, as well as baseline risk (the control event rate, CER), effect size (ES, usually expressed as relative risk (RR) or relative risk reduction (RRR)) and heterogeneity (usually expressed as I^2). Deciding which values to use can be difficult and is typically based on values observed or estimated from the meta-analysis or one of the included RCTs. Increased variation can also affect the OIS esimate, and there is currently no consensus about which value of heterogeneity should be used.

An OIS estimate can help determine the stability of an effect and whether treatment effect estimates are likely to differ based on further information. However, it is difficult to define in advance and there is no consensus regarding the alpha (significance) or power value to be used at the outset.

Since it is not currently known whether evidence accumulation and the associated OIS depend on the type of outcome studied or whether this varies by publication type (Cochrane or non-Cochrane review), the objective of this chapter was to estimate the proportion of systematic reviews with meta-analyses that meet the optimal information size (OIS) and to assess the impact of heterogeneity and ES on the OIS estimate by outcome type (all-cause mortality/semi-objective/subjective).

4.2. Methods

4.2.1. Literature search

Two sets of systematic reviews with meta-analyses were defined: Cochrane and non-Cochrane. For Cochrane reviews, the Archie Database (http://archie.cochrane.org) — which contains all Cochrane

published reviews and allows electronic searches — was searched for systematic reviews with meta-analyses published in the period 2010-2012; the Microsoft Excel random number generation feature was then used to select 120 of these reviews for inclusion in the study. For non-Cochrane reviews, Medline (PubMed) was searched for all systematic reviews with meta-analyses published in the top 5 general medical journals in 2010-2012 using the following search strategy: "BMJ"(Journal) OR "Ann Intern Med"(Journal) OR "JAMA"(Journal) OR "Lancet"(Journal) OR "N Engl J Med"(Journal) AND (systematic review (ti) OR meta-analysis (pt) OR meta-analysis (ti)).

4.2.2. Inclusion/exclusion criteria

One meta-analysis per systematic review was included and one outcome — the first outcome based on binary data from 2 or more RCTs — was selected for each Cochrane review (e.g. outcome 1.1) and each non-Cochrane review. If several outcomes were reported, the first outcome that met the inclusion criterion was selected (see Figure A4.1 in the Apprendix).

Meta-analyses that included observational studies and diagnostic interventions, network meta-analyses and meta-analyses with no effect (pooled effect = 1) or with no events in all the included RCTs were also excluded.

4.2.3. Data extraction

Full texts were obtained for meta-analyses whose abstracts met the inclusion criteria after assessment for eligibility. One researcher (JMGA) extracted the data into customized Microsoft Excel spreadsheets and a second researcher (RP or NP) checked the data. Extracted and calculated for each included meta-analysis were the following items: outcome type, comparison, number of included patients, number of RCTs, number of events in the control and intervention arms, CER, ES (usually RRR) and heterogeneity (usually I^2). Outcome types were as defined by Turner et al. (9),

namely, all-cause mortality, semi-objective outcome (cause-specific mortality, major morbidity event) and subjective outcome (pain, mental health outcomes).

4.2.4. Data analysis

Data were extracted from each RCT and the meta-analysis was repeated using a DerSimonian and Laird random-effects model to account for potential heterogeneity of effects. Estimates for RCTs with only one group and reporting zero events were adjusted with a constant continuity correction factor of 0.5 for each arm (as per the default adjustment in the Revman software). The estimates obtained for the pooled RRR and I^2 were compared with published results to detect differences, and, if necessary, the analyses were repeated to identify the source of the difference. Meta-analyses and OIS calculation were performed using TSA (v0.9) freeware (10) (www.ctu.dk/tsa). This software is designed to perform meta-analyses of dichotomous or continuous data under fixed-effects or random-effects models and has options for estimating OIS and stopping boundaries according to the O'Brien-Fleming and Lan-DeMets methods.

To evaluate the impact of changes in heterogeneity (I^2) and ES (RRR), OIS was estimated for different scenarios.

For heterogeneity (I^2), 3 scenarios were analysed, determined on the basis of estimates of predictive distributions published by Rhodes et al. (11): 'heterogeneity = rep' (i.e. as reported in the meta-analysis using a random-effects model or obtained from fitting a random-effects model if a fixed-effects model was originally used); 'heterogeneity = 0'; and 'heterogeneity = Q3' (upper quartile or 75th percentile). These last 2 estimates of heterogeneity (=0 and =Q3) were extreme scenarios, chosen to evaluate the impact on OIS. In line with Rhodes (11), outcome type (i.e. all-cause mortality, semi-objective and subjective) was considered, based, for simplicity sake, on assuming a mean study size of 50-200 participants.

For RRR, 2 scenarios were defined, namely, the RRR for each meta-analysis and — as an alternative more stringent value — an a priori conservative value of 5%, as reported by Djulbegovic et al. (12). Relative risk (RR) was transformed to RRR using the formula RRR =1-RR. If RR was >1 the RRR was taken to be negative. An alternative estimate for the semi-objective and subjective outcomes was not determined, as the distribution of possible effects would make the choice of an 'average effect' difficult to justify.

For CER (baseline risk), only one value was used per meta-analysis, namely, the median of the proportion of events in the RCTs included in each meta-analysis, as proposed by Hayden et al. (13).

Descriptive statistics and plots were used to quantify Cochrane review vs non-Cochrane review differences in CER, I^2 and RRR, stratified by outcome type (6 groups in total). Also determined for Cochrane reviews vs non-Cochrane reviews was the proportion of meta-analyses that achieved OIS, based on reported results and our 2 extreme scenarios (heterogeneity = 0 and heterogeneity = Q3), also stratified by outcome type.

SPSS v.22 software was used for the descriptive analysis of the characteristics of the included meta-analyses.

A simple linear regression analysis was conducted to investigate the correlation between the measures obtained for the different heterogeneity scenarios. These analyses were carried out for all outcomes combined as well as separately for mortality, semi-objective and subjective outcomes. The two scenarios modelled were: 'zero heterogeneity' against 'heterogeneity estimated for the review' and 'zero heterogeneity' against 'heterogeneity for Q3' as reported by Rhodes (11). The focus of these models was to determine if an OIS estimate with zero heterogeneity could provide adequate estimates of and OIS with a predefined level of heterogeneity as this would simplify their calculation.

R software was used for the linear regression analysis.

4.3. Results

4.3.1. Literature search

Figure 4.1 depicts a flow chart of the literature search results. Of the 120 randomly selected Cochrane reviews, 24 were excluded and of the 98 non-Cochrane reviews, 34 were excluded. Finally included therefore was a total of 137 meta-analyses from 218 potential reviews: 83 (61%) Cochrane reviews and 54 (39%) non-Cochrane reviews.

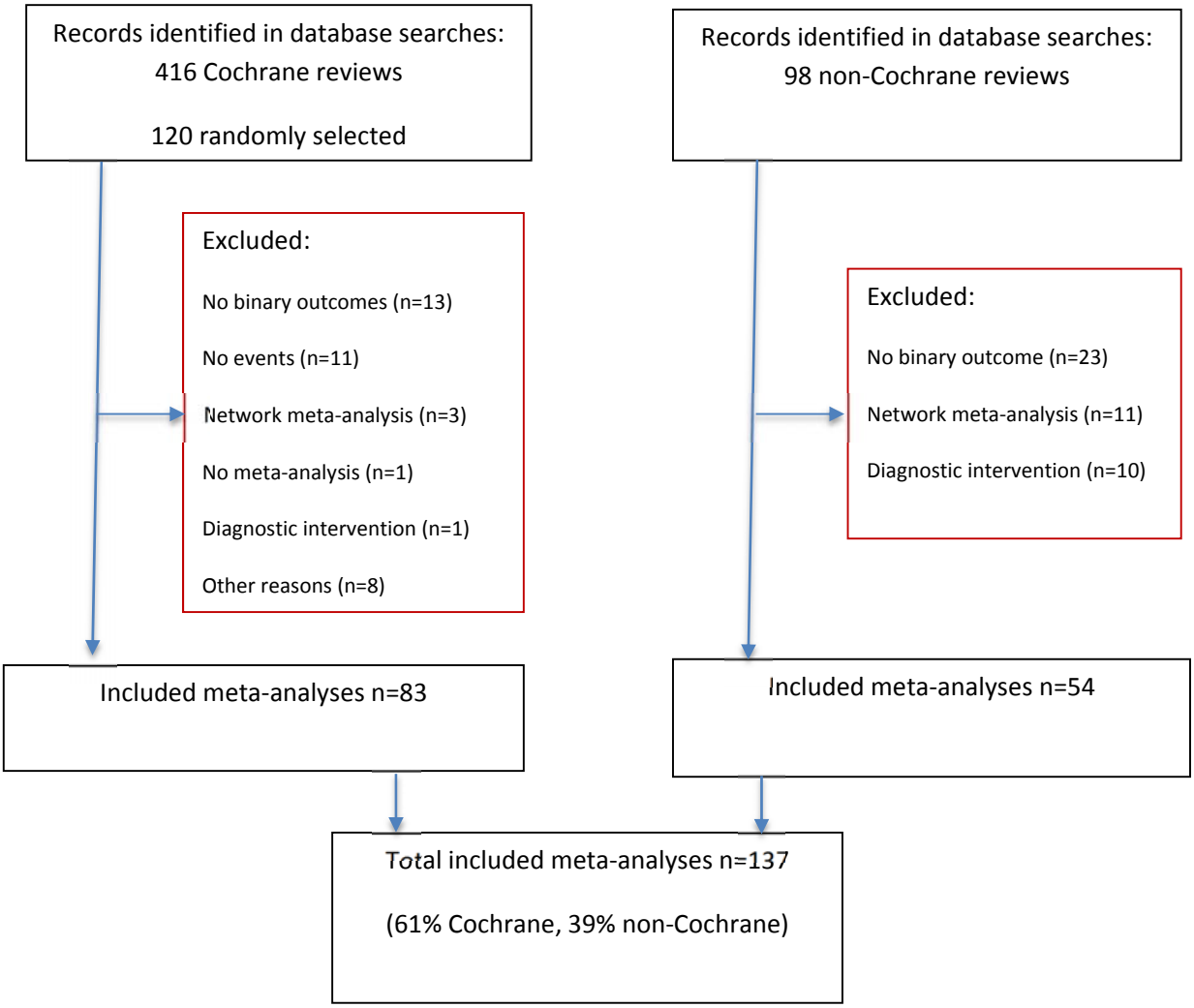


Figure 4.1. Flowchart depicting the literature search.

The mean (SD) of included RCTs was 6.5 (6.1) for the Cochrane reviews and 13.2 (10.2) for the non-Cochrane reviews. The mean (SD; median) number of included patients was 2619.1 (6245.8; 586.0) for the Cochrane reviews and 19888.5 (32925.7; 6566.5) for the non-Cochrane reviews.

4.3.2. Parameter estimation scenarios

Table 4.1 summarizes results for outcome types and intervention types (pharmacological and nonpharmacological) for the included meta-analyses broken down by review type (Cochrane and

non-Cochrane reviews). The most frequent outcome type was subjective (50.4%) followed by semi-objective (31%), and the most frequent intervention type was pharmacological (59%).

Table 4.1. Outcome and intervention types by Cochrane reviews and non-Cochrane reviews.

	Cochrane (n=83)	Non-Cochrane (n=54)	All reviews (n=137)
Outcome type	% (n/N)	% (n/N)	% (n/N)
All-cause mortality	16.8% (14/83)	22.2% (12/54)	19% (26/137)
Semi-objective	21.7% (18/83)	44.4% (24/54)	30.7% (42/137)
Subjective	61.4% (51/83)	33.3% (18/54)	50.4% (69/137)
Intervention type	% (n/N)	% (n/N)	% (n/N)
Pharmacological	56.6% (47/83)	62.9% (34/54)	59.1% (81/137)
Nonpharmacological	44.6% (36/83)	35.2% (20/54)	40.9% (56/137)

There were significant differences in the outcome type (χ^2=16.881; df=2; p=0.00021) but not in the intervention type (χ^2=1.4087 df=2; p=0.4944) between the Cochrane and non-Cochrane reviews.

Table 4.2 shows CER, RRR and I^2 values, the number of patients and OIS estimates, overall and by review type and broken down by outcome type. There were considerable variations in the data. For all outcomes analysed, the number of included patients was higher in the non-Cochrane reviews. The lowest mean CER was observed for all-cause mortality, with a distribution that differed depending on outcome type. The highest mean RRR and I^2 values were observed for subjective outcomes.

Table 4.2. Descriptive results for statistical assumptions in the included meta-analyses.

	CER	RRR*	I^2	Patients	OIS estimate¥
All reviews, n=137					
Mean (SD)	26.9 (26.1)	28.2 (31.5)	20.4 (26.1)	9426.0 (22753.9)	386441.1 (1645397.1)
Cochrane, n=83					
Mean (SD)	24.0 (27.9)	21.0 (36.6)	0.0 (25.5)	586 (6245.8)	2301.0 (1422086.6)
Non-Cochrane, n=54					
Mean (SD)	10.0 (21.7)	20.0 (20.6)	14.5 (26.9)	6566.5 (32925.7)	7299.5 (1946750.2)
All-cause mortality, n=26					
Mean (SD)	12.7 (16.5)	20.3 (22.1)	10.9 (17.6)	14314.6 (25880.1)	499090.7 (1966940.8)
Cochrane, n=14					
Mean (SD)	10.0 (11.2)	25.5 (27.3)	6.7 (13.7)	6902.5 (12971.1)	813301.7 (2678677.7)
Non-Cochrane, n=12					
Mean (SD)	15.8 (21.4)	14.3 (12.6)	15.7 (20.8)	22962.1 (34232.8)	132511.2 (201708.8)
Semi-objective, n=42					
Mean (SD)	17.2 (20.9)	19.0 (18.7)	18.8 (25.9)	18683.5 (33125.7)	828450.9 (2444468.1)
Cochrane, n=18					
Mean (SD)	18.5 (10.5)	21.7 (24.0)	12.8 (22.7)	3384.8 (4762.7)	536616.8 (1803091.2)
Non-Cochrane, n=24					
Mean (SD)	16.3 (20.0)	16.9 (13.7)	23.3 (27.5)	30157.0 (40233.9)	1047326.5 (2851698.9)
Subjective, n=69					
Mean (SD)	38.2 (27.1)	36.8 (37.9)	24.9 (28.1)	1948.9 (2971.0)	74944.0 (406792.0)
Cochrane, n=51					
Mean (SD)	41.6 (27.9)	36.5 (41.6)	23.5 (27.6)	1173.0 (2243.3)	78688.6 (449189.7)
Non-Cochrane, n=18					
Mean (SD)	28.5 (22.9)	37.6 (25.3)	28.9 (29.7)	4147.3 (3683.7)	64334.2 (261366.6)

*All values were considered to be positive.

¥ Estimate according to scenario 2 (heterogeneity = 0) and alpha 5%.

Figures 4.2-4.9 show regression lines fitted to the natural logarithm (ln) for the 2 regression scenarios ('zero heterogeneity' against 'heterogeneity estimated for the review' and 'heterogeneity estimated for the review' against 'heterogeneity for Q3') and for each of the 3 outcomes types (mortality, semi-objective and subjective).

Figure 4.2 depicts the regression line from the ln for OIS assuming 'zero heterogeneity' against 'heterogeneity estimated for the review' and Figure 4.3 depicts the regression line from the ln for OIS assuming 'zero heterogeneity' against 'heterogeneity for Q3'. In both cases, a strong positive linear correlation — y=0.9739x+0.5474; R^2=0.9753 in Figure 4.2. and y=0.9566x+0.8297; R^2=0.9941 in Figure 4.3 — can be observed, indicates that 99% of the variation in ln can be predicted by the regression line.

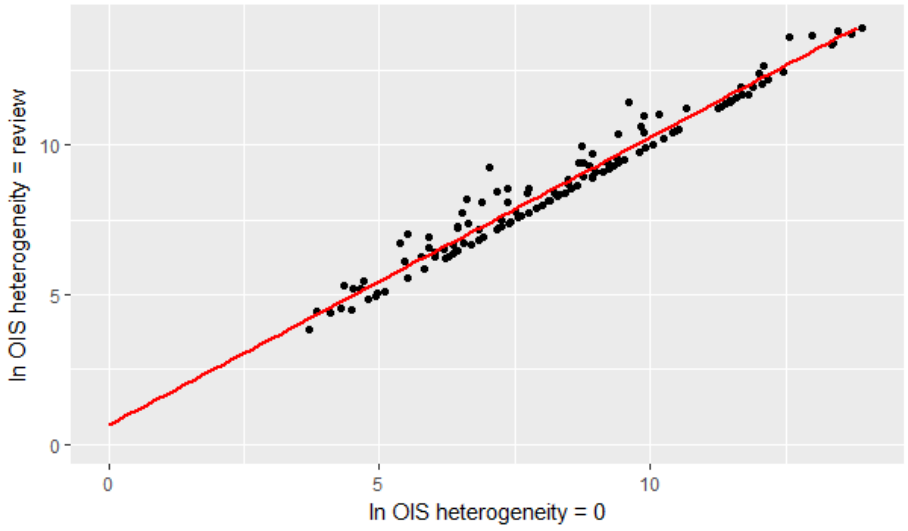


Figure 4.2. Correlation between OIS estimates assuming 'zero heterogeneity' against 'heterogeneity for the review': all reviews.

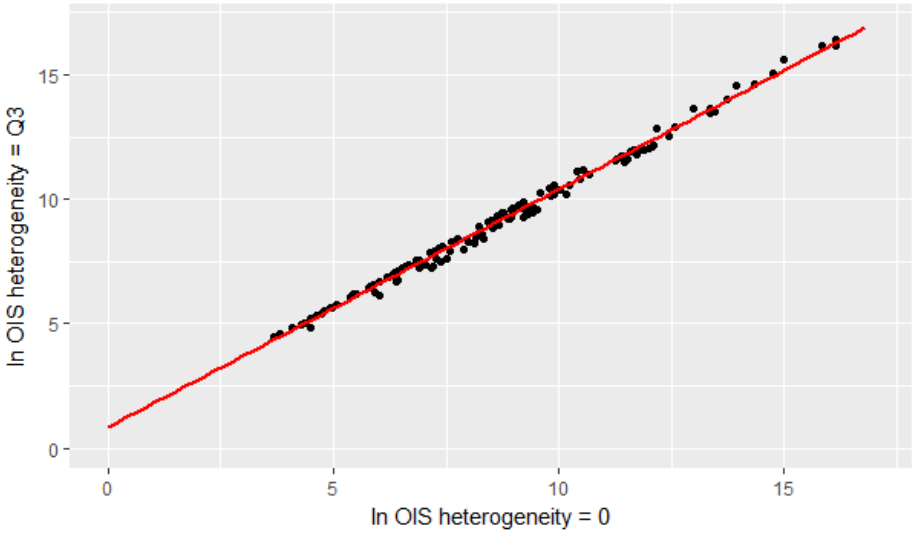


Figure 4.3. Correlation between OIS estimates assuming 'zero heterogeneity' against 'heterogeneity for Q3': all reviews.

Figure 4.4 depicts the regression line for the mortality outcome from the ln for OIS assuming 'zero heterogeneity' against 'heterogeneity for the review' and Figure 4.5 depicts the regression line for the mortality outcome from the ln for OIS assuming 'zero heterogeneity' against 'heterogeneity for Q3'. In both cases, the strong positive linear correlation — $y=1.031x-0.1787$; $R^2=0.9918$ in Figure 4.4 and $y=0.9995x+0.077$; $R^2=0.9918$ in Figure 4.5 — indicates that 99% of the variation in ln can be predicted by the regression line.

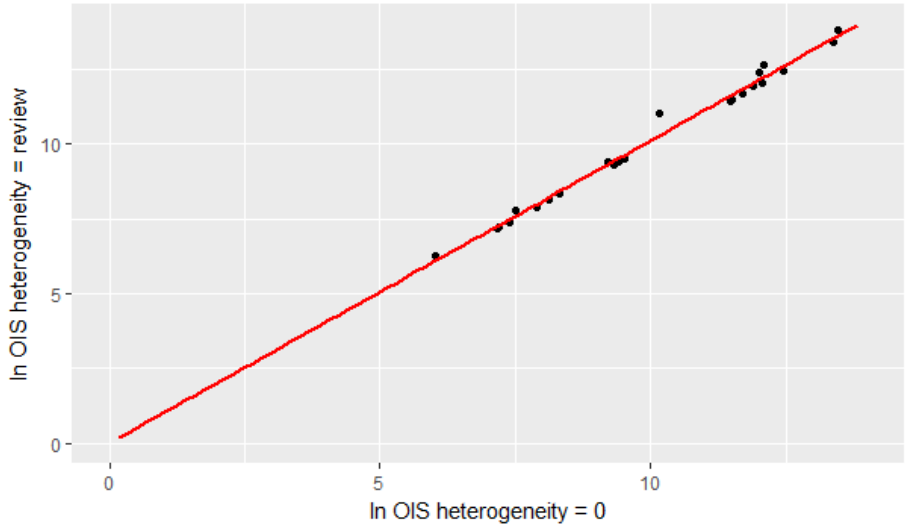


Figure 4.4. Correlation between OIS estimates assuming 'zero heterogeneity' against 'heterogeneity for the review': mortality outcome.

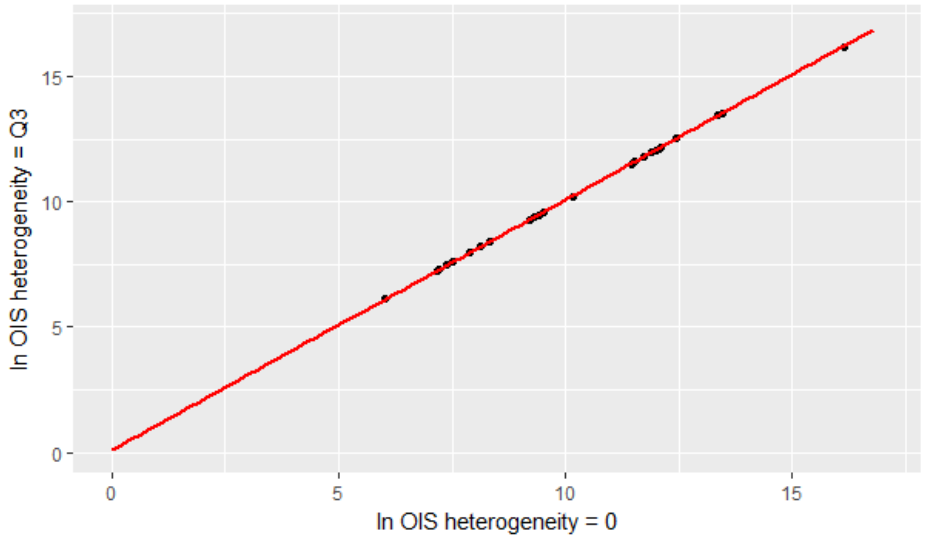


Figure 4.5. Correlation between OIS estimates assuming 'zero heterogeneity' against 'heterogeneity for Q3': mortality outcome.

Figure 4.6 depicts the regression line for the semi-objective outcome from the ln for OIS assuming 'zero heterogeneity' against 'heterogeneity for the review' and Figure 4.7 depicts the regression line for the semi-objective outcome from the ln for OIS assuming 'zero heterogeneity' against 'heterogeneity for Q3'. In both cases, the strong positive linear correlation — y=0.9746x+0.5401;

R^2=0.9739 in Figure 4.6 and y=0.9986x+0.327; R^2=1 in Figure 4.7 — indicates that 100% of the variation in ln can be predicted by the regression line.

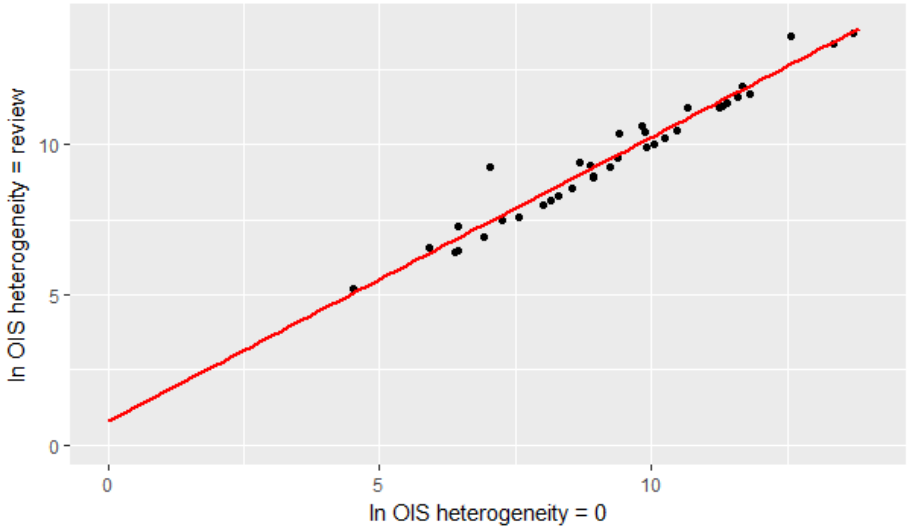


Figure 4.6. Correlation between OIS estimates assuming 'zero heterogeneity' against 'heterogeneity for the review': semi-objective outcome.

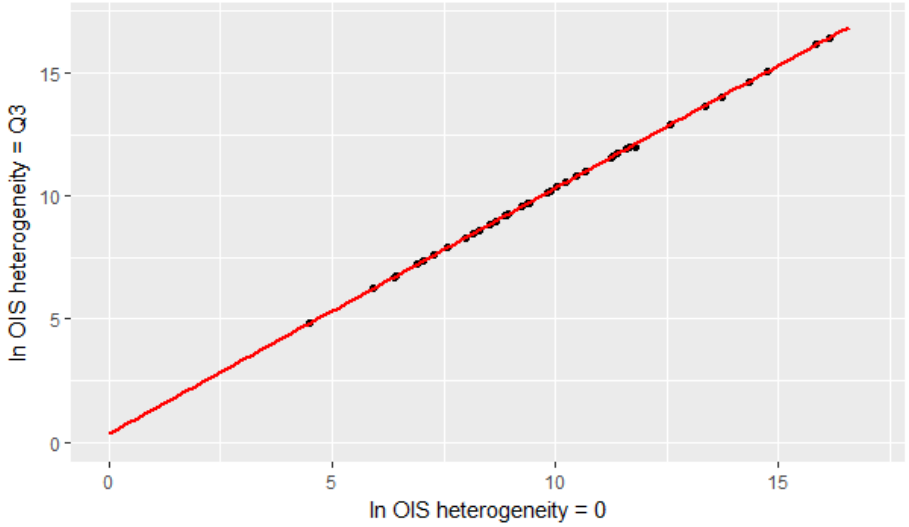


Figure 4.7. Correlation between OIS estimates assuming 'zero heterogeneity' against 'heterogeneity for Q3': semi-objective outcome.

Finally, Figure 4.8 depicts the regression line for the subjective outcome from the ln for OIS assuming 'zero heterogeneity' against 'heterogeneity for the review' and Figure 4.9 depicts the regression line for the subjective outcome from the ln for OIS assuming 'zero heterogeneity' against 'heterogeneity

for Q3'. In both cases, the strong positive linear correlation — y=0.9801x+0.5539; R^2=0.955 in Figure 4.8 and y=0.9958x+0.7128; R^2=1 in Figure 4.9 — indicates that 100% of the variation in ln can be predicted by the regression line.

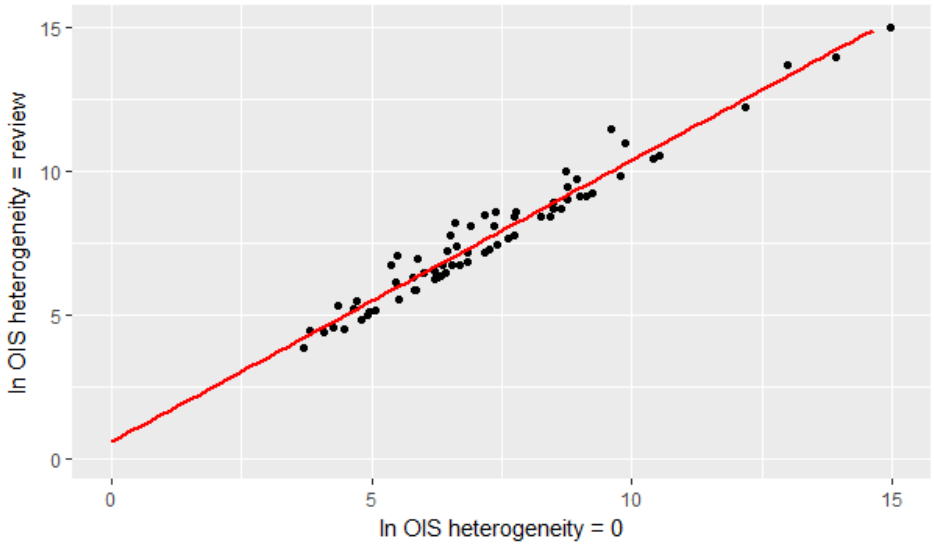


Figure 4.8. Correlation between OIS estimates assuming 'zero heterogeneity' against 'heterogeneity for the review': subjective outcome.

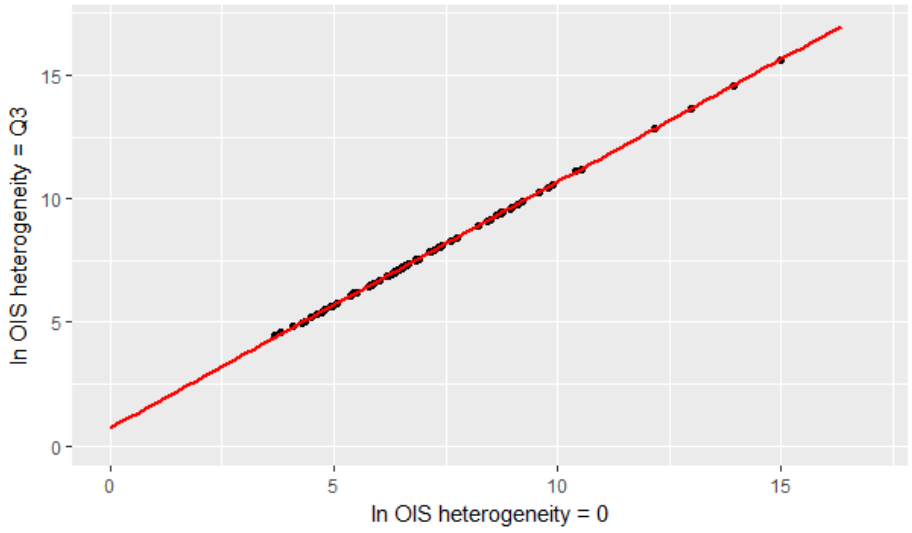


Figure 4.9. Correlation between OIS estimates assuming 'zero heterogeneity' against 'heterogeneity for Q3': subjective outcome.

4.3.3. Meta-analyses that achieved OIS

Figure 4.10(a) depicts the estimated OIS for each meta-analysis for the extreme scenario of heterogeneity = 0. All-cause mortality required the highest OIS in both Cochrane and non-Cochrane reviews, although only marginally more so than semi-objective outcomes. For the subjective outcomes, OIS estimates were considerably smaller due to higher CER and RRR values. Figure 4.10(b) shows the number of meta-analyses that achieved sample sizes equal to or greater than the estimated OIS, with — as illustrated in Figure 4.10(c) — more non-Cochrane reviews achieving this estimate.

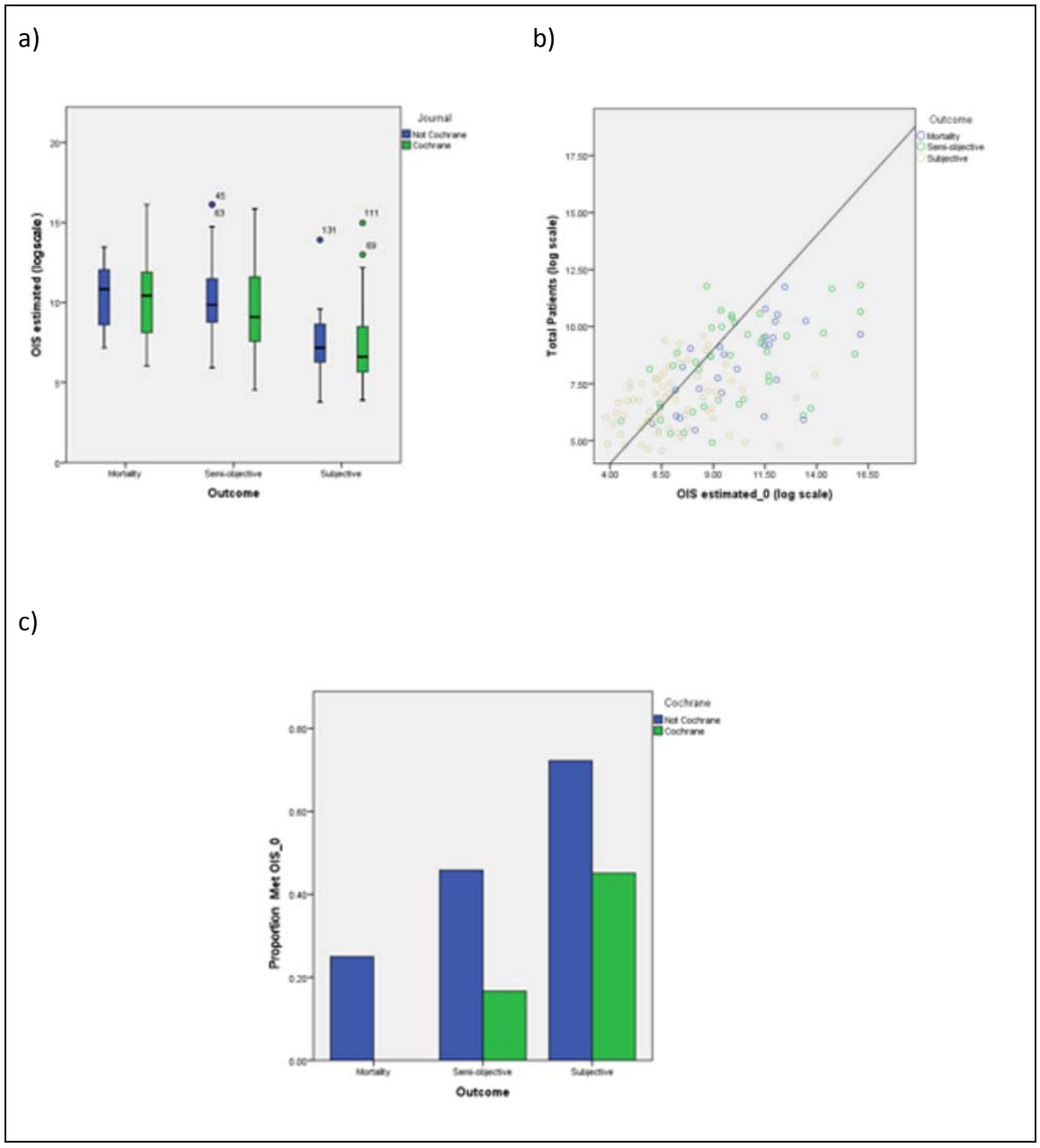


Figure 4.10. OIS estimated considering the best-case scenario ('zero heterogeneity')
(a) type of outcome and source, with the y-axis is on the log scale; (b) related to the total number of patients included in each review; and (c) proportion of reviews achieving the OIS by type of outcome and source.

OIS estimates based on reported I^2 show that the necessary sample size was only reduced for Cochrane reviews reporting subjective outcomes (Table 4.3). Further increasing I^2 (worst-case

scenario: heterogeneity = Q3) did not substantially change the proportion of meta-analyses achieving OIS.

Table 4.3. Meta-analyses achieving OIS according to assumed I^2 for different outcome types.

	All-cause mortality		
% (n/N)	Cochrane	Non-Cochrane	95% CI
OIS achieved with I^2=rep	0% 0/14	25% 3/12	(-0.074, 0.571)
OIS achieved with I^2=0	0% 0/14	25% 3/12	(-0.074, 0.571)
OIS achieved with I^2=Q3	0% 0/14	16.6% 2/12	(-0.134, 0.491)
	Semi-objective		
% (n/N)	Cochrane	Non-Cochrane	95% CI
OIS achieved with I^2=rep	11.1% 2/18	37.5% 9/24	(-0.043, 0.499)
OIS achieved with I^2=0	16.6% 3/18	45.8% 11/24	(-0.031, 0.534)
OIS achieved with I^2=Q3	11.1% 2/18	33.3% 8/24	(-0.079, 0.460)
	Subjective		
% (n/N)	Cochrane	Non-Cochrane	95% CI
OIS achieved with I^2=rep	31.4% 16/51	72.2% 13/18	(0.111, 0.616)
OIS achieved with I^2=0	45.1% 23/51	72.2% 13/18	(-0.024, 0.490)
OIS achieved with I^2=Q3	29.4% 15/51	61.1% 11/18	(0.027, 0.553)

When the more stringent 5% RRR estimate was used for all-cause mortality, no meta-analysis (0/14 Cochrane and 0/12 non-Cochrane) achieved the sample size necessary to comply with OIS.

4.4. Discussion

4.4.1. Statement of principal findings

The results point to wide variability in the range of values for RRR, I^2 and CER that impact on OIS calculation, irrespective of meta-analysis source (Cochrane or non-Cochrane). This variability is partially explained by the outcome type (all-cause mortality, semi-objective or subjective). OIS estimates could therefore be obtained for different types of outcomes, as previously proposed by Turner et al. (9) and Rhodes et al. (11); to my knowledge, this is the first time that accounting for outcome type in estimating OIS has been proposed. Also found was that the outcome type impacts on the range of I^2 levels observed, with this impact being particularly high for subjective outcomes. One possible explanation could be a higher number of smaller RCTs. Differences in I^2 levels were more marked in Cochrane reviews; non-Cochrane reviews showed more similar levels across all outcome types. The results show that, overall, few recently published meta-analyses in high-quality journals achieved OIS, which would indicate that they did not have adequate statistical power to draw firm conclusions.

As expected, estimating OIS assuming different I^2 and alpha values showed a strong correlation. Although specialist software can be used to estimate OIS (e.g. TSA v0.9), it is also possible to use any software that allows sample size estimates provided the I^2 level is assumed to be zero. Heterogeneity can also be incorporated using a simple adjustment proposed by Wetterslev et al. (5), who proposes using an alternative index to I^2, called the diversity statistic (D^2). However, there is still no consensus on what precise measure of heterogeneity to use in estimating OIS (4,14).

Published meta-analyses that estimate OIS often use one or more statistical assumptions, such as an RRR of 10% or the median RRR for RCTs with low risk of bias (15-17). This analysis shows that the

median RRR was 20% for all pooled reviews. However, because the distribution of RRR varies by outcome type, in some cases OIS may be either underestimated or overestimated.

Also observed was strong correlation in estimating OIS assuming 'zero heterogeneity' against 'heterogeneity for the review' and 'heterogeneity for the review' against 'heterogeneity for Q3'. Given this strong correlation, using the regression line formula makes it possible to estimate OIS, according to the outcome type, for the 'heterogeneity for the review' or 'heterogeneity for the Q3' scenarios from an estimate of OIS for 'zero heterogeneity'.

4.4.2. Limitations

Several statistics have been proposed to define a "desirable sample size in terms of numbers of participants across all studies" (4). OIS, as described in this chapter, involves relatively simple calculations, which, if anything, are likely to underestimate the information required to define whether sufficiently firm evidence has been obtained on which to draw robust conclusions (4,14). This definition of OIS was used as a measure to estimate what proportion of systematic reviews with meta-analyses met this minimum requirement.

A single threshold was calculated to define when/if a minimum level of evidence had been collected. However, retrospective analyses of results from meta-analyses are more commonly used to inform prospective studies — for instance, to determine the size of a new RCT aimed at definitively answering a question regarding efficacy. The use of trial sequential methods has been proposed to identify early signs of effect, with monitoring boundaries defined by frequentist, semi-Bayesian or fully Bayesian methods (4,18,19). Although there is still considerable uncertainty about estimates and the best method to use, empirical studies have provided examples to suggest that such methods could help detect benefit, harm or futility signals from early on (8,20). Note that determining the

sample size required for a new study or studies will depend on the method used in the meta-analysis (14).

Reviews conducted by the Cochrane Collaboration are considered to be of higher quality (21,22) and of greater methodological rigour than meta-analyses published in paper-based journals. Since this study only included meta-analyses from the top 5 medical journals, our results may not be applicable to other meta-analyses published in other journals. Nevertheless, this would bias our results towards better evidence being evaluated to what is currently being generated.

Another limitation is related to having used the values given in the article from Turner (9), as the predictive distribution for between-study heterogeneity only included meta-analyses from Cochrane reviews, which are not necessarily representative of meta-analyses in general.

These results, furthermore, cannot be generalized to network meta-analyses, a rapidly growing area of evidence synthesis (23). A recently published study demonstrated that substantial variation exists in network-based meta-analyses (24); also, the statistical methodology to estimate OIS in network meta-analyses is less developed than for traditional meta-analyses, hence our exclusion of these studies.

The use of correlation as a measure of agreement has been shown to be fatally flawed (25). However, the focus of the analysis presented in figures 4.2 to 4.9 is on predicting OIS estimates that incorporate heterogeneity (for example based on heterogeneity for Q3) from an OIS estimate with zero heterogeneity. By providing these regression values, OIS estimates that incorporate heterogeneity can be calculated from simple sample size calculators without the need for TSA or other specialized packages.

4.4.3. Implications for researchers and methodologists

This study has shown that, when estimating the OIS, outcome type can be used as a proxy for defining the basic required parameters, namely, CER, RRR and I^2. Systematic reviewers can use these results to estimate an OIS value for their primary outcome, independently of their confidence in the specific parameters obtained from their review. Reviewers are therefore encouraged to use sample size estimation as a measure of likely confidence in their results. The fact that over 50% of primary outcomes in recent systematic reviews appear to fall below this minimum requirement points to the need for further evidence to reduce uncertainty.

From the perspective of future research, it would also be interesting to explore the predictive distribution for between-study heterogeneity for a sample that also included non-Cochrane meta-analyses.

It would also be useful to further explore ES behaviour depending on the quality of individual RCTs included in the meta-analysis; this could be done by observing the impact on OIS of estimates for high-quality vs poor-quality RCTs, given that it has been clearly documented that the latter underestimate ES.

Another research topic of interest would be to analyse how ES varied as new RCTs are included in a meta-analysis, given the observation that ES varies over time in cumulative meta-analyses (26) and may decay from a large ES for initial results to a smaller ES for later results reported for a meta-analysis. Thus, if the ES used to calculate OIS is that for initial results, the OIS estimate may be low, with the consequence that further RCTs may not be designed and started, given that the OIS has been perceived or understood (erroneously) to be achieved.

4.5. Conclusions

Heterogeneity and ES impact on the OIS, which can be estimated using traditional sample size estimation software, with adjustments, if necessary, for heterogeneity. The above results demonstrate that outcome type is relevant to OIS estimation as well as to heterogeneity, CER and ES estimates. Currently, under half of the meta-analyses published to date in high-quality journals have achieved OIS, which would indicate that conclusions based on their results are subject to substantial uncertainty.

4.6. Chapter summary

- There is great variability in terms of the parameters necessary to calculate OIS according to outcome type.
- Few recently published meta-analyses in high-quality journals achieved OIS.
- The median ES (RRR) was 20% for all pooled reviews.
- As expected, OIS estimated assuming different I^2 values were strongly correlated.
- Reviewers are encouraged to use OIS as a measure of likely confidence in their meta-analysis results.

4.7. References

1. Pogue J, Yusuf S. Cumulating evidence from randomized trials: utilizing sequential monitoring boundaries for cumulative meta-analysis. Control Clin Trials 1997;18:580–93.
2. Pogue JM, Yusuf S. Overcoming the limitations of current meta-analysis of randomised controlled trials. Lancet 1998;351:47–52.
3. Brok J, Thorlund K, Gluud C, et al. Trial sequential analysis reveals insufficient information size and potentially false positive results in many meta-analyses. J Clin Epidemiol 2008;61:763–9.
4. Higgins JP, Whitehead A, Simmonds M. Sequential methods for random-effects meta-analysis. Stat Med 2011;30:903–21.
5. Wetterslev J, Thorlund K, Brok J, et al. Estimating required information size by quantifying diversity in random-effects model meta-analyses. BMC Med Res Methodol 2009;9:86.
6. Wetterslev J, Thorlund K, Brok J, et al. Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. J Clin Epidemiol 2008;61:64–75.
7. Turner RM, Bird SM, Higgins JP. The impact of study size on meta-analyses: examination of underpowered studies in Cochrane reviews. PLoS One 2013;8:e59202.
8. Imberger G, Thorlund K, Gluud C, et al. False-positive findings in Cochrane meta-analyses with and without application of trial sequential analysis: an empirical review. BMJ Open 2016;6:e011890.
9. Turner RM, Davey J, Clarke MJ, et al. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the cochrane database of systematic reviews. Int J Epidemiol 2012;41:818–27.
10. Thorlund K, Engstrøm J, Wetterslev J. User manual for trial sequential analysis (TSA). Trial Unit Available Internet 2011 (accessed 11 Jun 2014).

11. Rhodes KM, Turner RM, Higgins JP. Empirical evidence about inconsistency among studies in a pair-wise meta-analysis. Res Synth Methods 2016;7:346–70.
12. Djulbegovic B, Kumar A, Glasziou PP, et al. New treatments compared to established treatments in randomized trials. Cochrane Database Syst Rev 2012;10:MR000024.
13. Hayden JA, Tomlinson G, Ni A, et al. Approaches to estimate and present baseline risks considerations for cochrane reviews evidence-informed practice integration of: 19th Cochrane Colloquium Madrid, 2011.
14. Kulinskaya E, Wood J. Trial sequential methods for meta-analysis. Res Synth Methods 2014;5:212–20.
15. Imberger G, Orr A, Thorlund K, et al. Does anaesthesia with nitrous oxide affect mortality or cardiovascular morbidity? A systematic review with meta-analysis and trial sequential analysis. Br J Anaesth 2014;112:410–26.
16. Bangalore S, Kumar S, Wetterslev J, et al. Angiotensin receptor blockers and risk of myocardial infarction: meta-analyses and trial sequential analyses of 147 020 patients from randomised trials. BMJ 2011;342:d2234.
17. Afshari A, Wetterslev J, Brok J, et al. Antithrombin III in critically ill patients: systematic review with meta-analysis and trial sequential analysis. BMJ 2007;335:1248–51.
18. Wetterslev J, Jakobsen JC, Gluud C. Trial sequential analysis in systematic reviews with meta-analysis. BMC Med Res Methodol 2017;17:39.
19. Spence GT, Steinsaltz D, Fanshawe TR. A Bayesian approach to sequential meta-analysis. Stat Med 2016;35:5356–75.
20. AlBalawi Z, McAlister FA, Thorlund K, et al. Random error in cardiovascular meta-analyses: how common are false positive and false negative results? Int J Cardiol 2013;168:1102–7.

21. Remschmidt C, Wichmann O, Harder T. Methodological quality of systematic reviews on influenza vaccination. Vaccine 2014;32:1678–84.

22. Windsor B, Popovich I, Jordan V, et al. Methodological quality of systematic reviews in subfertility: a comparison of Cochrane and non-Cochrane systematic reviews in assisted reproductive technologies. Hum Reprod 2012;27:3460–6.

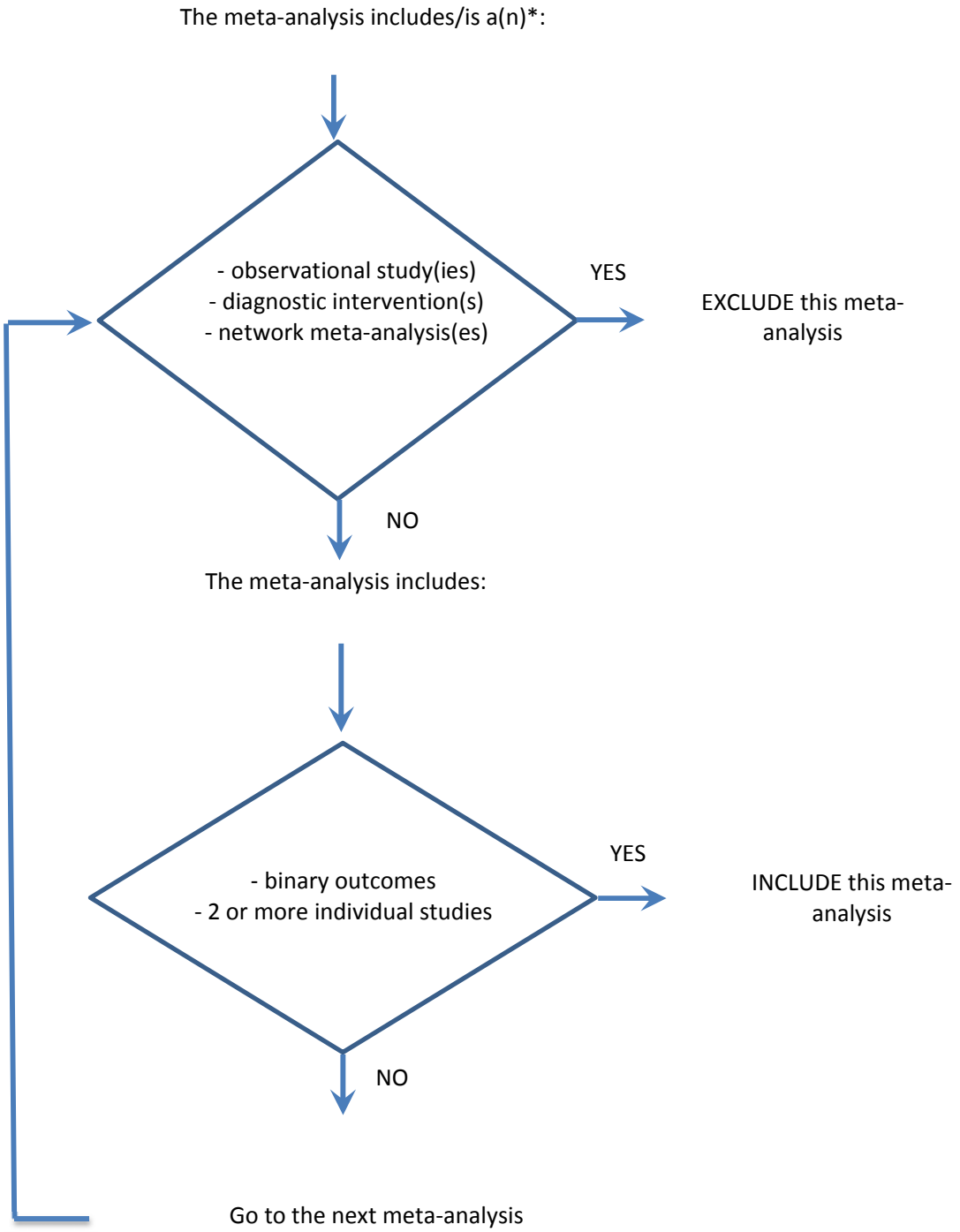
23. Hutton B, Salanti G, Chaimani A, et al. The quality of reporting methods and results in network meta-analyses: an overview of reviews and suggestions for improvement. PLoS One 2014;9:e92508.

24. Chambers JD, Naci H, Wouters OJ, et al. An assessment of the methodological quality of published network meta-analyses: a systematic review. PLoS One 2015;10:e0121715.

25. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986;1(8476):307-10.

26. Fanshawe TR, Shaw LF, Spence GT. A large-scale assessment of temporal trends in meta-analyses using systematic review reports from the Cochrane Library. Res Synth Methods 2017;8(4):404-415.

Chapter 4. Appendix



* Main outcome comparison as described by the authors.
If the authors did not describe a main outcome, the first comparison that appeared in the meta-analysis was used.

Figure A4.1. Algorithm for meta-analysis selection (main comparison).

Chapter 5. Impact of follow-up duration on the control event rate

Overview

Research objectives: To determine the impact of follow-up duration on the control event rate (CER) in randomized controlled trials (RCTs).

Background: The CER value is necessary to estimate optimum information size (OIS). Chapter 4 shows that there is an inverse association between CER and OIS levels. Theoretically CER should increase with study duration, which should result in longer RCTs requiring smaller numbers of participants. However, no empirical research has demonstrated this effect in RCTs summarized as part of systematic reviews.

Methods: A meta-epidemiological study with 2 sets of meta-analyses was defined to evaluate the impact of CER. The study combined data from 236 RCTs included in 34 meta-analyses. The impact of follow-up duration on CER was evaluated for 2 equal-sized sets of meta-analyses: Cochrane reviews from 2012, retrieved from the Archie Database; and non-Cochrane reviews published between 2012 and 2014 in the top 5 general medical journals, retrieved from Medline. Studies were included if they reported mortality outcome data from 2 or more individual RCTs. Meta-analyses that included observational or diagnostic intervention studies, network meta-analyses and meta-analyses with zero events were excluded. Descriptive analyses were performed for disease class and type of intervention for included systematic reviews and individual RCTs. Standardized measures for CER and follow-up duration were created based on within-review estimates with a view to reducing clinical and methodological variability. Multilevel linear regression analyses were run to quantify the

strength of the relationship between CER and follow-up duration and to explore other relationships of interest such as sample size and effect size (ES).

Results: 34 meta-analyses were included — 17 (50%) each representing Cochrane and non-Cochrane reviews — that included a total of 236 RCTs. Noncommunicable diseases and pharmacological interventions represented 79.5% and 73.5% of the sample, respectively. Mean follow-up duration was 13.1 months for Cochrane RCTs and 10.7 months for non-Cochrane RCTs. Mean CER was considerably higher for the Cochrane group (22.5) than for the non-Cochrane group (11.9). Median CER for the Cochrane group (12.7) was double that for the non-Cochrane group (6.3). The multilevel linear regression analysis showed no clear correlation between standardized follow-up duration and standardized CER; what correlation did exist tended to be positive, with larger CER values tending to be associated with longer follow-up duration. No correlation was observed between standardized CER and standardized sample size or standardized ES.

Conclusion: The findings do not point to any clear relationship between follow-up duration and CER within the context of RCTs included in systematic reviews with meta-analyses.

5.1. Introduction

Various parameters, including effect size (ES), homogeneity and baseline risk (usually measured as the control event rate (CER)), are commonly used in the calculation of optimal information size (OIS), yet there is little consensus on how these should be estimated. In addition to being related to ES, CER also depends on duration of follow-up, which typically varies across studies (1).

Baseline risk has been defined as the rate of occurrence of the event of interest when the standard treatment is used (2). It needs to be taken into account in calculating OIS, just as is done when

calculating sample size for a standard prospective randomized controlled trial (RCT). As mentioned above, baseline risk is generally measured in terms of CER, i.e. as the observed risk of an event in the control group, whether death, cure, infection, infarction, etc (3). Baseline risk is also related to the absolute ES of the intervention. Relative measures of ES — relative risk (RR), odds ratio (OR) or hazard ratio (HR) — tend to be consistent across risk groups. As illustrated in the linear regression plot in Figure 5.1, there is a weak negative linear relationship between the number of included patients in a meta-analysis and CER.

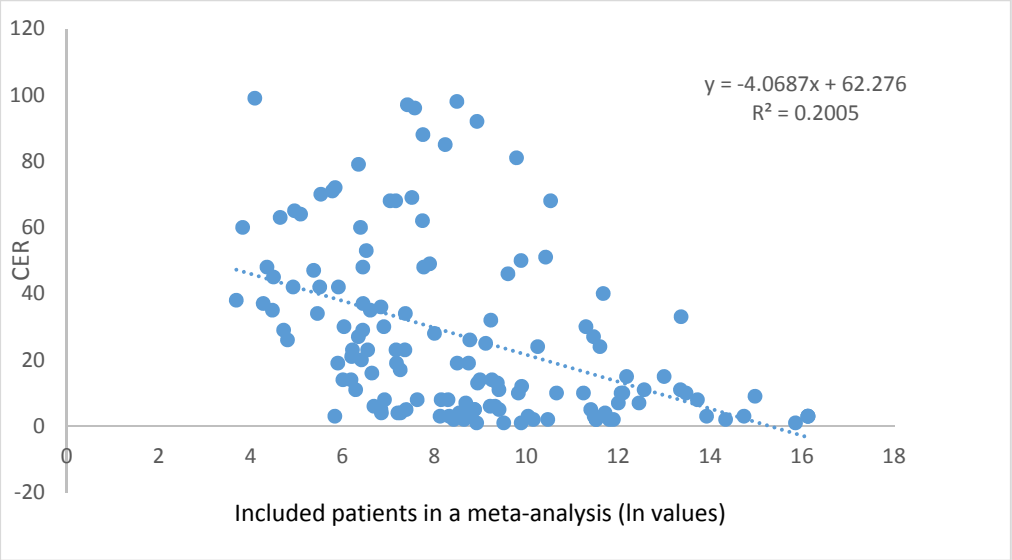


Figure 5.1. Association between CER and the number of patients in a meta-analysis.
The graph was constructed using data from 137 meta-analyses (those analysed in Chapter 4). CER was calculated as the median of the number of events in the control group for all RCTs in the meta-analysis and included patients reflects the total number of patients in all the RCTs in the meta-analysis.

As happens with the variability of effect estimates for an intervention, supporting estimates of baseline risk can vary, not only by intervention but also by population. The Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group, in an article published in 2011 (4), described several elements that could compromise baseline risk calculations, as follows: risk of bias related mainly to the study design (RCTs vs observational studies); publication bias (studies with a large number of events are less likely to be published); inconsistency (relevant

differences in effect estimates or the number of events between different studies evaluating the same intervention); and indirectness (when baseline risk estimates are derived from significantly different populations). Although there is obviously a need to ensure confidence in baseline risk estimates, there is still controversy on how exactly to make these estimates (5). The GRADE Working Group guidelines (6), for instance, indicate that baseline risk should ideally come from well-designed observational studies. This, however, is not always feasible, as indicated by the GRADE Working Group (6), which proposes a solution as follows:

> If high-quality observational studies are not available, we suggest using the median risk (rather than the weighted average) among the control groups in the included studies or, if it is available, the control group risk from a single trial with far larger sample size than other available trials. If there is important variation in control group risks, authors should consider presenting a range of risks within that observed in the included studies (that is, present a range of baseline risks) (GRADE Guideline 12).

The rationale for using the median rather than the weighted average is that baseline risk is likely to differ across settings. Thus, if the weighted average for the largest RCT were used, the resulting value might be an extreme one. Since no particular setting should be disproportionally represented, in choosing a single baseline risk, this should be the most typical for all settings.

Hayden et al. (7), referring to different approaches to estimating baseline risk, concluded that, overall and across all scenarios, a median estimate of baseline risk results in small and stable bias. Since all other methods tend to overestimate this risk when small (<0.5) or to underestimate it when large (>0.5), the median estimate seems to be the most satisfactory method for estimating baseline risk, and so is the main measure that will be used in this chapter.

Given that it is not known how CER behaves depending on RCT follow-up duration, the aim of this chapter was to analyse what impact, if any, RCT follow-up duration has on CER, first by examining median CER values and then by exploring further using regression analyses based on standardized within-review measures of CER and follow-up duration.

5.2. Methods

5.2.1. Literature search

To evaluate the impact of follow-up duration on CER, a meta-epidemiological study was conducted consisting of 34 systematic reviews with meta-analyses combining data from 236 RCTs and divided into 2 sets (Cochrane and non-Cochrane reviews). For definitions of the main concepts referred to in what follows, see Section 5.2.4.

Meta-analyses on noncommunicable and communicable diseases were included if they reported mortality outcome data from 2 or more individual RCTs. These disease classes were considered in order to analyse whether follow-up duration — which tends to be longer for noncommunicable diseases — had an impact on CER that differed according to disease class. Meta-analyses were also analysed according to pharmacological or nonpharmacological interventions in order to determine whether the intervention type had any impact on CER.

Excluded were meta-analyses referring to observational studies, diagnostic intervention studies and network meta-analyses, as also were meta-analyses with zero events.

The Archie Database (http://archie.cochrane.org), which contains all Cochrane reviews, was searched electronically for Cochrane reviews published in 2012 (the most recent year for which data were available at the time of commencing this research). To identify non-Cochrane reviews published in the top 5 general medical journals, Medline (PubMed) was electronically searched

using the following search strategy: ("BMJ"[Journal] OR "Ann Intern Med"[Journal] OR "JAMA"[Journal] OR "Lancet"[Journal] OR "N Engl J Med"[Journal]) AND (systematic review [ti] OR meta-analysis [pt] OR meta-analysis [ti]). Although the search was initially confined to 2012, given the small number of retrieved meta-analyses fulfilling the inclusion criteria, the search was subsequently broadened to include 2103 and 2104 so as to retrieve a similar number of meta-analyses as included in the Cochrane set.

5.2.2. Data extraction

Titles and abstracts of the retrieved studies were initially screened (by JMGA) according to the inclusion criteria described above to identify relevant meta-analyses. Full texts for meta-analyses that met the inclusion criteria and full texts of the RCTs included in those meta-analyses were further assessed for eligibility. Data were extracted into a Microsoft Excel spreadsheet by 2 researchers (JMGA and NP) and a consensus was reached with a third researcher (RP) in cases of disagreement.

5.2.3. Data analysis

Data for the analysed variables were recorded in tabulated form in a customized Microsoft Excel spreadsheet and frequencies, mean, standard deviation (SD), median and interquartile range were calculated. The results were also depicted as boxplots.

The analysis was conducted as follows: for the 34 meta-analyses overall; for the 236 RCTs included in those meta-analyses; for Cochrane and non-Cochrane reviews; for communicable/noncommunicable diseases; and for pharmacological/nonpharmacological interventions.

For the 34 meta-analyses, reported for Cochrane and non-Cochrane reviews are numbers and percentages broken down by noncommunicable/communicable diseases and by

pharmacological/nonpharmacological interventions, and means and SDs for events, patients, follow-up duration, CER and ES.

For the 236 individual RCTs included in Cochrane and non-Cochrane reviews, reported are numbers and percentages for noncommunicable/communicable diseases and pharmacological/nonpharmacological interventions, and means, SDs, medians and interquartile ranges for intervention and control group patients and events and, for follow-up duration (when available), CER and ES broken down by communicable/noncommunicable diseases and pharmacological/nonpharmacological interventions.

Also reported for the individual RCTs included in Cochrane and non-Cochrane reviews are numbers and percentages for follow-up in months stratified by duration $\leq$1 month or >1 month and numbers and percentages for truncated RCTs (tRCTs).

The above analyses were carried out using R software.

Simple linear regression analysis, performed using R, was used to model the relationship between the dependent variable (CER) and the explanatory variable (follow-up duration) based on the assumption of a linear relationship. This regression was performed with standardized variables, obtained by taking the value for sample size, ES, CER and follow-up duration for each RCT and dividing the value by the corresponding median for all RCTs.

Given that great CER variability was observed within reviews, to remove the cluster effect, a multilevel linear regression model was used to examine correlation between observations within clusters (i.e. within the same meta-analysis). As well as removing the cluster effect, multilevel linear regression model also enabled associations between dependent (standardized CER) and independent (explanatory) variables to be detected across different reviews.

Given that a fixed-effects model would be incapable of separating out group effects from covariate effects at the group level, a random-effects model was used to estimate the effect of covariates at the group level (e.g. according to Cochrane vs non-Cochrane groups). This random-effects multilevel linear regression model was used to evaluate the association between standardized CER and standardized follow-up duration, with Cochrane/non-Cochrane reviews and communicable/noncommunicable diseases as the confounding factors.

Multilevel analyses were performed with 224 RCTs nested within 33 systematic reviews (clusters). The multilevel linear regression analyses were carried out using STATA V.14 software.

5.2.4. Definitions

The variables compiled for the RCTs and meta-analyses are listed and defined as follows:

A. META-ANALYSES

Variable	Definition
ID-review	CD number for Cochrane reviews and author-journal-year for non-Cochrane reviews.
Disease class	Communicable and noncommunicable diseases (WHO definitions). Noncommunicable diseases (mainly cardiovascular diseases, cancers, chronic respiratory diseases and diabetes) are not transmitted from person to person, are of long duration and generally progress slowly. Communicable diseases are spread from one person to another through various means (contact with bodily fluids, airborne viruses, insect bites).
Type of intervention	Pharmacological and nonpharmacological interventions intended to improve health or well-being. Pharmacological interventions involve the use of a drug or medicine (e.g. chemotherapy for breast cancer). Nonpharmacological interventions include interventions classified as medical devices, surgical, complex, resources and infrastructure, behavioural, psychological, physical, complementary, educational, radiotherapy, vaccines, cellular and gene, and screening. Source: Turner et al. 2012 (8).
Year of publication	Year of publication of the meta-analysis (2012 for Cochrane reviews and 2012-2014 for non-Cochrane reviews).
Median CER	Median of the proportion of events in the RCTs included in each meta-analysis.
Pooled ES	Pooled relative risk (RR) obtained using either a random-effects or fixed-effects model. If the pooled ES was reported as OR or another measure, it was transformed to RR.
Standardized CER	CER for individual RCTs divided by the median CER for all RCTs.
Standardized follow-up duration	Follow-up duration for individual RCTs divided by the median follow-up duration for all RCTs, resulting in a person-year follow-up value.
Standardized ES	ES for the individual RCTs divided by the median ES for all RCTs.
Standardized sample size	Number of patients for individual RCTs divided by the median number of patients for all RCTs.

B. RCTs (one record for each trial included in each meta-analysis)

Variable	Definition
ID-RCT (plus ID-review)	RCT first author-year.
Year of publication	Year of publication.
Intervention group events	Number as reported in the meta-analysis.
Control group events	Number as reported in the meta-analysis.
Intervention group patients	Number as reported in the meta-analysis.
Control group patients	Number as reported in the meta-analysis.
Follow-up duration	Recorded in months, and if <1 month, recorded as follows: 3 weeks = 0.75 months; 2 weeks = 0.50 months; 1 week = 0.25 months; hours (1 hour, 24 hours, 72 hours, etc) = 0.25 months.
CER	Proportion of events in the control group.
ES	Relative risk (RR). If reported as OR or another measure, it was transformed to RR.
tRCT	Yes/no. No, if any of the following words/procedures suggesting early termination (truncation) were encountered. Words: halted, closed, closure, terminated, stopped, early, prematurely. Procedures: data and safety monitoring board (DSMB), interim analysis, and stopping rules or boundaries and their specific names (O'Brien-Fleming, Lan-DeMets, spending function(s), Haybittle-Peto boundary, Pocock boundary, triangular boundaries, prespecified p-value, spending function not reported, other boundaries). Source: Adapted from Montori et al. 2005 (9).

5.3. Results

5.3.1. Literature search

The initial electronic search strategy identified 446 potentially relevant meta-analyses in Medline (PubMed) and 1943 additional studies in the Cochrane Archie Database. After duplicates were removed, 1018 records remained for initial screening of titles and abstracts, and a further 945 were excluded. A total of 73 meta-analyses were retrieved for full-text review, of which 34 met the inclusion criteria: 17 (50%) Cochrane reviews and 17 (50%) non-Cochrane reviews (Figure 5.2).

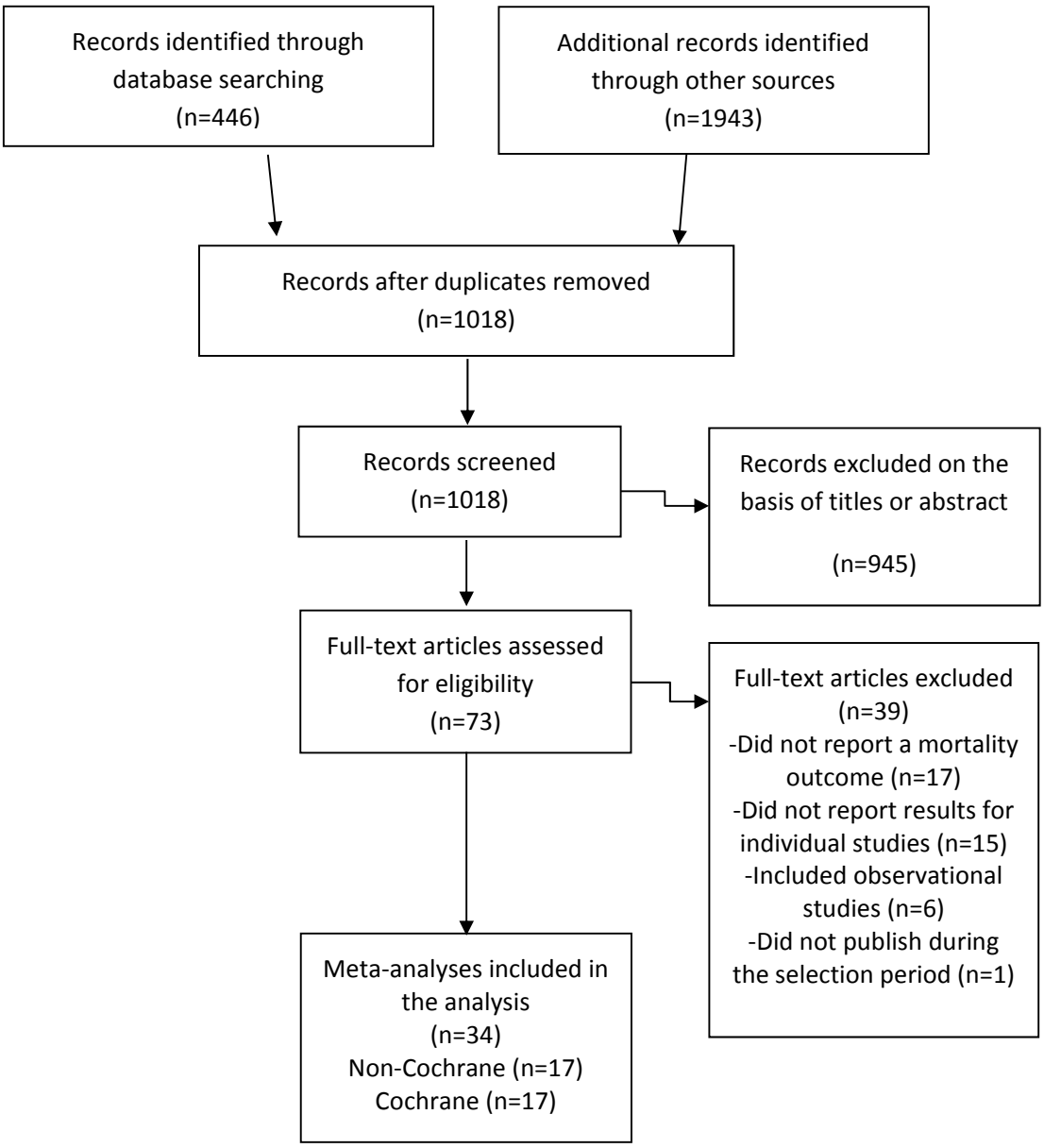


Figure 5.2. Flowchart of search results (PRISMA model).

The 34 meta-analyses included a total of 236 RCTs: mean (SD) 4.7 (3.1) for Cochrane reviews and 9.2 (5.6) for non-Cochrane reviews. Total included patients were mean (SD; median) 429.8 (1137.3; 88) for Cochrane reviews and 1142.2 (2859.2; 200) for non-Cochrane reviews.

The 34 included meta-analyses are listed in Table A5.1 in the Appendix.

5.3.2. Descriptive results

Table 5.1 shows results for disease class, intervention type, number of events, number of patients, follow-up duration, ES and CER for the 34 included meta-analyses by review type (Cochrane vs non-Cochrane). The majority of the reviews focused on noncommunicable diseases (79.4%), in similar proportions for Cochrane reviews (76.5%) and non-Cochrane reviews (82.4%). Likewise, pharmacological interventions were the focus in most reviews (73.5%), but noticeably more so in non-Cochrane reviews (88.2%) than Cochrane reviews (58.8%). The mean number of intervention and control group events was much greater in the non-Cochrane group (43.7 and 59.4, respectively) than in the Cochrane group (29.4 and 30.4, respectively). Mean sample size for non-Cochrane reviews was around 3 times greater than for Cochrane reviews (1130 vs 421 patients). Mean follow-up tended to be longer in Cochrane reviews compared to non-Cochrane reviews (16.6 vs 13.5 months). Finally, mean ES was considerably smaller for Cochrane reviews (0.82) than for non-Cochrane reviews (1.2), whereas mean CER for Cochrane reviews (22.5) was around double that for non-Cochrane reviews (11.9).

Table 5.2 shows numbers and percentages for noncommunicable/communicable diseases and pharmacological/nonpharmacological interventions for the 236 included RCTs by review type (Cochrane vs non-Cochrane). Noncommunicable diseases were the focus of the majority of RCTs (80.9%), but substantially less so in the Cochrane group vs the non-Cochrane group (63.8% vs 89.7%). Pharmacological interventions were also the focus in the majority of RCTs (79.7%), although slightly less so in the Cochrane group vs the non-Cochrane group (75% vs 82.1%).

Table 5.3 summarizes data on follow-up duration and tRCTs. Follow-up lasted >1 month in most RCTs (61.1%) — this percentage was lower in Cochrane reviews (52.2%) and higher in non-Cochrane reviews (65.1%). A higher proportion of non-Cochrane RCTs compared to Cochrane RCTs

had follow-up >1 month (65.1% vs 52.2%). There were just 12 trRCTs: 7 (9%) in the Cochrane group and 5 (3.2%) in the non-Cochrane group.

Table 5.1. Main characteristics of the 34 included meta-analyses.

	Cochrane reviews (n=17)	Non-Cochrane reviews (n=17)	All reviews (n=34)
Disease class, % (n)			
Noncommunicable	76.5 (13)	82.4 (14)	79.4 (27)
Communicable	23.5 (4)	17.6 (3)	20.6 (7)
Intervention type, % (n)			
Pharmacological	58.8 (10)	88.2 (15)	73.5 (25)
Nonpharmacological	41.2 (7)	11.8 (2)	26.5 (9)
Events and patients, mean (SD)			
Intervention group events	29.4 (71.7)	43.7 (132.6)	38.9 (115.7)
Control group events	30.4 (74.6)	59.4 (233.7)	49.6 (195.2)
Patients	421.3 (1127.5)	1130 (2845.7)	885.3 (2410.3)
Follow-up duration, ES and CER, mean (SD)			
Follow-up (months)	16.6 (17.9)	13.5 (17.6)	15 (17.5)
ES (RR)	0.82 (0.22)	1.2 (1.3)	1 (1.01)
CER (%)	22.5 (18.7)	11.9 (13.0)	17.2 (16.8)

Table 5.2. Disease classes and intervention types.

	Cochrane RCTs (n=80)	Non-Cochrane RCTs (n=156)	All RCTs (n=236)
Disease class, % (n)			
Noncommunicable	63.8 (51)	89.7 (140)	80.9 (191)
Communicable	36.2 (29)	10.3 (16)	19.1 (45)
Intervention type, % (n)			
Pharmacological	75 (60)	82.1 (128)	79.7 (188)
Nonpharmacological	25 (20)	17.9 (28)	20.3 (48)

Table 5.3. Follow-up duration and truncated RCTs.

	% (n)	% (n)	% (n)
Follow-up (months)	**Cochrane RCTs (n=69)**	**Non-Cochrane RCTs (n=152)**	**All RCTs (n=221)***
⩽ 1 month	47.8 (33)	34.9 (53)	38.9 (86)
> 1 month	52.2 (36)	65.1 (99)	61.1 (135)
tRCTs	**Cochrane RCTs (n=80)**	**Non-Cochrane RCTs (n=156)**	**All RCTs (n=236)**
Yes	9 (7)	3.2 (5)	5.1 (12)
No	91 (73)	96.8 (151)	94.9 (224)

* Information available only for 221 RCTs.

Tables 5.4-5.10 show mean, SD, median and interquartile range data for the 236 included RCTs, by review type (Cochrane/non-Cochrane), by disease class (noncommunicable/communicable) and by intervention type (pharmacological/ nonpharmacological).

Table 5.4 shows values for intervention group events. The mean number of events overall was substantially larger in the non-Cochrane group vs the Cochrane group (43.7 vs 29.4). For noncommunicable diseases, the mean event rate was similar for both groups (40.3 vs 44.8), whereas for pharmacological interventions, the mean event rate was substantially lower for the Cochrane group vs the non-Cochrane group (30.4 vs 49.4).

Table 5.4. Intervention group events.

Total	**Cochrane RCTs (n=80)**	**Non-Cochrane RCTs (n=156)**	**All RCTs (n=236)**
Mean (SD)	29.4 (71.7)	43.7 (132.6)	38.8 (115.6)
Median	4	6	5
Interquartile range	2-15.7	1-20.7	1-18
Noncommunicable	**Cochrane RCTs (n=51)**	**Non-Cochrane RCTs (n=140)**	**All RCTs (n=191)**
Mean (SD)	40.3 (87.4)	44.8 (137.8)	43.6 (126.1)
Median	4	6	6
Interquartile range	2-26.5	1-21	1-21
Communicable	**Cochrane RCTs (n=29)**	**Non-Cochrane RCTs (n=16)**	**All RCTs (n=45)**
Mean (SD)	10.3 (16.7)	34.1 (75.3)	18.8 (47.4)
Median	4	5	5
Interquartile range	3-10	3-7.7	3-10
Pharmacological	**Cochrane RCTs**	**Non-Cochrane RCTs**	**All RCTs**

	(n=60)	(n=128)	(n=188)
Mean (SD)	30.4 (79.3)	49.4 (145.6)	43.3 (128.3)
Median	4	4.5	4
Interquartile range	2-14.2	1-21	1-17.2
Nonpharmacological	**Cochrane RCTs (n=20)**	**Non-Cochrane RCTs (n=28)**	**All RCTs (n=48)**
Mean (SD)	26.5 (42.6)	17.8 (21.4)	21.4 (31.9)
Median	6.5	11.5	9.5
Interquartile range	3-32.7	7-18.5	5.7-20.2

Table 5.5 shows values for intervention group patients. The mean number of patients was around double in the non-Cochrane group vs the Cochrane group overall (696.1 vs 258.7), for noncommunicable diseases (759.1 vs 348.9) and for pharmacological interventions (815.6 vs 314.3).

Table 5.5. Intervention group patients.

Total	**Cochrane RCTs (n=80)**	**Non-Cochrane RCTs (n=156)**	**All RCTs (n=236)**
Mean (SD)	258.7 (623.1)	696.1 (1829.1)	547.8 (1542.8)
Median	56.5	115.5	100
Interquartile range	30-169.2	48.7-641.7	34.7-340.5
Noncommunicable	**Cochrane RCTs (n=51)**	**Non-Cochrane RCTs (n=140)**	**All RCTs (n=191)**
Mean (SD)	348.9 (765.7)	759.1 (1919.7)	649.6 (1698)
Median	63	126	104
Interquartile range	31-187	59-714.7	45.5-510
Communicable	**Cochrane RCTs (n=29)**	**Non-Cochrane RCTs (n=16)**	**All RCTs (n=45)**
Mean (SD)	100.1 (94.3)	144.7 (246.6)	116 (163.8)
Median	56	20.5	45
Interquartile range	30-169	13.7-167	25-169
Pharmacological	**Cochrane RCTs (n=60)**	**Non-Cochrane RCTs (n=128)**	**All RCTs (n=188)**
Mean (SD)	314.3 (708.9)	815.6 (1999.8)	655.6 (1711.6)
Median	72.5	129	104
Interquartile range	30-198.2	34-1011.2	30.7-542.2
Nonpharmacological	**Cochrane RCTs (n=20)**	**Non-Cochrane RCTs (n=28)**	**All RCTs (n=48)**
Mean (SD)	91.6 (130)	149.5 (135)	125.6 (134.6)
Median	40.5	104	85
Interquartile range	28-70.7	72.5-142.5	44-142

Table 5.6 shows values for control group events. The mean number of events overall was much smaller in the Cochrane group vs the non-Cochrane group (30.4 vs 59.4). For noncommunicable diseases, means for both groups were reasonably close to the overall mean (56.9), although higher for the non-Cochrane group vs the Cochrane group (62.4 vs 42.1). For pharmacological interventions, the mean for the non-Cochrane group was around double that of the Cochrane group (67.6 vs 31.1). Table 5.7 shows values for control group patients. The mean number of patients was more than double in the non-Cochrane group vs the Cochrane group overall (696 vs 258.8) and this difference was similar for noncommunicable diseases (667.7 vs 363). For pharmacological interventions, the mean for the non-Cochrane group was nearly triple that of the Cochrane group (815.6 vs 314.3).

Table 5.6. Control group events.

Total	**Cochrane RCTs (n=80)**	**Non-Cochrane RCTs (n=156)**	**All RCTs (n=236)**
Mean (SD)	30.4 (74.6)	59.4 (233.8)	49.6 (195.2)
Median	4	7	6
Interquartile range	1-15.7	1-24	1-22
Noncommunicable	**Cochrane RCTs (n=51)**	**Non-Cochrane RCTs (n=140)**	**All RCTs (n=191)**
Mean (SD)	42.1 (90.6)	62.4 (245.7)	56.9 (215.5)
Median	5	7	6
Interquartile range	2-26	1-24	1-24
Communicable	**Cochrane RCTs (n=29)**	**Non-Cochrane RCTs (n=16)**	**All RCTs (n=45)**
Mean (SD)	9.9 (19)	32.8 (65.5)	18 (42.6)
Median	3	8.5	6
Interquartile range	1-9	5.7-12	2-11
Pharmacological	**Cochrane RCTs (n=60)**	**Non-Cochrane RCTs (n=128)**	**All RCTs (n=188)**
Mean (SD)	31.1 (82.6)	67.6 (257.3)	55.9 (217.7)
Median	3	6	4.5
Interquartile range	1-13	0-24	1-21.2
Nonpharmacological	**Cochrane RCTs (n=20)**	**Non-Cochrane RCTs (n=28)**	**All RCTs (n=48)**
Mean (SD)	28.3 (44.2)	21.7 (25.3)	24.5 (34.2)
Median	6	14.5	11.5
Interquartile range	3.7-36.2	8.7-20.2	6-24.5

Table 5.7. Control group patients.

Total	Cochrane RCTs (n=80)	Non-Cochrane RCTs (n=156)	All RCTs (n=236)
Mean (SD)	258.8 (623.5)	696 (1289.1)	498.7 (1096.2)
Median	56.5	115.5	101.5
Interquartile range	30-169.2	48.7-641.7	37-380
Noncommunicable	**Cochrane RCTs (n=51)**	**Non-Cochrane RCTs (n=140)**	**All RCTs (n=191)**
Mean (SD)	363 (762.3)	667.7 (1316)	586. (1199.3)
Median	75	136	115
Interquartile range	32-202.5	58.2-683.5	44.5-530
Communicable	**Cochrane RCTs (n=29)**	**Non-Cochrane RCTs (n=16)**	**All RCTs (n=45)**
Mean (SD)	115.1 (131.9)	148.6 (252.3)	127 (181.7)
Median	56	20	50
Interquartile range	30-167	12.7-168.5	25-167
Pharmacological	**Cochrane RCTs (n=60)**	**Non-Cochrane RCTs (n=128)**	**All RCTs (n=188)**
Mean (SD)	314.3 (708.9)	815.6 (1999.8)	655.6 (1711.5)
Median	72.5	129	104
Interquartile range	30-198.2	34-1011.2	30.7-542.2
Nonpharmacological	**Cochrane RCTs (n=20)**	**Non-Cochrane RCTs (n=28)**	**All RCTs (n=48)**
Mean (SD)	92.1 (130)	149.5 (135)	125.6 (134.6)
Median	40.5	104	85
Interquartile range	28-70.7	72.5-142.5	44-142

Table 5.8 shows values for follow-up duration. Mean follow-up duration was longer by about 2 months for the Cochrane group vs the non-Cochrane group. For noncommunicable diseases, the mean for the Cochrane group was almost double that of the non-Cochrane group (19.7 vs 11.6) and, for pharmacological interventions, the mean for the Cochrane group was greater than that of the non-Cochrane group (15.2 vs 11.8).

Table 5.9 shows values for ES. Mean ES overall was very similar for the Cochrane group compared to the non-Cochrane group (1.3 vs 1.2). For noncommunicable diseases, the mean for the Cochrane group was smaller than that of the non-Cochrane group (1.0 vs 1.2) and, for pharmacological

interventions, the mean for the Cochrane group was higher than that of the non-Cochrane group (1.4 vs 1.2).

Table 5.8. Follow-up duration for intervention groups.

Total	Cochrane RCTs (n=72)	Non-Cochrane RCTs (n=155)	All RCTs (n=227) *
Mean (SD)	13.1 (21.2)	10.7 (16.8)	11.5 (18.3)
Median	1.4	4	3
Interquartile range	1-15.8	1-12	1-12
Noncommunicable	**Cochrane RCTs (n=45)**	**Non-Cochrane RCTs (n=139)**	**All RCTs (n=184)**
Mean (SD)	19.7 (24.4)	11.6 (17.5)	13.6 (19.6)
Median	6	6	6
Interquartile range	1-36	1-12	1-13
Communicable	**Cochrane RCTs (n=27)**	**Non-Cochrane RCTs (n=16)**	**All RCTs (n=43)**
Mean (SD)	2.2 (4.9)	2.6 (4)	2.4 (4.5)
Median	1	0.2	1
Interquartile range	0.8-1	0.2-3	0.2-1.4
Pharmacological	**Cochrane RCTs (n=54)**	**Non-Cochrane RCTs (n=109)**	**All RCTs (n=181)**
Mean (SD)	15.2 (22.6)	11.8 (18.3)	12.8 (19.7)
Median	2.2	4	3
Interquartile range	1-23	1-12	1-12
Nonpharmacological	**Cochrane RCTs (n=18)**	**Non-Cochrane RCTs (n=28)**	**All RCTs (n=46)**
Mean (SD)	7.1 (15.1)	5.8 (3.5)	6.3 (9.7)
Median	1	6	3
Interquartile range	0.5-3	3-6.5	1-6

* Information available only for 227 RCTs.

Table 5.9. ES values for intervention groups.

Total	Cochrane RCTs (n=76)	Non-Cochrane RCTs (n=137)	All RCTs (n=213) *
Mean (SD)	1.3 (1.2)	1.2 (1.5)	1.2 (1.4)
Median	1	0.8	0.9
Interquartile range	0.6-1.3	0.5-1.2	0.5-1.2
Noncommunicable	**Cochrane RCTs (n=48)**	**Non-Cochrane RCTs (n=121)**	**All RCTs (n=169)**
Mean (SD)	1 (0.8)	1.2 (1.5)	1.2 (1.4)
Median	0.9	0.8	0.9
Interquartile range	0.5-1.1	0.5-1.1	0.5-1.1
Communicable	**Cochrane RCTs (n=28)**	**Non-Cochrane RCTs (n=16)**	**All RCTs (n=44)**
Mean (SD)	1.7 (1.7)	0.8 (0.3)	1.4 (1.4)
Median	1.2	0.9	1
Interquartile range	0.8-1.9	0.5-1	0.7-1.3
Pharmacological	**Cochrane RCTs (n=56)**	**Non-Cochrane RCTs (n=109)**	**All RCTs (n=165)**
Mean (SD)	1.4 (1.3)	1.2 (1.6)	1.3 (1.5)
Median	1	0.9	1
Interquartile range	0.7-1.5	0.5-1.1	0.5-1.2
Nonpharmacological	**Cochrane RCTs (n=20)**	**Non-Cochrane RCTs (n=28)**	**All RCTs (n=48)**
Mean (SD)	0.8 (0.4)	0.9 (0.6)	0.8 (0.5)
Median	0.9	0.7	0.7
Interquartile range	0.4-1	0.5-1.1	0.4-1

* Information available only for 213 RCTs.

Table 5.10 shows values for CER. Mean CER overall was considerably higher for the Cochrane group compared to the non-Cochrane group (19 vs 12.2). For noncommunicable diseases the mean for the Cochrane group was more than double that of the non-Cochrane group (19.6 vs 8.7) and, for pharmacological interventions, the mean for the Cochrane group was also higher than that of the non-Cochrane group (14.8 vs 11.8). Given this chapter's focus on median CER — guided by the conclusion of Hayden et al. (7) that a median estimate of baseline risk results in small and stable bias overall and across all scenarios — highlighted are median CER values. Thus, comparing the Cochrane group vs the non-Cochrane group, median CER was double (12.7 vs 6.3) overall, was more

than double (14 vs 5.8) for noncommunicable diseases and was slightly higher (7 vs 5) for pharmacological interventions.

Table 5.10. CER values for intervention groups.

Total	Cochrane RCTs (n=69)	Non-Cochrane RCTs (n=125)	All RCTs (n=194) *
Mean (SD)	19 (20.7)	12.2 (14.3)	14.6 (17.1)
Median	12.7	6.3	7.4
Interquartile range	2.5-29	2.8-17.1	2.8-19.6
Noncommunicable	**Cochrane RCTs (n=45)**	**Non-Cochrane RCTs (n=109)**	**All RCTs (n=154)**
Mean (SD)	19.6 (20.6)	8.7 (8.9)	11.9 (14.3)
Median	14	5.8	6.5
Interquartile range	4.1-29	2.3-13.2	14.3
Communicable	**Cochrane RCTs (n=24)**	**Non-Cochrane RCTs (n=16)**	**All RCTs (n=40)**
Mean (SD)	17.7 (21.2)	36.2 (20.4)	25.1 (22.6)
Median	6.1	37.9	21.1
Interquartile range	1.9-30.2	22.2-54	3.2-39.1
Pharmacological	**Cochrane RCTs (n=49)**	**Non-Cochrane RCTs (n=97)**	**All RCTs (n=146)**
Mean (SD)	14.8 (17.2)	11.8 (15.9)	12.8 (16.4)
Median	7	5	5.2
Interquartile range	2-21	2-16	2-17.9
Nonpharmacological	**Cochrane RCTs (n=20)**	**Non-Cochrane RCTs (n=28)**	**All RCTs (n=48)**
Mean (SD)	29.2 (25.1)	13.4 (6.2)	20 (18.4)
Median	21.1	12.7	14.2
Interquartile range	12.9-36	8.4-17.2	9.3-22.1

* Information available only for 194 RCTs.

Figures 5.3–5.10 depict boxplots for median CER values and median follow-up durations for Cochrane and non-Cochrane reviews and for noncommunicable and communicable diseases.

Figures 5.3 and 5.4 show median CER distributions for the 17 Cochrane reviews (ranging from 1.9 to 63.2) and 17 non-Cochrane reviews (ranging from 0.39 to 40 months), respectively, reflecting great variability in follow-up duration for the RCTs in each review. Figures 5.5 and 5.6 show median CER distributions for the 27 noncommunicable disease reviews (ranging from 0.39 to 63.2) and 7

communicable disease reviews (ranging from 1.9 to 51 months), respectively, again reflecting great variability in follow-up duration for the disease classes.

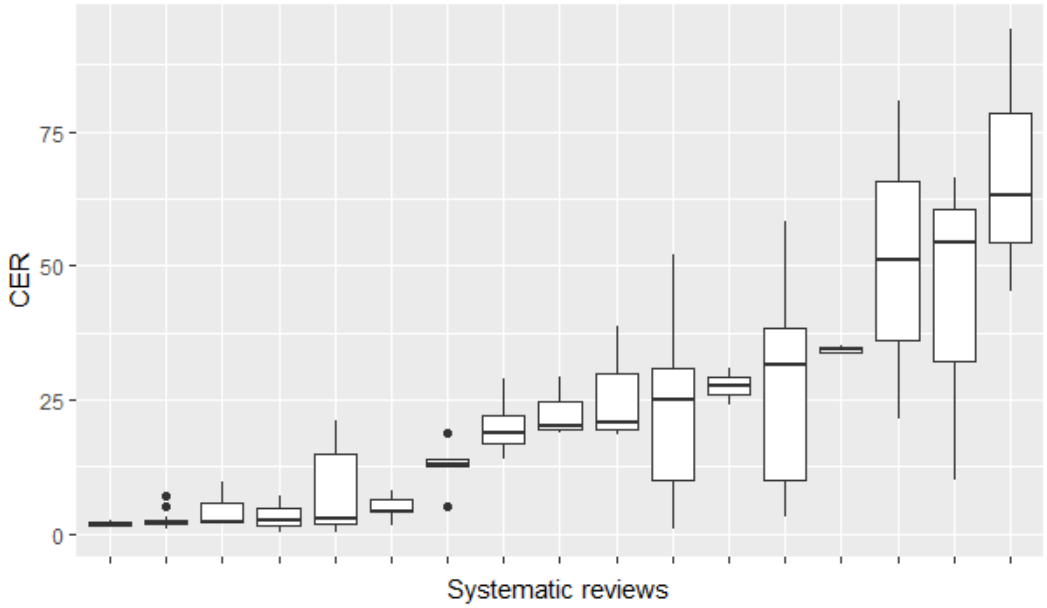


Figure 5.3. Boxplot for CER for Cochrane reviews.
Ordered (left to right) from smallest to largest median CER values.

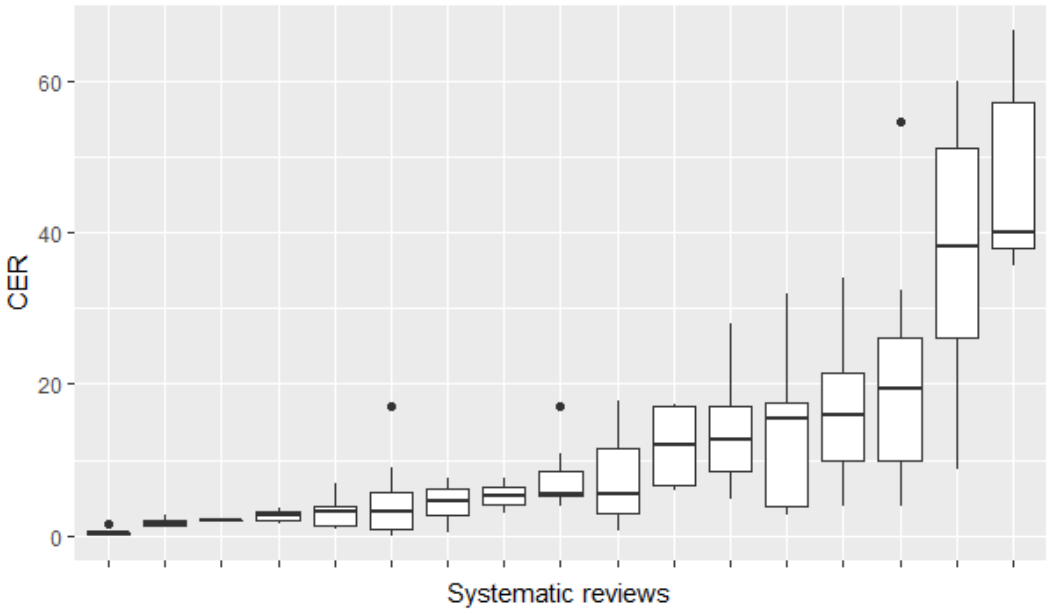


Figure 5.4. Boxplot for CER for non-Cochrane reviews.
Ordered (left to right) from smallest to largest median CER values.

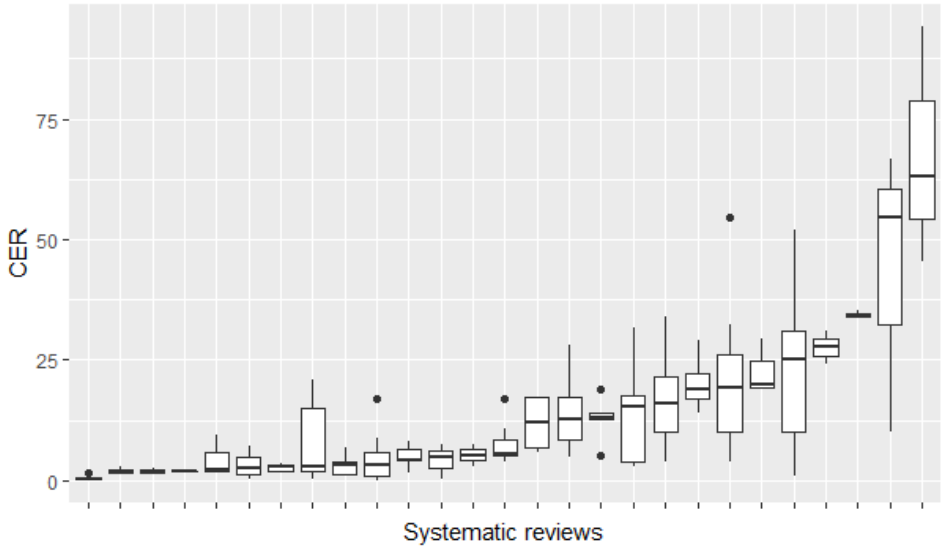


Figure 5.5. Boxplot for CER for noncommunicable disease reviews.
Ordered (left to right) from smallest to largest median CER values.

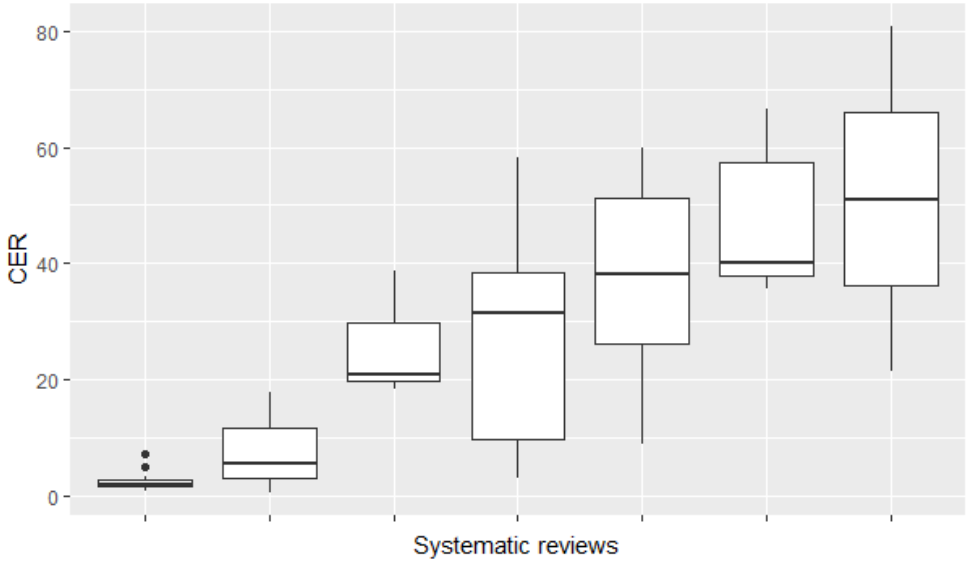


Figure 5.6. Boxplot for CER for communicable disease reviews.
Ordered (left to right) from smallest to largest median CER values.

Figures 5.7 and 5.8 show median follow-up (months) distributions for the 17 Cochrane reviews and 17 non-Cochrane reviews, respectively. Figures 5.9 and 5.10 show distributions for median follow-up duration (in months) for the 27 noncommunicable diseases and 7 noncommunicable disease reviews, respectively. In all 4 cases, great variability is again evident in the distributions.

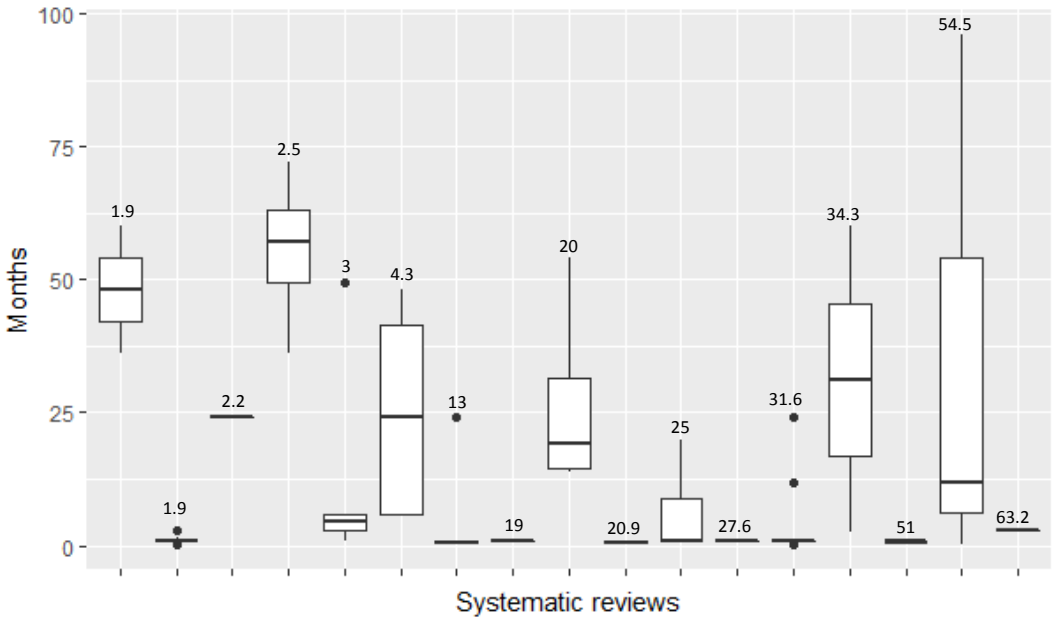


Figure 5.7. Boxplot of median follow-up values for Cochrane reviews.
Ordered (left to right) from smallest to largest median CER values (numbers).

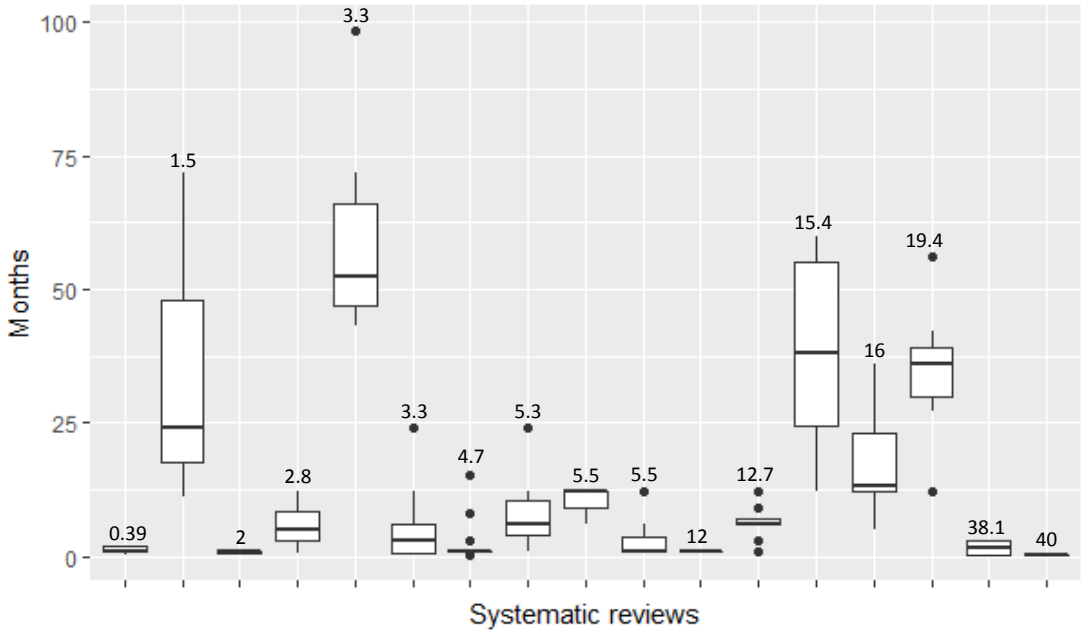


Figure 5.8. Boxplot of median follow-up values for non-Cochrane reviews.
Ordered (left to right) from smallest to largest median CER values (numbers).

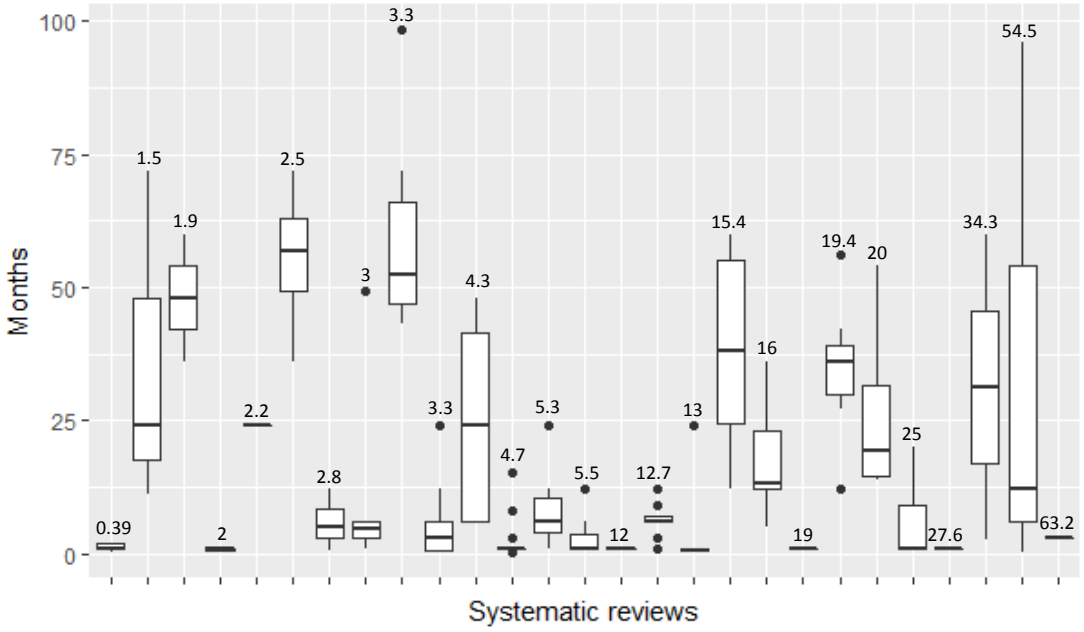


Figure 5.9. Boxplot of median follow-up values for noncommunicable disease reviews. Ordered (left to right) from smallest to largest median CER values (numbers).

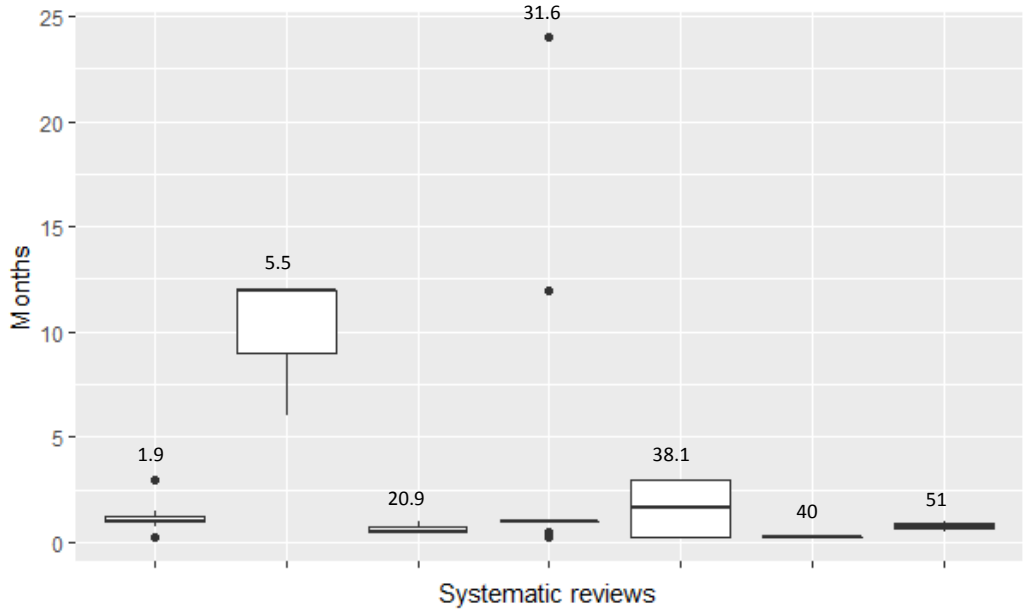


Figure 5.10. Boxplot of median follow-up values for communicable disease reviews. Ordered (left to right) from smallest to largest median CER values (numbers).

Figures 5.11 and 5.12 show standardized CER distributions for the 17 Cochrane reviews and 17 non-Cochrane reviews, respectively. Figures 5.13 and 5.14 show standardized CER distributions for the

27 noncommunicable diseases and 7 noncommunicable disease reviews, respectively. It can be observed that standardization achieves a similar median of values between the different reviews.

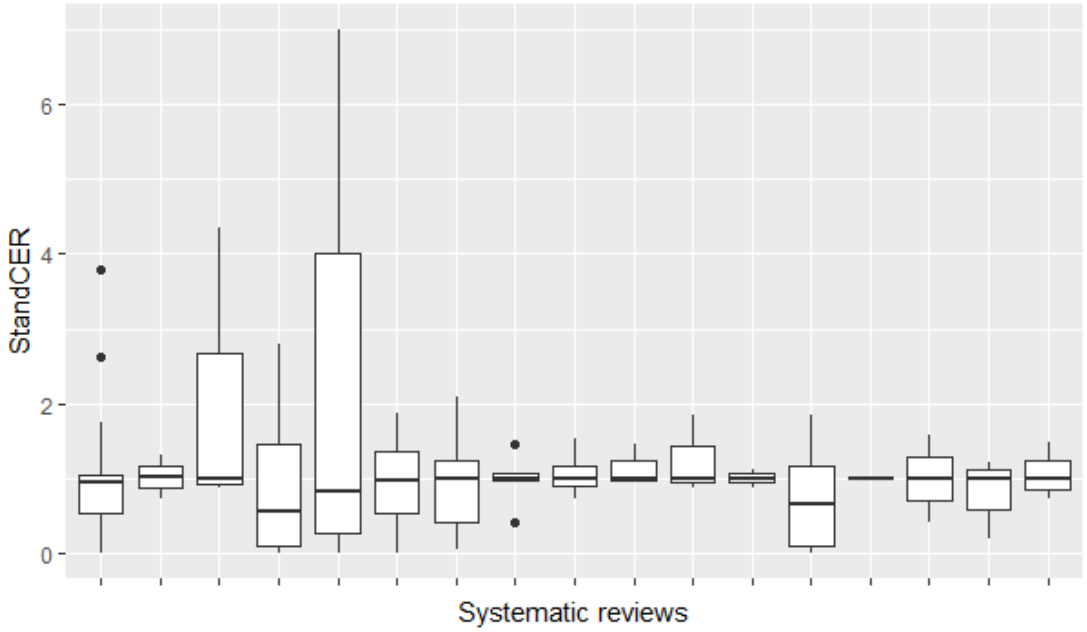


Figure 5.11. Boxplot of standardized CER for Cochrane reviews.
Ordered (left to right) from smallest to largest median CER values.

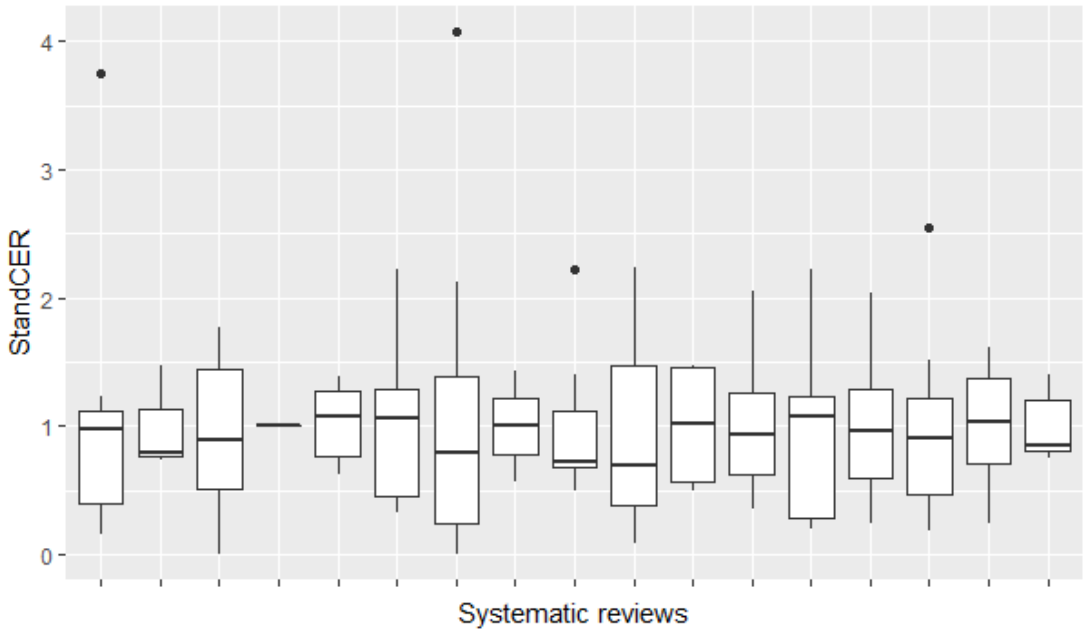


Figure 5.12. Boxplot of standardized CER for non- Cochrane reviews.
Ordered (left to right) from smallest to largest median CER values.

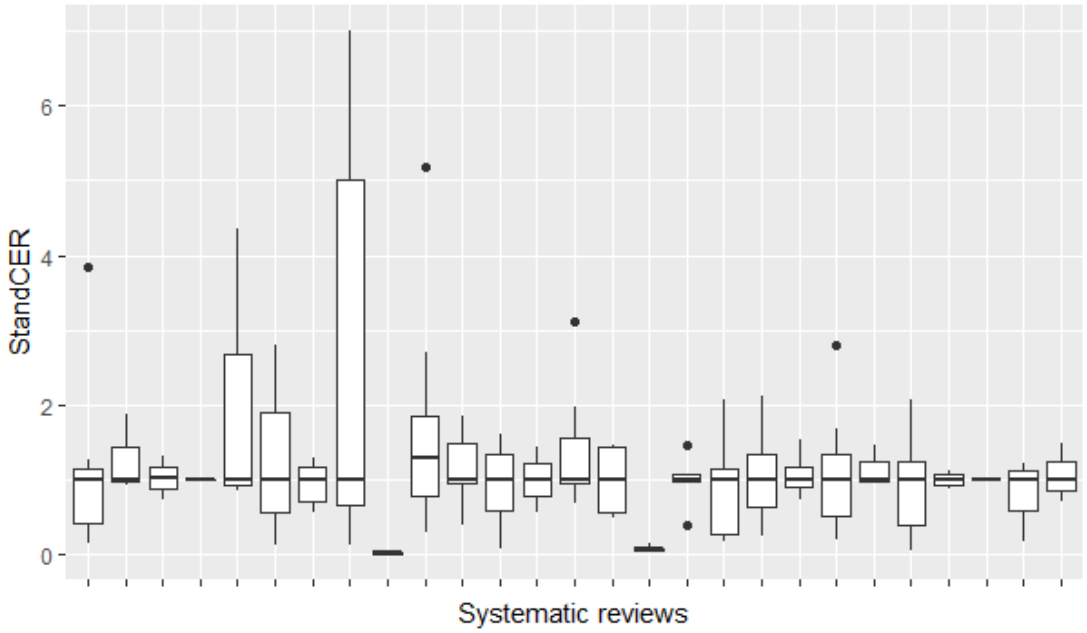


Figure 5.13. Boxplot of standardized CER for noncommunicable disease reviews. Ordered (left to right) from smallest to largest median CER values.

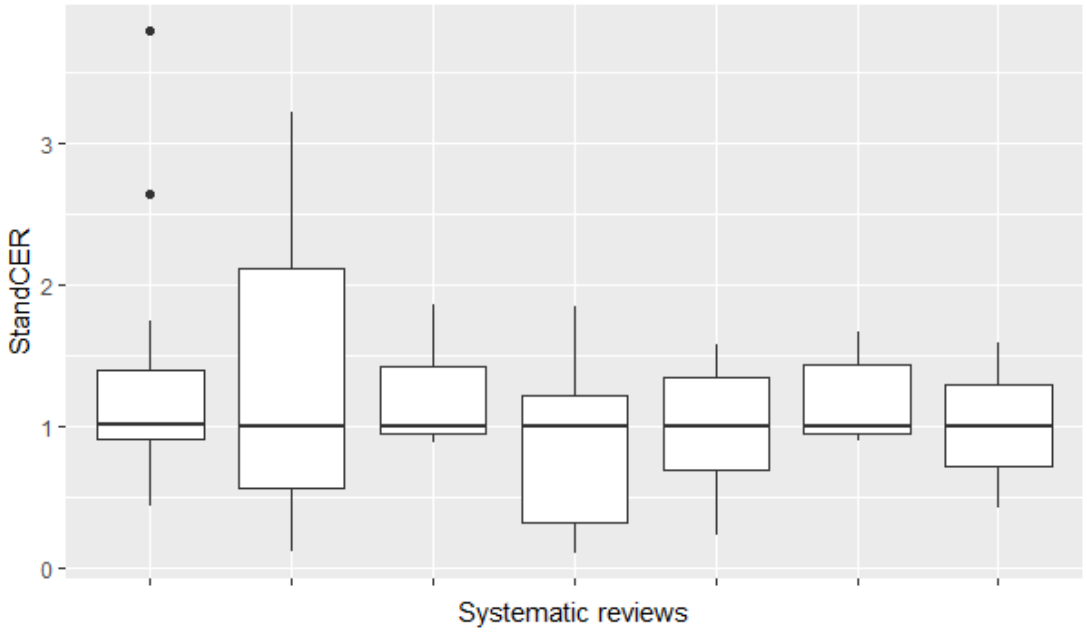


Figure 5.14. Boxplot of standardized CER for communicable disease reviews. Ordered (left to right) from smallest to largest median CER values.

Figures 5.15 and 5.16 show standardized follow-up duration distributions for the 17 Cochrane reviews and 17 non-Cochrane reviews, respectively. Figures 5.17 and 5.18 show standardized

follow-up duration distributions for the 27 noncommunicable diseases and 7 noncommunicable disease reviews, respectively. It can be observed that standardization achieves a similar median of values between the different reviews.

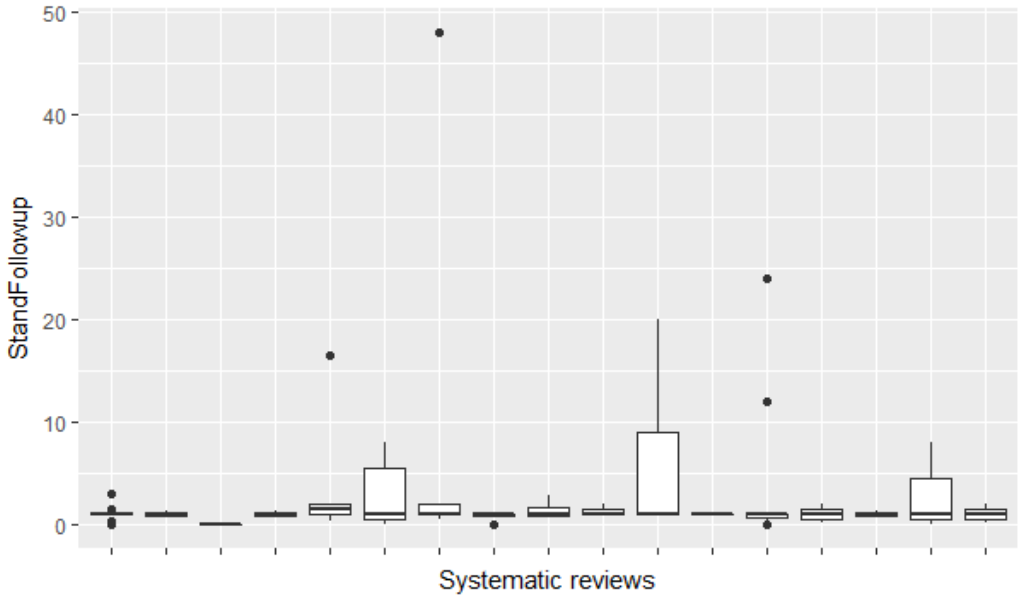


Figure 5.15. Boxplot of standardized follow-up duration for Cochrane reviews. Ordered (left to right) from smallest to largest median CER values.

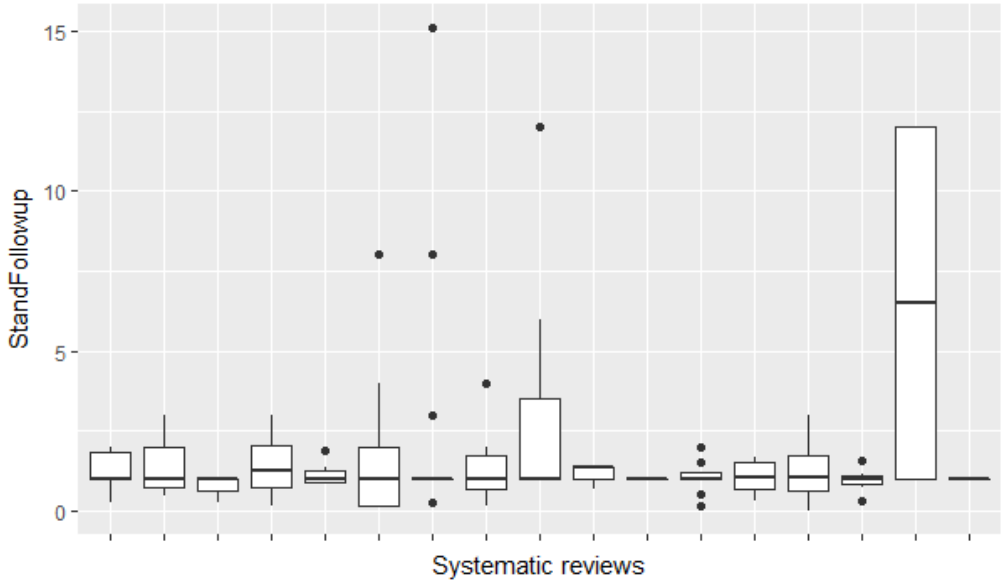


Figure 5.16. Boxplot of standardized follow-up duration for non-Cochrane reviews. Ordered (left to right) from smallest to largest median CER values.

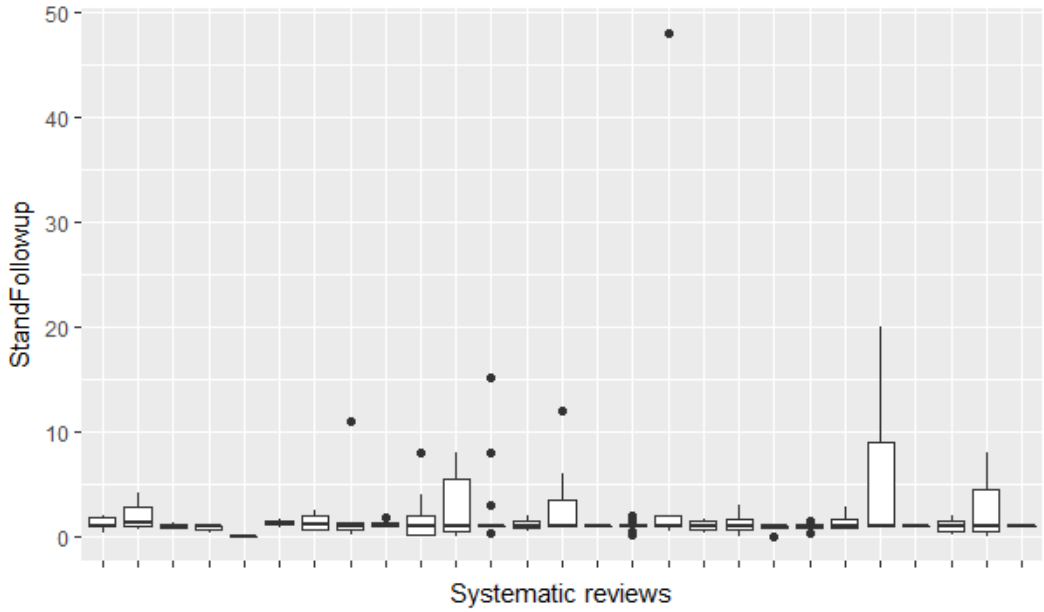


Figure 5.17. Boxplot of standardized follow-up duration for noncommunicable disease reviews. Ordered (left to right) from smallest to largest median CER values.

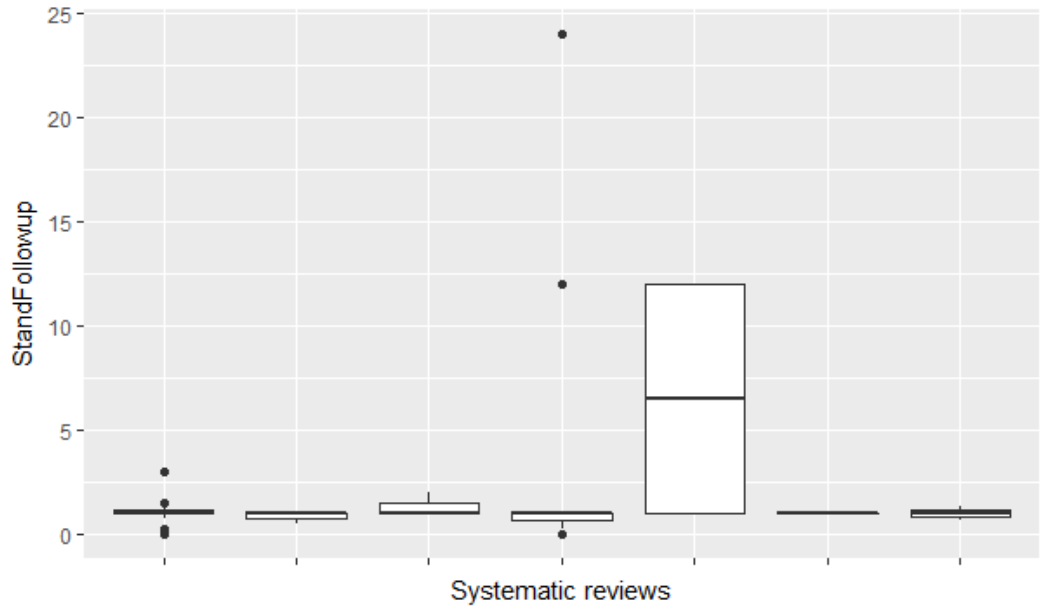


Figure 5.18. Boxplot of standardized follow-up duration for communicable disease reviews. Ordered (left to right) from smallest to largest median CER values.

5.3.3. Linear regression analysis

Figures 5.19-5.21, which depict regression lines for the standardized CER against standardized follow-up duration, standardized sample size (log number of patients) and standardized ES

(expressed as RR) for the 236 RCTs, respectively, show that these relationships were not linear. Regression for standardized CER against standardized follow-up duration (person-year follow-up) showed that CER increased as follow-up duration increased (y=0.0097x+1.0643; R^2=0.0052), with only 0.5% of standardized CER variation predicted by the regression line (Figure 5.19). Regression for standardized CER against standardized sample size showed that CER decreased as sample size increased (y=-0.0094x+1.0357; R^2=0.0036), with only just over 3% of CER variation predicted by the regression line (Figure 5.20). Finally, regarding standardized CER and standardized ES (RR), the regression analysis showed that standardized CER increased as ES increased (y=0.0315x+0.9783; R^2=0.0015), with just 0.1% of standardized CER variation predicted by the regression line (Figure 5.21).

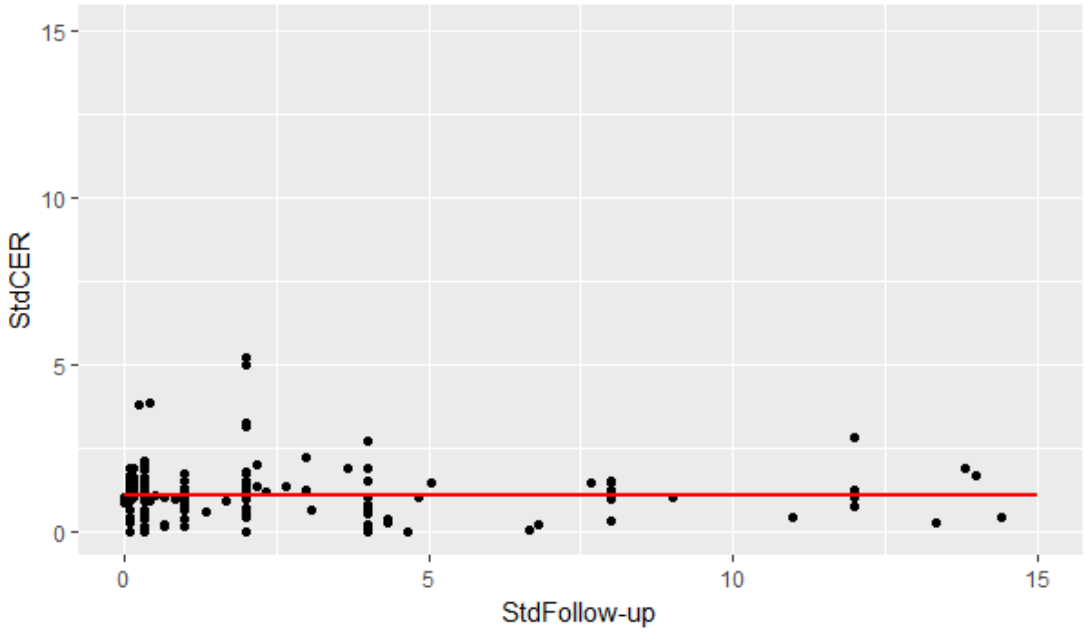


Figure 5.19. Nonlinear correlation between standardized follow-up duration and CER values for RCTs.

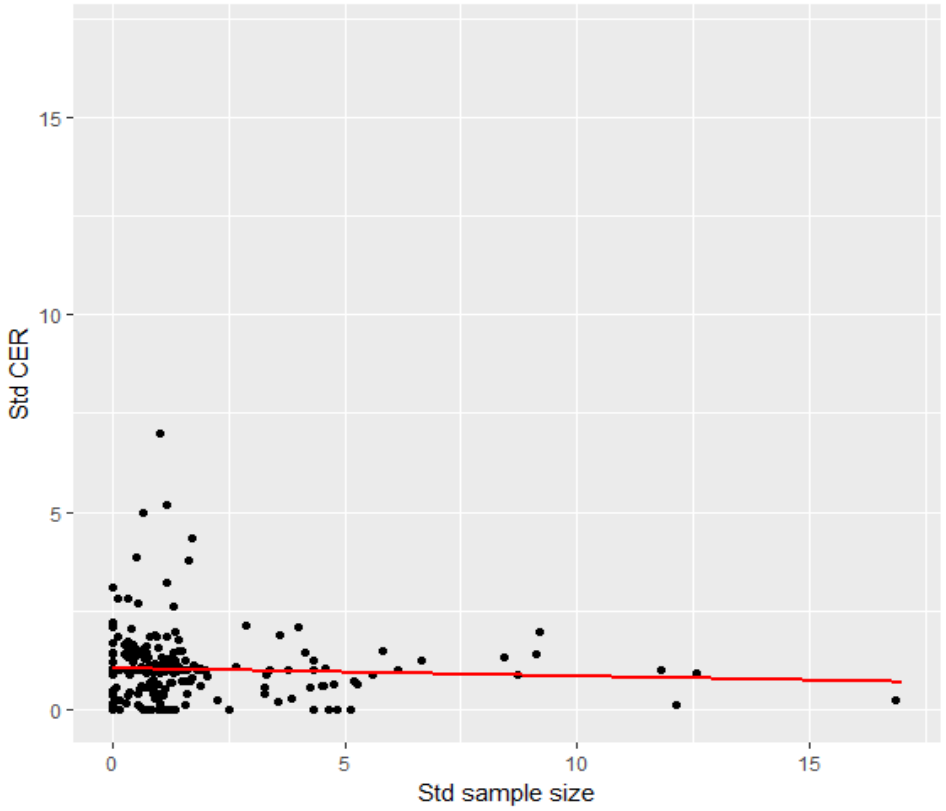


Figure 5.20. Nonlinear correlation between standardized sample size and CER values for RCTs.

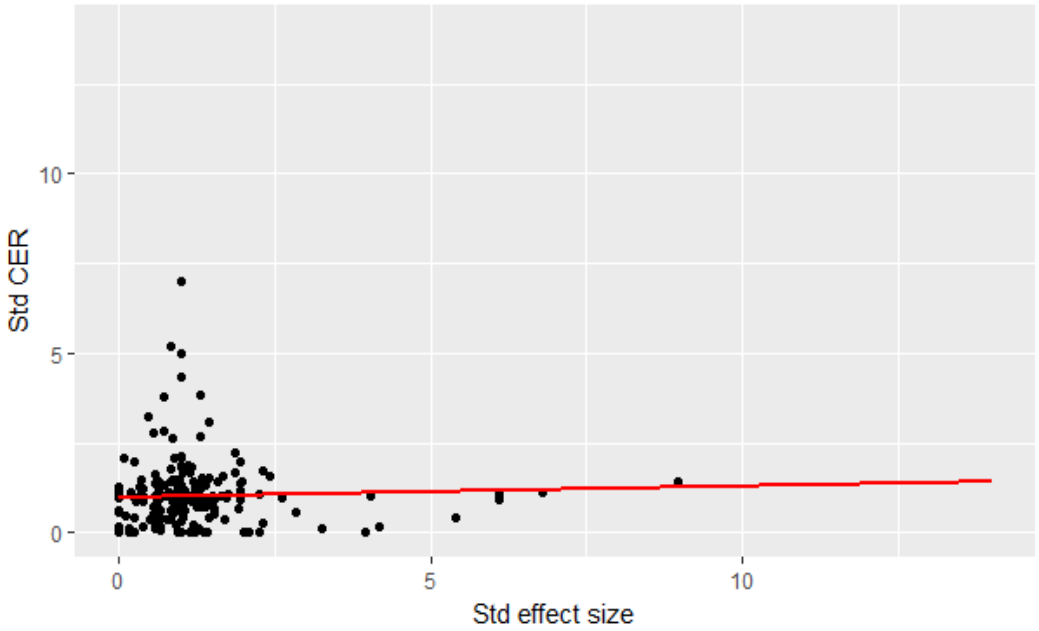


Figure 5.21. Nonlinear correlation between standardized ES and CER values for RCTs.

The great variability in CER and follow-up duration values observed in the RCTs, which reflect the descriptive results (mean, SD and interquartile range), however, meant that simple linear regression analysis was unable to detect any cluster effect.

5.3.4. Multilevel linear regression analysis

Since the simple linear regression was unable to detect correlations — because of highly variable follow-up duration and CER values for RCTs within meta-analyses (as indicated by Figures 5.3-5.6 for CER and Figures 5.7-5.10 for follow-up duration) — multilevel linear regression analysis, using the reviews as the index factor, was applied in an attempt to identify correlations for subgroups that were not identified using simple linear regression.

The multilevel linear regression output (Table 5.11) suggests that there was no direct association between standardized follow-up duration and standardized CER.

Table 5.11. Multilevel linear regression analysis output.

R^2 overall = 0.0296						
Standardized follow-up	**Coeff.**	**Std. Err.**	**z**	**P>(z)**	**[95% CI]**	
Standardized CER	.495237	.3238697	1.53	0.126	-.139536	1.13001
Cochrane	1.165755	.7376957	1.58	0.114	-.2801022	2.611612
Communicable diseases	.123519	.8855748	0.14	0.889	-1.612176	1.859214
_Cons	1.207933	.538873	2.24	0.025	.1517614	2.264105
Sigma_u	.97474593					

The direction of the association, however, did show a very slight tendency for larger CER values to be associated with greater follow-up duration; nonetheless, this result, even if slightly positive, has to be interpreted with care.

5.4. Discussion

5.4.1. Statement of principal findings

The highest proportion of the 34 analysed meta-analyses (79.4%) corresponded to noncommunicable diseases, which was only to be expected, as these conditions are generally chronic and so have a longer follow-up to outcome (mortality). Pharmacological interventions (73.5%) were predominant in both the Cochrane and non-Cochrane reviews, although substantially less so in the Cochrane group (58.8%). The number of control group events was much greater in the non-Cochrane group (59.4) than in the Cochrane group (30.4). The mean number of patients per review was 885, and non-Cochrane reviews on average tripled Cochrane reviews in terms of sample size.

Mean follow-up was 15 months (slightly longer for the Cochrane group) and lasted >1 month in 61.1% of all reviews. The fact that the percentage of tRCTs was under 10% in both the Cochrane and non-Cochrane groups meant that a differentiated analysis for this subgroup was not possible.

Mean CER in the Cochrane reviews was almost double that of the non-Cochrane reviews (22.5 vs 11.9), which would indicate that – despite a higher number of patients in the non-Cochrane group (n=1130) than in the Cochrane group (421.3) — if the statistical parameters had the same value, then non-Cochrane reviews would require a larger number of patients to achieve OIS.

No correlation was initially observed between standardized CER and standardized follow-up duration, sample size or ES. It was observed, however, that CER seemed to decline as follow-up duration increased. This may be due, however, to the ecological fallacy, i.e. drawing conclusions about individuals based on analyses of group data; alternatively, it may be due to the great variability in the reviews, both in terms of follow-up duration and CER values. Furthermore, as

sample size increased, CER tended to decrease, indicating that larger RCTs would have a lower CER value. From an OIS estimation perspective, this issue may represent a penalization of larger meta-analyses, as demonstrated in Chaper 4 of this thesis: meta-analyses with larger samples (e.g. >10000 included patients) associated with small CER values did not achieve OIS. In contrast, as ES increased, CER tended to decrease.

Although CER seemed to decline as follow-up duration increased, when these parameters were standardized according to sample size, an opposite correlation trend was observed, i.e. CER seemed to increase as follow-up duration increased.

5.4.2. Limitations

A limitation in relation to sample size is that, despite the inclusion of over 200 RCTs, the overall number of 34 meta-analyses was small, which would suggest a lack of representativeness. Regarding the disease class variable, around 80% of reviews covered noncommunicable diseases. This fact could introduce bias, given that mean CER was more than double for communicable diseases compared to noncommunicable diseases. This difference was also observed for mean follow-up duration, which was, logically, longer for noncommunicable diseases than for communicable diseases. Regarding type of intervention, nearly three quarters of the reviews covered pharmacological interventions. Mean CER for nonpharmacological interventions was nearly double that for pharmacological interventions. These imbalances would suggest that the inclusion of more meta-analyses that evaluated communicable diseases and nonpharmacological interventions would improve the representativeness of the sample.

Another issue related to representativeness was that a differentiated analysis of tRCTs was not possible, as 95% of the included RCTs were non-truncated. It would have been useful to have been able to evaluate the impact of early stopping on CER and the number of outcomes.

Another limitation was that only RCTs were included, despite recommendations regarding the usefulness of observational studies for estimating baseline risk (10,11). It would therefore be useful to analyse possible difference in baseline risk for RCTs in comparison to observational studies.

Yet another limitation regarding the sample is that the only outcome type evaluated was mortality. Indeed, this issue limited the possibility of locating a larger number of meta-analyses, as a significant proportion of meta-analyses corresponding to the period analysed did not include the mortality outcome. On the basis that possible differences may exist in the number of events, baseline risk is likely to vary and follow-up duration may be prolonged until the outcome of interest is achieved. Assuming possible differences in the number of events, the baseline risk is likely to vary and follow-up duration may be prolonged until the outcome of interest is achieved. For this reason, the results of this analysis cannot be extrapolated to outcomes other than mortality.

5.4.3. Implications for research and unanswered questions

The results obtained do not point to any clear relationship between tCER and follow-up duration. One reason why this relationship is difficult to demonstrate is because of the way RCTs are run, and, although standardization is of some help, the power of the sample to detect a linear association is poor.

From the perspective of estimating OIS, one important issue regarding CER is the 'moving target phenomenon' affecting the values used for OIS estimation; these values change as new RCTs are included in a cumulative meta-analysis. CER is therefore affected because its value will vary as data from new patients are incorporated.

While the use of informative priors is useful for establishing heterogeneity values (whose behaviour is covered in Chapter 4 of this thesis) for OIS estimation, this is not the case for CER. On the basis of

the results obtained here, it is difficult to consider how to develop informative priors for CER, e.g. according to disease classes or follow-up duration.

Given the great variability in CER values in the RCTs within meta-analyses, as demonstrated in the results obtained here, it is also necessary to consider what other variables (e.g. population characteristics, outcome type or disease stage) could explain CER behaviour in individual RCTs and in meta-analyses. It would also be potentially useful to analyse large observational studies in order to compare their CER values with those for RCTs.

In conclusion, based on these findings, no clear recommendation on how to estimate CER for OIS calculations is possible. This conclusion would indicate the need to further investigate this issue for a larger sample of RCTs. A larger sample would enable an analysis of the impact of stopping RCTs early on CER estimation. Another benefit of a larger sample of RCTs would be the possibility of exploring, using multilevel linear regression analysis, other independent variables (different types of outcomes and diseases, types of outcome reporting and analyses) that might explain CER behaviour.

5.5. Conclusions

Despite the fact that CER is an essential parameter in estimating OIS, there is no consensus regarding how to calculate it. Using the median value, rather than the mean value, would appear to be the most recommended approach. The CER value appears to vary depending on whether the intervention is pharmacological or nonpharmacological and also according to whether the disease is communicable or noncommunicable. As for a possible relationship between CER and follow-up, no clear association is evident, although multilevel linear regression suggests that CER increases as follow-up duration increases.

5.6. Chapter summary

- A difference in CER was observed between Cochrane reviews and non-Cochrane reviews that could potentially affect OIS calculations.
- The great variability in CER and follow-up duration values within reviews would suggest that the ecological fallacy plays a role.
- Multilevel linear regression and use of standardized CER and follow-up duration values would seem to indicate that CER tends to increase as follow-up duration increases.
- No clear recommendation is possible regarding how to estimate CER for the purpose of calculating OIS.

5.7. References

1. Sharp SJ, Thompson SG, Altman DG. The relation between treatment benefit and underlying risk in meta-analysis. BMJ. 1996; 313(7059):735-8.
2. Newcombe RG, Bender R. Implementing GRADE: calculating the risk difference from the baseline risk and the relative risk. Evid Based Med. 2014; 19(1):6-8.
3. Sutton AJ, Welton NJ, Ades AE, Cooper N, Abrams KR. Evidence Synthesis for Decision Making in Healthcare. First edition. United Kingdom: John Wiley &Sons, Ltd; 2012.
4. Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, Montori V, Akl EA, Djulbegovic B, Falck-Ytter Y, Norris SL, Williams JW Jr, Atkins D, Meerpohl J, Schünemann HJ. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). J Clin Epidemiol. 2011; 64(4):407-15.
5. Spencer FA, Iorio A, You J, Murad MH, Schünemann HJ, Vandvik PO, Crowther MA, Pottie K, Lang ES, Meerpohl JJ, Falck-Ytter Y, Alonso-Coello P, Guyatt GH. Uncertainties in baseline risk estimates and confidence in treatment effects. BMJ. 2012; 345:e7401.
6. Guyatt GH, Oxman AD, Santesso N, Helfand M, Vist G, Kunz R, Brozek J, Norris S, Meerpohl J, Djulbegovic B, Alonso-Coello P, Post PN, Busse JW, Glasziou P, Christensen R, Schünemann HJ. GRADE guidelines: 12. Preparing summary of findings tables-binary outcomes. J Clin Epidemiol. 2013; 66(2):158-72.
7. Hayden JA, Tomlinson G, Ni A, Parker R, Cartwright J, van Tulder M, et al. Approaches to estimate and present baseline risks: Recommendations for Cochrane reviews 'Summary of findings' tables. Cochrane Database Syst Rev Suppl 3–233 (2011).

8. Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JP. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. Int J Epidemiol. 2012;41(3):818-27. 29.
9. Montori VM, Devereaux PJ, Adhikari NK, Burns KE, Eggert CH, Briel M, Lacchetti C, Leung TW, Darling E, Bryant DM, Bucher HC, Schünemann HJ, Meade MO, Cook DJ, Erwin PJ, Sood A, Sood R, Lo B, Thompson CA, Zhou Q, Mills E, Guyatt GH. Randomized trials stopped early for benefit: a systematic review. JAMA. 2005;294(17):2203-9.
10. Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. Trials. 2015;16:495.
11. Prasad V. Baseline event rate, the Concorde fallacy, and the topography of cardiac risk. Med Hypotheses. 2010;75(6):639-41.

Chapter 5. Appendix

Table A5.1. Descriptive characteristics of the 34 meta-analyses.

Author, year, journal, country.	Disease class	Intervention type	Included RCTs	Included patients	Median CER
Rojas-Reyes MX, 2012, Cochrane Database Syst Rev, Colombia.	Noncommunicable	Pharmacological	4	2571	19
Rudic JS, 2012, Cochrane Database Syst, Serbia.	Noncommunicable	Pharmacological	7	829	4.3
Bahadue FL, 2012, Cochrane Database Syst, USA.	Noncommunicable	Pharmacological	5	3845	25
Heran BS, 2012, Cochrane Database Syst, Canada.	Noncommunicable	Pharmacological	6	9926	3
Jamal SA, 2013, Lancet, USA.	Noncommunicable	Pharmacological	10	2913	16
Lv J, 2012, Cochrane Database Syst, Australia	Noncommunicable	Pharmacological	4	9227	2.5
Eliakim-Raz N, 2012, Cochrane Database Syst, Israel.	Communicable	Pharmacological	14	4180	1.9
Fang Y, 2012, Cochrane Database Syst, China.	Noncommunicable	Non-pharmacological	5	675	13

Sajid MS, 2012, Cochrane Database Syst, UK.	Noncommunicable	Non-pharmacological	3	245	2.2
Masters B, 2012, Cochrane Database Syst, UK.	Noncommunicable	Non-pharmacological	2	108	27.6
Berton DC, 2012, Cochrane Database Syst, Brasil.	Communicable	Non-pharmacological	3	1541	20.9
Zhang Y, 2012, Cochrane Database Syst, China.	Communicable	Non-pharmacological	2	155	51
Szakmany T, 2012, Cochrane Database Syst, UK.	Communicable	Pharmacological	10	953	31.6
Diao D, 2012, Cochrane Database Syst, Canada.	Noncommunicable	Pharmacological	3	9072	1.9
Chaudhari T, 2012, Cochrane Database Syst, UK.	Noncommunicable	Pharmacological	3	151	54.5
Jun M, 2012, Cochrane Database Syst, Australia.	Noncommunicable	Pharmacological	4	1817	20
Ricci S, 2012, Cochrane Database Syst, Italy.	Noncommunicable	Non-pharmacological	3	268	63.2
McCarthy K, 2012, Cochrane Database Syst, UK.	Noncommunicable	Non-pharmacological	2	1779	34.3

Bellemain-Appaix A, 2012, JAMA, France.	Noncommunicable	Pharmacological	7	8759	2
Fox BD, 2012, BMJ, Canada.	Noncommunicable	Pharmacological	8	14545	2.8
Hemmingsen B, 2012, BMJ, Denmark.	Noncommunicable	Pharmacological	16	1552	5.3
Neumann I, 2012, Ann Intern Med, Canada.	Noncommunicable	Pharmacological	10	21462	0.39
Palmer SC, 2012, Ann Intern Med, Italy.	Noncommunicable	Pharmacological	7	6093	5.5
Rothwell PM, 2012, Lancet, UK.	Noncommunicable	Pharmacological	7	42364	3.3
Silvain J, 2012, BMJ, France.	Noncommunicable	Pharmacological	18	30188	4.7
Al-khatib SM, 2014, Ann Intern Med, USA.	Noncommunicable	Pharmacological	6	7254	15.4
Feltner C, 2014, Ann Intern Med, USA.	Noncommunicable	Non-pharmacological	24	6177	12.7
Makani H, 2013, BMJ, USA.	Noncommunicable	Pharmacological	7	65475	19.4
Haase N, 2013, BMJ, Denmark.	Communicable	Pharmacological	6	4070	38.1
Schmolzer, 2013, BMJ, Canada.	Noncommunicable	Non-pharmacological	4	3083	12
Ford AC, 2014, BMJ, UK.	Noncommunicable	Pharmacological	3	3063	1.5
Patel A, 2014, BMJ, UK.	Communicable	Pharmacological	7	261	40
Udell JA, 2013, JAMA, Canada.	Communicable	Pharmacological	3	1434	5.5
Chatterjee S, 2014, JAMA, USA.	Noncommunicable	Pharmacological	13	1841	3.3

Chapter 6. International survey: awareness and perceptions of optimal information size

Overview

Research objectives: To determine awareness and perceptions of optimal information size (OIS) among authors of systematic reviews and authors of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement (non-PRISMA and PRISMA authors, respectively).

Background: A number of published studies recommend using OIS estimates in all published systematic reviews and meta-analyses. However, it is not known to what extent authors are aware of OIS or how it is perceived.

Methods: A total of 500 authors of published meta-analysis and 32 PRISMA Statement authors were surveyed by means of a questionnaire consisting of 10 questions designed to compile data on awareness and perceptions of OIS estimation.

Results: A total of 88 responses were received (71 non-PRISMA and 17 PRISMA). Response rates for the 500 non-PRISMA authors and the 32 PRISMA authors contacted were 14.2% and 53.1%, respectively. High proportions of authors from both groups agreed that meta-analyses may overestimate treatments or falsely identify false positive results (non-PRISMA 80.6% vs PRISMA 93.8%). The main view regarding whether meta-analysis results were conclusive was neutral (non-PRISMA 66.7% vs PRISMA 42.1%). As for wanting to know more about OIS, the strongest position was affirmative (non-PRISMA 75% vs PRISMA 60%). Similar percentages also responded that OIS would be useful or necessary (non-PRISMA 72.8% vs PRISMA 64.7%). As for possible reasons for not using OIS, the majority response for non-PRISMA authors (over 40%) was a lack of familiarity with

OIS; PRISMA authors also mentioned a lack of familiarity with OIS (30%) but also stated that OIS would not add further to a review (30%). Non-PRISMA and PRISMA authors tended to diverge more when it came to responses that reflected methodological knowledge. Thus, more non-PRISMA authors compared to PRISMA authors would consider using OIS (80.5% vs 47.1%), but were less familiar with statistical procedures, including those necessary to calculate OIS (37.1% vs 52.9%). Finally, divergence was also evident in responses in favour of inclusion of OIS in PRISMA recommendations (non-PRISMA 47.1% vs PRISMA 35.3%) and whether OIS would be used if recommended (non-PRISMA 55.7% vs PRISMA 29.4%).

Conclusion: While the concept of OIS is of substantial interest to both non-PRISMA and PRISMA authors, the latter tend to be less enthusiastic, possibly due to the lack of methodological development of the technique.

6.1. Introduction

The number of systematic reviews and meta-analyses published each year is growing (1), as is the number of reviews that incorporate OIS estimates, although these are still a tiny proportion of all reviews with meta-analyses.

First published in 2009 (2), the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statements consists of a checklist and flow diagram aimed at helping authors improve how they report systematic reviews and meta-analyses. Although not intended as an instrument for quality assessment of systematic reviews, they may be useful for their critical appraisal. PRISMA has also developed several extensions, for instance, PRISMA-P for systematic review protocols (3), for other methodological designs such as network meta-analysis (4) and for other outcomes such as harms (5). Many indexed journals now require adherence to PRISMA by authors of systematic

reviews and meta-analyses. Given the impact and influence of PRISMA and its extensions, the opinions of PRISMA Statement authors are likely to have a bearing on the inclusion of OIS estimation in PRISMA recommendations and, therefore, on its subsequent inclusion in systematic reviews and meta-analyses.

Furthermore, although a number of authors recommend the use of OIS in all published meta-analyses (6-8), no data are available on awareness and perceptions regarding this recommendation by systematic review authors and methodologists.

The objective of this chapter was to determine awareness and perceptions of OIS estimation among PRISMA Statement authors and among a sample of authors of meta-analyses.

6.2. Methods

6.2.1. Description

The survey consisted of a questionnaire addressed to authors of meta-analyses published in 2015 and current authors of the PRISMA Statement (non-PRISMA authors and PRISMA authors, respectively). The data were collected using an anonymous online Google Forms survey created specifically for this study. The questionnaire was sent to 500 non-PRISMA authors and 32 PRISMA authors (total n=532).

6.2.2. Participants

Randomly selected were 500 potential authors of systematic reviews with meta-analyses as well as all 32 members of the PRISMA group that drew up the PRISMA Statement according to a BMJ publication from 2009 (9). For the non-PRISMA authors, using simple random sampling and the term ‘meta-analysis [pt]’, 500 names and emails were extracted from abstracts indexed in 2015 in

Medline (PubMed). For the PRISMA authors, name and email details were obtained from the corresponding author of the PRISMA Statement (David Moher).

The introductory email to authors consisted of an explanation of the purpose of the research and the survey and an invitation to participate in the online survey. Those who agreed to participate were informed that participation was voluntary, that they could withdraw at any time and that their information would be treated as confidential. Completion and submission of the online questionnaire was considered to be adequate evidence of consent.

6.2.3. Pilot survey

The questionnaire, consisting of 6 main questions, was designed to assess author perceptions regarding use of OIS estimates in meta-analyses. The questions covered general awareness of OIS, prevalence of OIS use and rationale for incorporating the OIS methodology in meta-analyses.

To fully test the questionnaire before carrying out the definitive survey, 8 researchers attached to the University of Oxford's Nuffield Department of Primary Care Health Sciences were invited via email to participate in a pilot study. These researchers, all of whom agreed to participate, had different levels of experience and expertise regarding systematic reviews and meta-analyses. Five of the 8 researchers were initially selected to complete the survey and provide feedback on each question. The questionnaire was then further revised and sent to all 8 piloting researchers, who also provided feedback on the basis of which the definitive questionnaire was drawn up.

The pilot survey was designed and conducted using the Google Forms online survey tool.

6.2.4. Questionnaire

The final Google Forms survey, reproduced in full in Table A6.1 in the Appendix, is briefly described below in terms of its key sections.

Introductory text: An introduction to OIS consisting of a brief explanation and an example.

Information for participants: A brief explanation of the research and of the purpose of the survey, details of how the invited participant was selected and details of ethical and legal aspects of data processing and information confidentiality.

Questionnaire: 6 questions with 4 sub-questions, equivalent to 10 questions as follows: general awareness of OIS and specific meta-analysis concepts (4 questions); perceptions of OIS by the participants (3 questions); technical aspects of OIS calculation (1 question); and reactions to a possible recommendation regarding OIS in future PRISMA Statement updates (2 questions).

6.2.5. Survey procedure

The online questionnaire was initially sent to participants on 11 July 2016. Reminders were sent on 17 September, 2 October and 1 November 2016. Whenever an automatic reply indicating non-availability was received, the name and email of a replacement non-PRISMA author was randomly selected from the list of authors identified in Medline (PubMed).

In an endeavour to improve the response rate for non-PRISMA authors, on 11 February 2017, the same questionnaire was sent via a single email to 200 further authors of non-PRISMA meta-analyses published in 2015, randomly selected, again, from the list of non-PRISMA authors previously identified in Medline (PubMed).

6.2.6 Ethics

The study was reviewed by and received ethical clearance from the University of Oxford Central University Research Ethics Committee (see Table A6.2 in the Appendix).

6.2.7. Analysis

The different sections of the questionnaire were analysed descriptively. The chi-squared test was used to compare non-PRISMA and PRISMA groups. Statistical significance was set to $p<0.05$. All statistical analyses were performed using SPSS V.22 for Windows.

6.3. Results

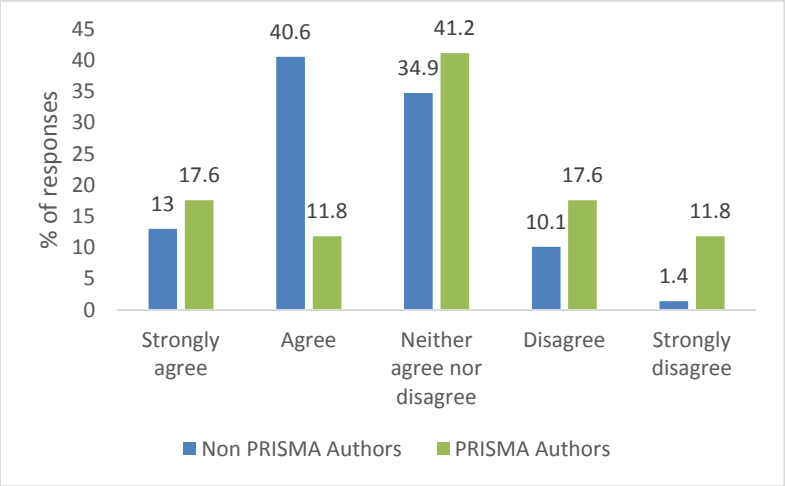
6.3.1. Response rate

A total of 5265 bibliographical references corresponding to possible meta-analyses published in 2015 were identified through Medline (PubMed). Of the 500 non-PRISMA authors asked to participate, 71 (14.2%) responded; of the 200 additional authors invited to participate, although 11 (5.5%) responded, they were excluded from the analysis due to the low response rate. Finally, of the 32 PRISMA authors invited to participate, 17 (53.1%) responded. In total, therefore, 88 responses were received (17 PRISMA and 71 non-PRISMA).

6.3.2. Responses

Question 1 (Figure 6.1)

Nearly double the proportion of non-PRISMA authors compared to PRISMA authors strongly agreed/agreed (53.6% vs 29.4%) that the terms OIS and trial sequential analysis should be used in the context of meta-analyses, whereas results for neither agree/disagree were broadly similar, at 34.9% vs 41.2%, respectively.



χ^2=8.4291, df=4, p=0.0770

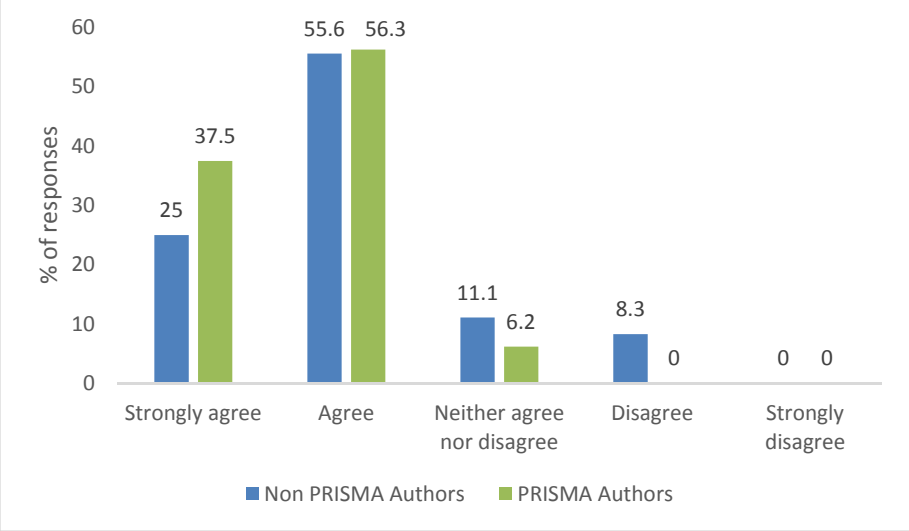
Figure 6.1. Responses (n=86) to Q1.
"The following terms should be used in the context of meta-analysis: trial sequential analysis and optimal information size."

Questions 1a, 1b and 1c (Figures 6.2, 6.3 and 6.4)

Respondents who strongly disagreed/disagreed with the statement in Q1 were asked 3 further questions. Regarding opinions as to whether meta-analyses may overestimate treatment effects or falsely identify positive results, results were broadly similar for both PRISMA and non-PRISMA authors. Very high proportions of non-PRISMA authors (80.6%) but especially of PRISMA authors (93.8%) strongly agreed/agreed. No PRISMA authors disagreed with the statement, in contrast with 8.3% of non-PRISMA authors (Figure 6.2).

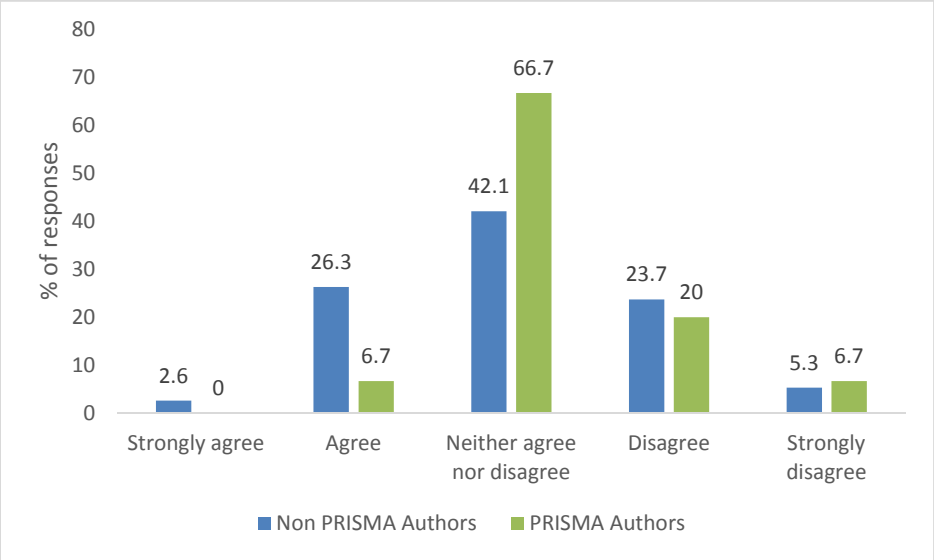
The majority of responses as to whether meta-analysis results are conclusive were neutral for authors from both groups: 66.7% for the PRISMA authors and 42.1% for the non-PRISMA authors. Of the small proportion of authors who strongly agreed/agreed, the proportion of non-PRISMA authors (28.9%) quadrupled PRISMA authors (6.7%). Disagreement levels were broadly similar for non-PRISMA authors (29%) and PRISMA authors (26.7%) (Figure 6.3).

As for wanting to know more about OIS or trial sequential analysis, the strongest position was agreement, with a larger proportion of non-PRISMA authors (75%) compared to PRISMA authors (60%) responding that they strongly agreed/agreed (95% CI: -0.2 to 0.3; difference in proportions 15%). However, within overall agreement, a higher share of non-PRISMA authors (22.5%) compared to PRISMA authors (6.7%) strongly agreed (95% CI: -0.1 to 0.2; difference in proportions 15.8%). Interestingly, 20% of PRISMA authors strongly disagreed with the statement, whereas the corresponding share for non-PRISMA authors was 0% (Figure 6.4).



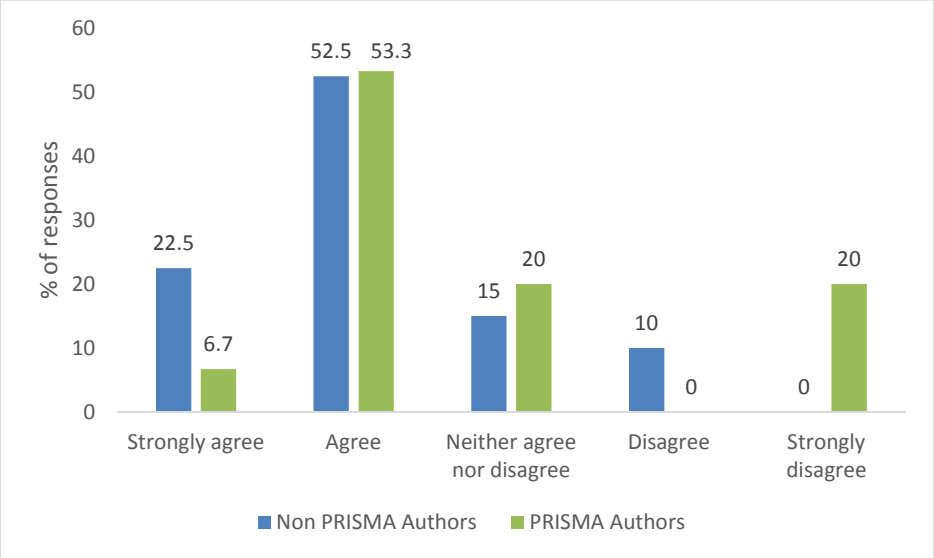
χ^2=5.487, df=4, p=0.2409

Figure 6.2. Responses (n=52) to Q1a.
"Meta-analyses may overestimate treatment effects or falsely identify positive results (type I error)."



χ^2=3.8198, df=4, p=0.4309

Figure 6.3. Responses (n=54) to Q1b.
"I consider that the results of meta-analysis are conclusive."



χ^2=3.0164, df=4, p=0.5551

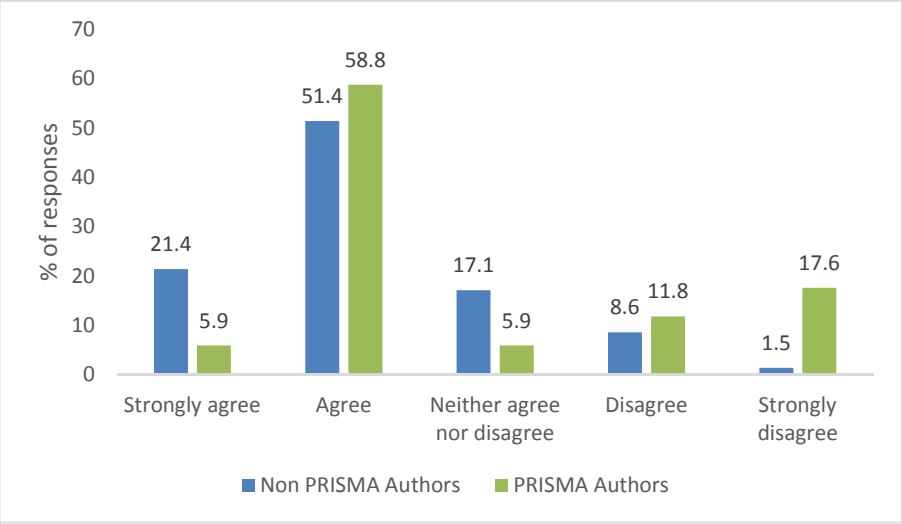
Figure 6.4. Responses (n=56) to Q1c.
"I would like to know more about optimal information size or trial sequential analysis before making up my mind."

Question 2 (Figure 6.5)

Regarding opinions as to whether OIS or similar methods would be useful or necessary, the strongest position was agreement, with a large proportion of both non-PRISMA authors (72.8%) and PRISMA authors (64.7%) responding that they strongly agreed/agreed; within that agreement, just over a fifth of non-PRISMA authors (21.4%) responded that they strongly agreed. Nearly a third (29.4%) of PRISMA authors responded that they strongly disagreed/disagreed — 3 times the proportion of non-PRISMA authors (10.1%).

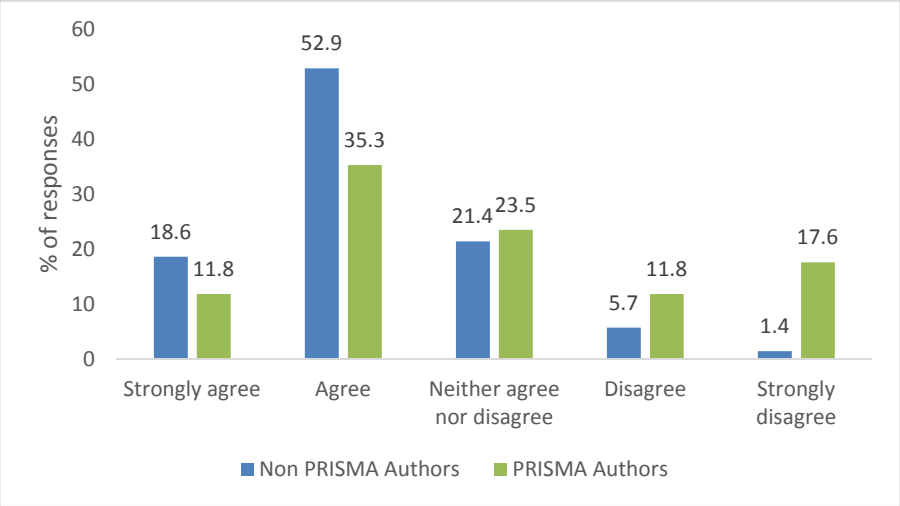
Question 3 (Figure 6.6)

As for whether respondents would consider using OIS, while a neutral stance was adopted by roughly a quarter of each group, non-PRISMA authors (80.5%) were substantially more in agreement than PRISMA authors (47.1%). As for the level of disagreement, that of the PRISMA authors (29.4%) was 4 times that of the non-PRISMA authors (7.1%).



χ^2=11.077, df=4, p=0.02571

Figure 6.5. Responses (n=87) to Q2.
"Using optimal information size or similar methods to determine the required number of participants for a meta-analysis is a useful or necessary part of a systematic review."



χ^2=9.8003, df=4, p=0.04393

Figure 6.6. Responses (n=87) to Q3.
"I would consider using optimal information size in my next systematic review."

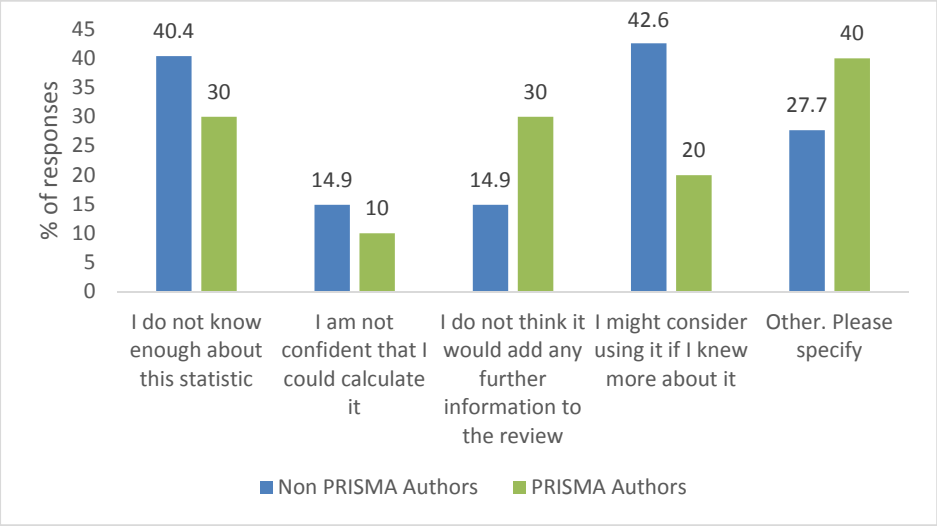
Question 4 (Figure 6.7)

Of the 4 specific response options (i.e. excluding the option 'other', responded to by 40% and 27.7% of the PRISMA and non-PRISMA authors, respectively), the 2 majority responses (over 40% each) for non-PRISMA authors reflected not knowing enough about OIS, whereas the 2 majority responses (30% each) for PRISMA authors reflected a lack of familiarity with OIS and the opinion that it would not add further to a review. These responses by the PRISMA authors would probably explain why they ticked the 'other' response option (40%) more often than any of the other 3 options (maximum 30% each). The comments included in the 'other' responses are reviewed in the discussion section.

Question 5 (Figure 6.8)

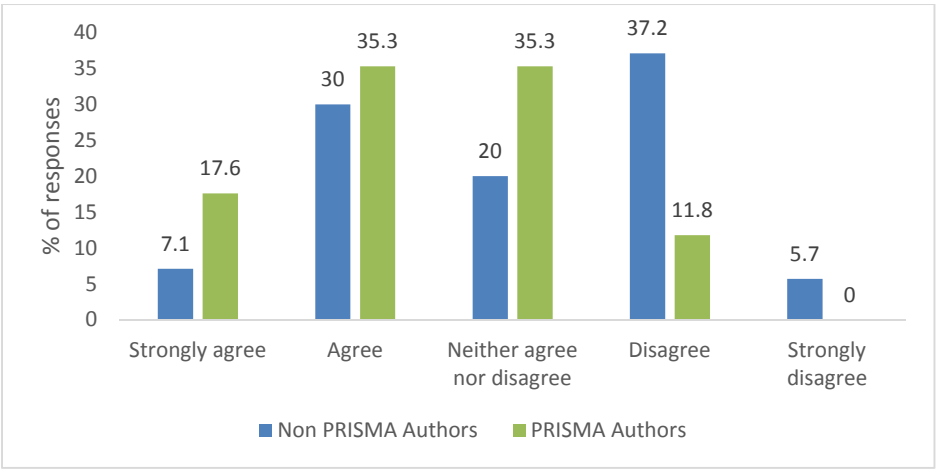
Regarding familiarity with statistical developments in meta-analysis and with OIS methods, more PRISMA authors (52.9%) than non-PRISMA authors (37.1%) strongly agreed/agreed with the statement. The share of non-PRISMA authors (42.9%) who responded that they strongly disagreed/disagreed nearly quadrupled that of the PRISMA authors (11.8%). Around a third of

PRISMA authors (35.3%) adopted a neutral stance — nearly twice the share of the non-PRISMA authors (20%).



χ^2=3.0394, df=4, p=0.5513

Figure 6.7. Responses (n=57) to Q4.
"If you are not sure whether you would use this method what are the main reasons for this? (Please select all the options that apply)."



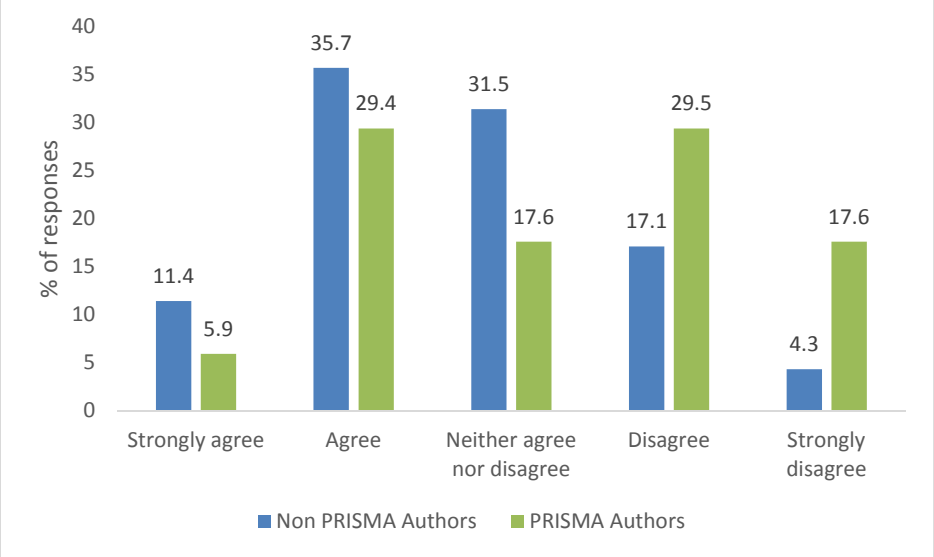
χ^2=6.8652, df=4, p=0.1432

Figure 6.8. Responses (n=87) to Q5.
"I am familiar with statistical developments in meta-analysis and with methods used to estimate optimal information size for meta-analyses."

Question 6 (Figures 6.9 and 6.10)

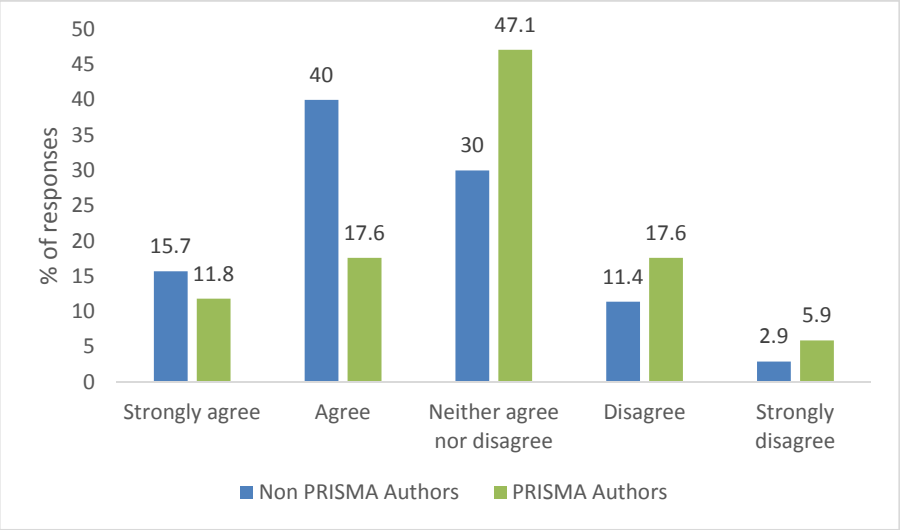
As to whether OIS estimation should be included in PRISMA recommendations, 47.1% of PRISMA authors strongly disagreed/disagreed; particularly noteworthy was the fact that 17.6% strongly disagreed. In contrast, 35.3% of PRISMA authors strongly agreed/agreed and 17.6% adopted a neutral stance. Of the non-PRISMA authors, nearly half (47.1%) agreed and just over a fifth (21.4%) disagreed (Figure 6.9).

The final question asked for opinions as to whether it would be easy to apply OIS if it were a PRISMA recommendation. For PRISMA authors the majority position was neutral (47.1%), contrasting with fairly balanced stances on agreement (29.4%) and disagreement (23.5%). For non-PRISMA authors, the majority position was agreement (55.7%), followed by a neutral stance (30%) (Figure 6.10).



χ^2=6.0628, df=4, p=0.1945

Figure 6.9. Responses (n=87) to Q6a.
"Trial sequential analysis or optimal information size should be included in the PRISMA recommendations."



χ^2=4.0363, df=4, p=0.4011

Figure 6.10. Responses (n=87) to Q6b.
"If the optimal information size were to be included in the PRISMA recommendations I would find it easy to incorporate it into my next review."

6.3.3. Statistical significance

A statistically significant difference was observed between PRISMA and non-PRISMA groups for 2 of the questions, namely Q2 ("Using optimal information size or similar methods to determine the required number of participants for a meta-analysis is a useful or necessary part of a systematic review") and Q3 ("I would consider using optimal information size in my next systematic review"), for which significance values were p=0.02571 and p=0.04393, respectively.

6.4. Discussion

6.4.1. Statement of principal findings

There were mixed opinions regarding whether the terms OIS or trial sequential analysis should be used in the context of meta-analyses. Only a third (29.4%) of PRISMA authors, compared to over half of the non-PRISMA authors (53.6%), stated they were in favour, with high and fairly similar proportions from both groups neither agreeing or disagreeing with the statement (34.9% and 41.2%,

respectively) (Figure 6.1). Note that a possible confounding factor in relation to this question is discussed as a limitation in Section 6.4.2.

Terminology issues related to information size were raised by respondents, with one PRISMA author indicating that such notions "might better be described as minimum information size and sequential meta-analysis" (see Table A6.3 in the Appendix). Alternative terms have also been discussed in the literature, e.g. 'required information size' (10). The term OIS, initially proposed by Pogue and Yusuf (11), is defined as the "minimum amount of information required for reliable conclusions"; accordingly, an information size greater than the OIS cannot be perceived as suboptimal. According to the survey respondents, more appropriate terminology to designate the statistical approach to sample size calculation for meta-analyses still needs to be better defined.

Although meta-analysis is considered to be an approach that ensures precision in the estimation of results (12, 13), a very high proportion of respondents to the survey agreed that meta-analyses can overestimate treatment effects (non-PRISMA authors 80.6% and PRISMA authors 93.8%, respectively) and only around a third of respondents (non-PRISMA authors 29% and PRISMA authors 26.7%, respectively) felt that the results of meta-analyses are conclusive (Figures 6.2 and 6.3). This would indicate there is a need to improve the level of confidence in the results of meta-analyses. One way to do that would be to reach a consensus about OIS and for information to be made available regarding certain underlying statistical assumptions for meta-analyses.

There was an evident general interest in this type of methodology (Figures 6.4, 6.5 and 6.6), as indicated by the relatively high proportion of authors wanting to know more about OIS (non-PRISMA authors 75% and PRISMA authors 60%), considering OIS to be useful and necessary (non-PRISMA authors 72.8% and PRISMA authors 64.7%), and that would consider using OIS in the future (non-PRISMA authors 80.5% and PRISMA authors 47.1%).

Interestingly, but perhaps not surprisingly, the PRISMA authors were more sceptical about OIS thyan the non-PRISMA authors, with fewer than half (47.1%) of PRISMA authors in favour of, but nearly a third (29.4%) against, using this methodology in the future (Figure 6.6). One possible explanation could be their greater methodological or statistical knowledge of meta-analysis procedures and their better understanding of the possible statistical weaknesses of the OIS approach. Indeed, a number of the PRISMA Statement authors are methodologists (14). Thus, although PRISMA authors may acknowledge that it is conceptually important to calculate the number of participants, their reservations may be related to the current methodologies available for OIS estimation — a concern that is shared elsewhere (15). This conclusion is corroborated by responses regarding the main reasons for not using OIS estimates (see Figure 6.7 referring to Q4, which allowed multiple responses) — essentially, a general lack of knowledge of the OIS methodology (the responses to options a, b and d add up to 60%). Those responses probably explain why 30% of PRISMA authors were of the opinion that OIS estimates would not add further information to a review (responses to Q4, 'other' option).

Respondents from both groups provided qualitative information that also tended to point to technical reservations regarding OIS estimation (Q4, the 'other' option; all comments in response to this option are reproduced in Table A6.4 in the Appendix):

> "The approach is based on the presumption that one might define what might be considered a worthwhile effect. This is specific and cannot be defined in absolute terms." (non-PRISMA author)

> "Meta-analysis is a retrospective investigation. Why should sample size estimation help?" (non-PRISMA author)

"While I recognise that sequential methods may have a role in prospective meta-analyses, they are inappropriate when applied retrospectively." (PRISMA author)

"I do not like the focus on statistical significance. I view the meta-analysis as a tool to give me prediction interval of the treatment effect. That is all I need to know. OIS would merely confuse me." (PRISMA author)

Technical aspects were also reflected in responses regarding familiarity with statistical developments in meta-analyses and with OIS estimates (Figure 6.8), with the PRISMA authors generally acknowledging greater familiarity than the non-PRISMA authors. Indeed, the lack of familiarity of nearly half the non-PRISMA authors is relevant, but hardly surprising, given that no recommendation as to OIS estimation is included in the PRISMA Statement.

Obviously, several issues would need to be considered before a recommendation to use OIS estimates could be included in the PRISMA Statement, for instance:

- whether OIS is a statistically sound parameter
- whether OIS might lead to confusion
- whether OIS would prevent further investigations that might be useful in the future
- whether OIS would ultimately enhance the information coming from meta-analyses
- (and most importantly) whether OIS would ultimately help make more informed clinical or health policy decisions.

Even though Cochrane reviews represent only about 15% of all published meta-analyses (16), a methodological recommendation by the Cochrane Collaboration would likely influence meta-analysis authors and methodologists. Regarding the question as to whether OIS should be a PRISMA recommendation (Figure 6.9), the PRISMA authors adopted quite a polarized position, with around

a third in favour and nearly half against; for the non-PRISMA authors opinions were more widely distributed between the for, against and neutral positions.

As for whether the OIS would be used if recommended (Figure 6.10), around a third of PRISMA authors were in favour, leaving, however, around 70% indifferent or against. As for the non-PRISMA authors, these were largely in favour. Again, it may be the case that a drawback to using OIS is its lack of methodological development as perceived by PRISMA authors, which, in turn (given the absence of a recommendation in the PRISMA Statement), is likely to affect eventual uptake among non-PRISMA authors.

Referring to the PRISMA Statement, OIS estimation methods and results could be included in checklist items #16 (currently "Describe methods of additional analyses (e.g. sensitivity or subgroup analyses, meta-regression), and if done, indicating which were pre-specified.") and #23 (currently "Give results of additional analyses, if done (e.g. sensitivity or subgroup analyses, meta-regression)"), respectively (17). The recent PRISMA-P extension — a checklist for evaluating the effects of interventions for meta-analysis protocols — could also usefully include a mention of OIS estimation (in item #15c: "Describe any proposed additional analyses (such as sensitivity or subgroup analyses, meta-regression)") (18). Including this information in review protocols would ensure better adherence by enhancing the description of methodologies and of underlying statistical assumptions, including those affecting OIS estimation.

The GRADE (Grading of Recommendations, Development and Evaluations) Working Group considers that OIS estimation is relevant to evaluating quality of evidence, and especially imprecision, in meta-analysis results. It recommends, for instance, as follows: "If the total number of patients included in a review is less than the number of patients generated by a conventional sample size calculation for a single adequately powered trial, consider rating down for imprecision" (19). However, it also

points out that OIS estimation is not advisable when the outcome refers to a low event rate for a large sample size (19).

Statistical reporting guidelines like SAMPL (Statistical Analyses and Methods in the Published Literature) do not as yet include any mention of the OIS (20), nor does the more recent ROBIS (Risk of Bias in Systematic Reviews) tool (21).

6.4.2. Limitations, unanswered questions and future research

One limitation of the survey was that data were not collected on the profiles of the respondents (such as whether they were clinical practitioners, methodologists, etc), on their experience in conducting systematic reviews or the number they published per year. Future studies — including a qualitative approach based on focus groups — could profile authors and their reviews more clearly so as to better understand why they might reject or accept OIS estimation.

Another limitation was that, although the survey was anonymous, respondents may have thought that they could be identified indirectly; consequently, they may not have given accurate or honest answers that they felt might have placed them in an unfavourable light.

Even though the response rates for PRISMA Statement authors and non-PRISMA review authors were similar to those of other similar surveys, bias may have arisen due to missing data in relation to non-respondents to the survey, as only authors more familiar with the research topic (i.e. OIS) may have responded. A more specific area of missing data was the lower response rate observed for Q4 (“If you are not sure whether you would use this method what are the main reasons for this?”).

A limitation that affected another question in the survey (referred to in Section 6.4.1) was a confounding factor in relation to the phrasing of Q1 (“The following terms should be used in the

context of meta-analysis: trial sequential analysis and optimal information size") — despite the survey having been piloted, prior to issue, in 2 phases (first with 5 researchers who were specifically asked to provide feedback on content and then with 8 researchers). Since it is not entirely clear whether the respondents understood "terms" in the literal sense of 'terminology' or in the broader sense of 'concept', the responses to this question should be interpreted with care. This issue — which points to how small variations in the wording of a question may produce ambiguous or even unexpected results — would suggest that potential language interference needs to be taken into account in the design of international surveys. One way to avoid this problem would be to consult a bilingual expert who would ensure that questions were phrased clearly and unambiguously with international respondents in mind.

Finally, although the survey was piloted (with researchers with differing levels of experience of systematic reviews and meta-analyses) in order to check understanding of the survey and its individual questions, survey validity — i.e. whether the questions genuinely measured what they claimed to measure — was not psychometrically or statistically tested.

Although the above limitations may seem to undermine the absolute validity of the reported results, the survey does manage to reflect to what degree PRISMA Statement and review authors are aware of, and how they perceive, OIS.

6.5. Conclusions

A majority of PRISMA and non-PRISMA authors expressed an interest in knowing more about OIS, while mainly indicating a lack of familiarity with the methodology as a reason for not using this technique. PRISMA authors tended to be rather more sceptical about the benefits of OIS, with a third stating that OIS estimates would not add further information to a review. Nonetheless, the responses to the survey overall would indicate that the main concern regarding OIS is a perceived lack of methodological development of the technique. The fact that high proportions of PRISMA and non-PRISMA authors were of the opinion that meta-analyses may overestimate treatments or falsely identify false positive results would suggest that several broader issues would also need to be considered — most importantly, whether OIS would ultimately help make more informed clinical or health policy decisions.

6.6. Chapter summary

- There was an evident general interest in the OIS methodology among both PRISMA Statement authors and systematic review authors.
- A very high proportion of respondents were of the opinion that meta-analyses may overestimate treatment effects.
- PRISMA Statement authors were more sceptical about OIS than systematic review authors.
- The main reservations of PRISMA Statement authors appear to be related to the current methodologies available for OIS estimation.

6.7. Acknowledgements

Thanks to University of Oxford researchers from the Nuffield Department of Primary Care Health Sciences for their assistance with the pilot survey and to all authors who assisted with this study by responding to the online survey.

6.8. References

1. Ioannidis J. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. 2016;94(3):485–514.
2. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. 2009.
3. Petticrew M, Shekelle P, Stewart LA, Group P. meta-analysis protocols (PRISMA-P) 2015: elaboration. 2015;7647(January):1–25.
4. Hutton B, Catalá-López F, Moher D. La extensión de la declaración PRISMA para revisiones sistemáticas que incorporan metaanálisis en red: PRISMA-NMA. The PRISMA statement extension for systematic reviews incorporating network. Med Clin (Barc) 2016;147(6):262-6.
5. Zorzela L, Loke YK, Ioannidis JP, Golder S, Santaguida P, Altman DG, Moher D, Vohra S; PRISMA Harms Group. PRISMA harms checklist: improving harms reporting in systematic reviews. BMJ. 2016;352:i157.
6. Wetterslev J, Thorlund K, Brok J, Gluud C. Estimating required information size by quantifying diversity in random-effects model meta-analyses. BMC Med Res Methodol. 2009;9:86.
7. Thorlund K, Imberger G, Walsh M, Chu R, Gluud C, Wetterslev J, et al. The number of patients and events required to limit the risk of overestimation of intervention effects in meta-analysis: a simulation study. PLoS One [Internet]. 2011 Jan [cited 2014 Mar 24];6(10):e25491.
8. Thorlund K, Imberger G, Walsh M, Chu R, Gluud C, Wetterslev J, et al. The number of patients and events required to limit the risk of overestimation of intervention effects in meta-analysis--a simulation study. PLoS One [Internet]. 2011 Jan [cited 2014 May 17];6(10):e25491.

9. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ. 2009;339:b2535.
10. Wetterslev J, Jakobsen JC, Gluud C. Trial Sequential Analysis in systematic reviews with meta-analysis. BMC Med Res Methodol. 2017; 17(1):39.
11. Pogue J, Yusuf S. Overcoming the limitations of current meta-analysis of randomised controlled trials. Lancet. 1998; 351(9095):47-52.
12. Chalmers TC, Berrier J, Sacks HS, Levin H, Reitman D, Nagalingam R. Meta-analysis of clinical trials as a scientific discipline. II: Replicate variability and comparison of studies that agree and disagree. Stat Med. 1987;6(7):733-44.
13. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ. 2009; 339:b2700.
14. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol. 2009; 62(10):1006-12.
15. Kulinskaya E, Wood J. Trial sequential methods for meta-analysis. Res Synth Methods. 2014; 5(3):212-20.
16. Page MJ, Shamseer L, Altman DG, Tetzlaff J, Sampson M, Tricco AC, Catalá-López F, Li L, Reid EK, Sarkis-Onofre R, Moher D. Epidemiology and Reporting Characteristics of Systematic Reviews of Biomedical Research: A Cross-Sectional Study. PLoS Med. 2016;13(5):e1002028.
17. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med. 2009; 151(4):264-9.

18. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA; PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ. 2015; 349:g7647.
19. Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, Devereaux PJ, Montori VM, Freyschuss B, Vist G, Jaeschke R, Williams JW Jr, Murad MH, Sinclair D, Falck-Ytter Y, Meerpohl J, Whittington C, Thorlund K, Andrews J, Schünemann HJ. GRADE guidelines 6. Rating the quality of evidence--imprecision. J Clin Epidemiol. 2011; 64(12):1283-93.
20. Lang TA, Altman DG. Basic statistical reporting for articles published in biomedical journals: the "Statistical Analyses and Methods in the Published Literature" or the SAMPL Guidelines. Int J Nurs Stud. 2015; 52(1):5-9.
21. Whiting P, Savović J, Higgins JP, Caldwell DM, Reeves BC, Shea B, Davies P, Kleijnen J, Churchill R; ROBIS group. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. J Clin Epidemiol. 2016; 69:225-34.

Chapter 6. Appendix

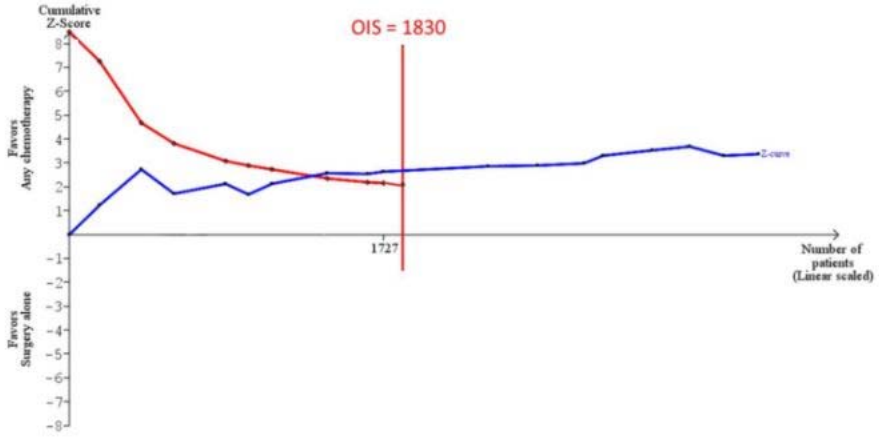
Table A6.1. Email and questionnaire

The use of optimal information size (OIS) in systematic reviews that use meta-analysis: An international survey.

Introduction

As you know, meta-analyses are used to provide more powerful, more precise estimates of the size of intervention effects. Optimum information size (OIS) is a concept first described by Pogue et al. in 1998 (Lancet. 1998; 351:47-52); it is defined as the minimum amount of information required to reach reliable conclusions in a meta-analysis. One suggestion is that the OIS is at least equal to the number of participants required for a single adequately powered trial. The methods most commonly used to estimate OIS are similar to those used for sample size calculation for a single randomized controlled trial, with adjustment for the expected heterogeneity of a meta-analysis.

Estimating the OIS should help to determine when we have sufficient data to draw reliable conclusions.



The above figure shows how evidence for a published meta-analysis accumulates over trials; the cumulative Z-curve (blue line) represents. The monitoring boundary (red line) can be used to determine when sufficient evidence has been obtained to indicate whether or not an intervention is effective. In this figure the

cumulative Z-curve crosses the monitoring boundary at the 8th trial; this indicates that enough evidence has been accumulated to conclude that the intervention is effective (the overall risk of type I error is less than 5%). After nine trials the OIS (1830 patients) has been reached.

Information sheet for participants

Project: The use of optimal information size (OIS) in systematic reviews that use meta-analysis: An international survey.

Dear Researcher (Participant)

If you accept to take part in this international survey, please ensure you have read all of the following sections correctly and thoroughly before electing to continue. Thank you for your time and participation.

This research intends to understand how sample sizes influence systematic reviews and meta-analyses. One approach to this is to use trial sequential methods to estimate the optimal information size (the number of participants needed to reliably detect an effect in a meta-analysis).

The aims are to evaluate how well-known these methods are and the reasons for their use or lack of use. The value of this research will be to describe the current worldwide knowledge about this methodology and the reasons for their use or lack of use in clinical research.

This research will involve you initially completing this online questionnaire, which will ask you basic questions regarding your knowledge and opinion about the optimal information size in a meta-analysis.

This research has targeted a sample from authors of systematic reviews that published their research on Medline during 2015, authors like yourself will have the option to participate in the survey or not.

As a voluntary participant you do have the right to withdraw at any point during the process. Participants may withdraw without penalty. Whether that is now or at a future point. There is no problem if you have a sudden change of mind just do not complete the survey. If you have put yourself forward for the interviews, then contact the investigator to inform them of your decision.

As a participant you can leave this research study at any time. As a participant you can provide the research team with the reason(s) for leaving this study, but is not required to provide your reason(s).

To take part all you have to do is complete the online questionnaire.

You may feel as if the researcher is judging your knowledge or opinions based on the answer you give; that is not true and full anonymity and confidentiality is guaranteed.

Participating in this study is a great chance to understand the knowledge's, penetration and acceptance of the optimal information size in meta-analysis. It is a relatively original study to take part in and the results should be very informative.

Every part of information given throughout this process by you the participant will remain strictly confidential, your name will not appear on the survey response and if you do reveal any personal information the data will be reviewed and anonymized. Only the investigator team of this research will have access to the data.

The data of this questionnaire is logically protected in agree with Google Forms privacy. Unauthorized parties cannot access this data. All questionnaires are protected by this secure architecture that ensures that one participant cannot see another participant's data. For additional details see: [https://support.google.com/work/answer/6057301].

The findings of this research will be used in a thesis and published as a paper in a journal. After the publication of data the results will be stored only for five years. After the publication of data the results will be stored only for five years and after this period all data will be destroyed.

This study has been reviewed by, and received ethics clearance through, the University of Oxford Central University Research Ethics Committee.

Contact information: Josep García (DPhil student)

QUESTIONNAIRE

Please respond to the items below by selecting the answer that best reflects your position.

1) The following terms should be used in the context of meta-analysis: trial sequential analysis and optimal information size

- o Strongly agree
- o Agree
- o Neither agree nor disagree
- o Disagree
- o Strongly disagree

If you answered "disagree" or "strongly disagree" please answer the following additional questions:

1a) Meta-analyses may overestimate treatment effects or falsely identify positive results (type I error)

- o Strongly agree
- o Agree
- o Neither agree nor disagree
- o Disagree
- o Strongly disagree

1b) I consider that the results of meta-analysis are conclusive

- o Strongly agree
- o Agree
- o Neither agree nor disagree
- o Disagree
- o Strongly disagree

1c) I would like to know more about optimal information size or trial sequential analysis before making up my mind

- o Strongly agree
- o Agree
- o Neither agree nor disagree
- o Disagree
- o Strongly disagree

2) Using optimal information size or similar methods to determine the required number of participants for a meta-analysis is a useful or necessary part of a systematic review

- o Strongly agree

- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

3) I would consider using optimal information size in my next systematic review

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

4) If you are not sure whether you would use this method what are the main reasons for this? (Please select all options that apply)

- I do not know enough about this statistic
- I am not confident that I could calculate it
- I do not think it would add any further information to the review
- I might consider using it if I knew more about it
- Other. Please specify

5) I am familiar with statistical developments in meta-analysis and with methods used to estimate optimal information size for meta-analyses

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

6a) Trial sequential analysis or optimal information size should be included in the PRISMA recommendations

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

6b) If the optimal information size were to be included in the PRISMA recommendations I would find it easy to incorporate it into my next review

- Strongly agree
- Agree
- Neither agree nor disagree
- Disagree
- Strongly disagree

Table A6.2. Ethics committee approval

MEDICAL SCIENCES
INTER-DIVISIONAL RESEARCH ETHICS COMMITTEE

Research Services, University of Oxford, Wellington Square, Oxford, OX1 2JD
Tel: +44(0)1865 616577 Fax: +44(0)1865 280467
ethics@medsci.ox.ac.uk

CONFIDENTIAL

Ref: R45973/RE001

6th July 2016

Mr Jose Maria Garcia-Alamino
Nuffield Department of Primary Care Health Sciences
University of Oxford
Radcliffe Observatory Quarter
Woodstock Road
Oxford

Dear Josep

Research Ethics Approval - CUREC 1

Project title: The use of optimal information size (OIS) in systematic reviews that use meta analysis: An international study

The above application has been considered on behalf of the Medical Sciences Inter-divisional Research Ethics Committee (IDREC) in accordance with the procedures laid down by the University for Ethical Approval of all research involving human participants.

I am pleased to inform you that, on the basis of the information provided to the IDREC, the proposed research has been judged as meeting appropriate ethical standards, and approval has been granted for a period of 3 months, commencing on 6th July 2016. The reference number for this project is **R45973/RE001**.

This is **subject to:**

a) **it is your responsibility to comply with the requirements for administering any tests or questionnaires and if in doubt to contact the publisher of those tests or questionnaires.**

b) **if new staff are engaged MS IDREC should be informed of their names, status and ethics training.**

Please may I remind you that your project may be reviewed at some stage during an annual audit of projects.

Amendments

Should there be any subsequent changes to the project, which raise ethical issues not covered in the original application, you should submit details to the IDREC for consideration and approval.

Please do not hesitate to contact me if you have any queries.

Yours Sincerely

Dr. Helen Barnby-Porritt
Research Ethics Manager, Medical Sciences

Table A6.3. Responses to Question 4e by PRISMA Statement authors

- Requires statement of an MCID [minimum clinically important difference] which we do not have good science to do. Similarly, it is restricted by the need to incorporate the observed heterogeneity, hence is a (partially) data dependent method.
- I am not sure how important this might be.
- While I recognise that sequential methods may have a role in prospective meta-analyses, they are inappropriate when applied retrospectively. Also, 'trial sequential analysis' and 'optimal information size' are really stupid terms for notions that might better be described as minimum information size and sequential meta-analysis.
- Generally, my data doesn't permit meta-analysis. I am interested in whether OIS can be used to determine when a meta-analysis should be updated. on Question 6, PRISMA is the minimum reporting standard.

Table A6.4. Responses to Question 4e by non-PRISMA Statement authors

- Reviewers are not familiar with this method. In my experience, they ask to remove it.
- Depends on the purpose of the SR [systematic review] – I see it as potentially useful when purpose is to determine efficacy/effectiveness of an intervention.
- My assumption in doing any meta-a/r worth doing would be that there is at least a handful of good trials. If one needed to use this then it tells me that the individual trials are small to begin with and the quality of the meta is then limited by non-optimal quality of the trials it contains. There are other methodological issues that may then be exacerbated – random-effects analysis may not be a cure as outlined in Cochrane Handbook.
- Estimating effect size and thus optimal size from the meta-analysis, where the included studies are heterogenous, can be difficult.
- Even if it may not add further information, it would be worth trying during the review.
- There are certainly areas where this is useful. But two counterarguments are i) making the best use of available evidence, which by definition is always incomplete; ii) problems with rare/niche research areas (e.g. uncommon cancers, MND [motor neuron disease], Duchenne's) where interventions are typically very small trials (I work in these areas).
- The trials included in many/most of my systematic reviews are often at high risk of bias or do not answer the question we posed therefore not been at point when can apply this statistic.
- GRADE has written about this concept, but for me to use it, I would like to see Cochrane endorse it.
- It depends on the specific objectives of the systematic review and meta-analysis.
- The approach is based on the presumption that one can define what might be considered a worthwhile effect. This is context specific and so cannot be defined in absolute terms. The output of interest here should be an estimate of effect size that a) will then b) used by a clinical or commissioning decision maker can use to inform their decisions. Thus, setting this in advance is not helpful. What may be more helpful is an approach that tells us we have enough information that no new trials will change overall conclusions, or where there are insufficient data how much more data are needed before results are conclusive; i.e. a post hoc consideration of what might be

needed. Finally, of course you cannot make a silk purse out of a sow's ear; if, as is often the case, the underlying trials are of very poor quality, it does not matter how clever one's meta-analytical techniques may be – the results still cannot be trusted.

- I do not like the focus on statistical significance. I view a meta-analysis as a tool to give me prediction interval of the treatment effect. That is all I need to know. OIS would merely confuse me and distract from the message I can give given the data. It is like performing a sample size calculation after performing the trial: bad practice.
- I do not favour using a single metric such as OIS in determining that a meta-analysis is conclusive.
- Meta-analysis is a retrospective investigation. Why should sample size estimation help?
- Aiming for a certain sample size in systematic reviews may cause researchers to be overly strict in inclusion, which may introduce bias and/or reduce generalisability. I think study inclusion/selection in every analysis should be as broad as possible within the context of that specific analysis in the interest of said generalisability and to reduce the risk of selection bias. Thus, all reported pooled outcomes should reflect ALL the RELEVANT literature. Reporting results of meta-analysis of a subset of the studies relevant to that analysis may be misleading, especially to readers unfamiliar with the methodology.

Chapter 7. Discussion and conclusions

7.1. Summary of research findings

In the healthcare sector, systematic reviews, which combine and analyse results for available randomized controlled trials (RCTs), are theoretically a reliable resource for decision-making regarding health interventions. Meta-analyses are frequently included in systematic reviews as a statistical approach to summarizing the results of individual RCTs. Nonetheless, as documented in the literature review (**Chapter 2**), their results are increasingly being called into question.

There is a growing body of evidence that indicates that statistically significant results in early meta-analyses have a high proportion of type I errors. A main cause of type I error in meta-analyses is statistical multiplicity; thus, the more analyses are done in a meta-analysis (each time a further RCT is included or another outcome is analysed), the more likely it is that some tests will be found to be statistically significant.

Another issue is statistical power, which is highly correlated with the cumulative number of patients and events. Empirical studies suggest that more evidence accumulated over time may point to many apparently large intervention effects being substantially overestimated in early studies. Thus, even though more RCTs are available and more patients are included, it is not clear that meta-analyses have the necessary power to yield conclusive results or to detect or refute even large intervention effects. Indeed, the problem of reliability is even acknowledged by authors, who often conclude

their systematic review with comments such as 'further RCTs are required before any recommendations can be made for the possible routine use of X intervention in patients with Y' or 'due to the small number of events, we cannot be certain whether X intervention results in a clinically important decreased risk of new events'. It is clear that there is a need to remove — or at least minimize — uncertainty regarding whether a sample is of a sufficient size to enable reliable conclusions to be drawn regarding a health intervention.

Although sample size calculation is well established for individual RCTs, in relation to meta-analyses this issue received little attention until publication of a seminal article by Pogue and Yusuf in 1997, proposing and describing the concept of optimal information size (OIS) as the minimum amount of information required for reliable conclusions to be drawn about an intervention. From this point, the literature has included a growing number of publications on what authors variously refer to as 'optimal information size', 'optimum information size', 'trial sequential analysis' or 'required information size'.

The OIS concept in practice has not been widely studied and documented, so little is known about it in terms of 2 main issues explored in this thesis. First, while OIS is calculated using 3 main statistical parameters — baseline risk (the control event rate, CER), heterogeneity (usually I^2 or D^2), and effect size (ES, usually expressed as relative risk, RR, or relative risk reduction, RRR) — no guidance is available regarding what values to use for those primary parameters or how they affect OIS. Second, little is known regarding the number of published reviews that include OIS estimates, trends in the use of OIS, awareness and perceptions of OIS in organizations that develop methodologies and publish guidelines on reviews, or familiarity with the concept among authors of systematic reviews.

Therefore, following up on Pogue and Yusuf's seminal publication of 1997, it was considered important to review uptake of OIS in published meta-analyses, analyse what methods and statistical

assumptions were used, and determine Cochrane Review Group (CRG) opinions regarding OIS (**Chapter 3**). Although only a very small proportion of meta-analyses published annually included OIS estimates (0.1%-0.8%), significantly, a very distinctive upward trend is evident from 2010. As for the statistical assumptions used for OIS estimates, no explicit information has been provided in many reviews regarding the underlying assumptions, nor is there any apparent consensus on which parameters or even values to use. Furthermore, although only a small number of CRGs have any specific policy regarding OIS estimates, around 40% of the meta-analyses that included OIS estimates were Cochrane reviews. However, in absolute terms, no increase was observed in the number of CRGs with an interest in OIS, and what interest there existed was concentrated in 2 specific CRGs and most especially in the Hepato-Biliary Group.

In order to assess what proportion of meta-analyses achieved OIS, a sample of 137 recently published meta-analyses in high-quality journals was investigated, together with the impact of heterogeneity and ES on OIS estimates (**Chapter 4**), broken down by type of systematic review (Cochrane review vs non-Cochrane review) and by outcome type (all-cause mortality, semi-objective outcome or subjective outcome). The fact that few of these recently published meta-analyses achieved OIS might suggest that they did not have adequate statistical power to allow firm conclusions to be drawn. Differences in heterogeneity were more marked in Cochrane reviews compared to non-Cochrane reviews. Evident also was wide variability — partially explained by outcome type — in the range of values used for the statistical parameters that impact on OIS estimates, with fewer than half of the primary outcomes synthesized in the meta-analyses achieving the OIS. The results demonstrate that — as well as heterogeneity, ES and CER— outcome type is also relevant to OIS estimation. To my knowledge, this is the first time that outcome type has been accounted for in an analysis of OIS estimates.

CER also is fundamental to estimating OIS, yet the literature does not document how this parameter might behave depending on RCT follow-up duration, nor is there any consensus regarding how to calculate CER for OIS estimation purposes. The impact of CER on follow-up duration was therefore analysed via a meta-epidemiological study of 236 RCTs included in 34 meta-analyses (**Chapter 5**). While no correlation was initially observed between standardized CER and standardized follow-up duration, sample size or ES, what correlation did exist tended to be positive, with larger CER values tending to be associated with longer follow-up durations. No clear and unequivocal relationship could ultimately be claimed, however; as happened with effect estimates for interventions, supporting estimates of CER varied, not only by intervention but also by population. Although there is obviously a need to ensure confidence in CER estimates, how exactly to obtain this estimate remains unclear.

Since the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement provides crucial guidance to authors of systematic reviews and meta-analyses, it was considered important to determine awareness and perceptions of OIS among PRISMA Statement authors and to contrast this with awareness and perceptions of OIS among systematic review authors (**Chapter 6**). PRISMA Statement authors, while acknowledging the conceptual importance of calculating OIS, were rather sceptical about OIS — perhaps because, as expert methodologists, they may have reservations regarding current methods available for OIS estimation. As for the non-PRISMA authors, a relatively high proportion wanted to know more about OIS, considered OIS to be useful and necessary and would consider using OIS in the future. It may be that these authors (not necessarily expert methodologists) objectively perceive the value of OIS or of any procedure that might enhance confidence in meta-analysis results.

7.2. Strengths and limitations

Strengths

The first strength of the thesis is that it is an extensive review of research to date into the relatively novel concept of OIS as used for systematic reviews with meta-analyses. Furthermore, the thesis analyses an aspect of meta-analyses not previously explored, which is the behaviour of the 3 main parameters for OIS estimation — namely, heterogeneity, ES and CER — according to the type of outcome analysed in the interventions. This analysis reveals that wide variability in the range of values used for the parameters that impact on OIS estimates is largely explained by the type of outcome.

To obtain information on general awareness of OIS and on policy regarding OIS use by benchmark bodies, surveys were administered to Cochrane Collaboration CRGs and PRISMA Statement authors. The Cochrane Collaboration was surveyed — on 2 different occasions in the last decade (2010 and 2016) — as the source of reputable systematic reviews, whereas PRISMA Statement authors were surveyed because adherence to the PRISMA Statement by authors wishing to publish systematic reviews with meta-analyses is prescribed by a growing number of scientific journals. Useful information was collected on the perceptions of OIS by these benchmark figures, particularly regarding the methodology and possible use of OIS in the future.

Limitations

One limitation of the thesis is its focus on a frequentist model of analysis in terms of calculating a sample size that would ensure adequate information size (number of patients) for a meta-analysis and that would, ultimately, ensure a reliable estimation of the intervention effect. The fact of focusing exclusively on OIS is because OIS involves relatively simple calculations and can therefore be easily implemented in systematic reviews with meta-analyses. Researchers are also generally

more familiar with frequentist approaches. Nonetheless, semi-Bayesian or fully Bayesian methods could also contribute useful information.

Another limitation refers to the survey of CRGs that, in the interest of brevity and of maximizing response rates, only included 4 questions. This meant that certain issues of potential interest were not analysed, such as the non-availability of guidance regarding how to estimate the OIS, which statistical methods to apply or proposals to promote the use of OIS estimation by Cochrane authors. Another issue was the poor response in 2016 compared to 2010, attributable to a new Cochrane policy that restricted possibilities for making direct contact with individual CRGs. This new policy would need to be taken into account in any future surveys of CRGs.

Another limitation was that the analysis of the impact of heterogeneity and ES on OIS estimates was confined to the top 5 medical journals. Reviews conducted by the Cochrane organization are considered to be of higher quality (1) and of greater methodological rigour than meta-analyses published in paper-based journals, yet Cochrane systematic reviews are not representative of systematic reviews published by non-Cochrane journals (2). Since this thesis only included meta-analyses from the top 5 medical journals, it may not be possible to extrapolate findings to meta-analyses published in other journals. That said, however, this approach would bias results towards better evidence being evaluated to what is currently being generated. The results reported here, furthermore, cannot be generalized to the rapidly growing area of evidence synthesis that is network meta-analysis (3), for which the statistical methodology to estimate OIS is less developed than for traditional meta-analyses (hence our exclusion of these studies from the thesis).

The fact that the analyses were conducted for binary outcomes (Chapter 4) or for the mortality outcome (Chapter 5) meant that a wide range of intervention areas was covered; however, the

included meta-analyses obviously may not be representative of all healthcare meta-analyses, so the findings may not be generalizable to meta-analyses with continuous outcomes.

The meta-epidemiological study combined data from 236 RCTs included in 34 meta-analyses. However, this small overall number of meta-analyses would suggest a lack of representativeness. Furthermore, certain imbalances within the sample also affect its representativeness and potentially introduced bias. Regarding the disease class variable, for instance, around 80% of meta-analyses referred to noncommunicable diseases, but mean CER for communicable diseases was more than double that for noncommunicable diseases. This difference was also observed for mean follow-up duration, which was, logically, longer for noncommunicable diseases compared to communicable diseases. As for intervention type, nearly three quarters of the reviews covered pharmacological interventions, for which mean CER was half that for nonpharmacological interventions.

Another issue was that a differentiated analysis of truncated RCTs (tRCTs) was not possible, as 95% of the included RCTs were non-truncated. It would be useful to evaluate the impact of early stopping on CER and the number of outcomes. An additional limitation was that only RCTs were included, despite the usefulness of observational studies in estimating baseline risk (4). It would therefore be of interest to analyse possible differences in baseline risk for RCTs in comparison to observational studies.

Regarding the survey addressed to PRISMA Statement authors and non-PRISMA authors of reviews, data were not collected on the profiles of the respondents (such as whether they were clinical practitioners, methodologists, etc), on their experience in conducting systematic reviews or on the number of reviews they published per year. Future studies — which could include a qualitative approach based on focus groups — could profile authors and their reviews more clearly so as to better understand their reasons for rejecting or accepting OIS estimation. Furthermore, although

the survey was anonymous, respondents may have thought that they might be identified indirectly and so may not have given answers that they felt might have placed them in an unfavourable light.

Even though the response rates for PRISMA Statement authors and non-PRISMA review authors were similar to those of other similar surveys, bias may have arisen due to missing data in relation to non-respondents to the survey, as only authors more familiar with the research topic (i.e. OIS) may have responded.

7.3. Implications of findings

Given the variability that exists in how OIS estimates are arrived at, further research is required into the statistical assumptions underlying OIS estimates and into how to determine CER, ES and heterogeneity values.

Although use of OIS in relation to the total number of published reviews remains anecdotal, there has been an evident rising trend since 2010, with 2015 alone accounting for almost a third of meta-analyses with OIS estimates published in the period 2006-2015. Given the growing interest in OIS, it would seem necessary for a benchmark body like the Cochrane Collaboration or PRISMA to provide explicit recommendations on OIS estimates — specifically regarding what statistical parameters to include and on how to calculate them. A complementary initiative could be to include a description of OIS calculation in PROSPERO (International Prospective Register of Systematic Reviews; www.crd.york.ac.uk/prospero), which, as a permanent register of systematic review protocols, allows completed reviews to be compared with their protocols (5). The inclusion of recommendations in journal editorial policies and instructions for authors would also be useful.

Regarding the software used, OIS estimation represents an area to be further developed, given that the included meta-analyses mainly used TSA software, developed by the Copenhagen Trial Unit (CTU) in Denmark, which has made an important contribution to increasing knowledge and

improving understanding of the OIS concept. However, its software adopts a frequentist approach, so it would be of interest to see the development of software that could be used for Bayesian and semi-Bayesian approaches. Another important consideration with TSA software is the poor quality of its graphic output. R, in contrast, produces excellent graphs and could be adapted to OIS estimation (6) using either frequentist or Bayesian approaches. It is also suggested that OIS calculation could be included in the Cochrane RevMan software (http://community.cochrane.org/tools/review-production-tools/revman-5) provided as a tool for preparing and maintaining reviews.

In relation to the CRG survey for 2016, it would be useful to re-run it — provided a higher response rate could be ensured — for comparison with the results from the 2010 survey, when the response rate was close to 70%. This could be done in partnership with the Cochrane Collaboration or, preferably, could be promoted to CRGs and run directly by the Cochrane Collaboration. It would also be useful to include the Cochrane Methods Groups (CMGs) in the survey so as to be able compare and analyse differences in responses from CRGs and CMGs.

Another interesting avenue for exploration would be to apply a methodology to mine author relationship networks for systematic reviews with OIS estimates (both Cochrane and non-Cochrane) so as to study relationships between authors.

This thesis has shown that outcome type can be used as a proxy to define the basic required parameters (CER, ES and heterogeneity) for an estimate of the OIS for the primary outcome, independently of the confidence regarding those specific parameters as obtained from the review. Reviewers are therefore encouraged to use sample size estimates as a measure of likely confidence in their results. The fact that over 50% of primary outcomes in recent systematic reviews appear to

fall below this minimum requirement of information points to the need for further evidence aimed at reducing uncertainty.

As future research, it would be of interest to explore predictive distributions for between-study heterogeneity for meta-analyses from a broad sample of journals, so as to replicate what has been done for Cochrane systematic reviews (7). It would also be useful to further explore ES behaviour according to the quality of individual RCTs included in a meta-analysis (high-quality vs low-quality RCTs), given that low-quality RCTs have been shown to overestimate ES (8).

Another research topic of interest would be to analyse how ES varies as new RCTs are included in a meta-analysis, given the observation that ES in cumulative meta-analyses varies over time (9) and may even decay, i.e. go from a large ES for initial results to a smaller ES for later results. The danger of using a low ES based on initial results is that further RCTs may not be implemented because of an erroneous perception that OIS has been achieved.

7.3.1. Sequential methods to estimate sample size for meta-analyses

While most Cochrane reviews identify residual uncertainty and are a rich source of suggestions for further healthcare research (10), only 3.2% of reviews explicitly suggest that no further research is necessary or feasible. OIS estimation, however, can contribute information to clarify what further research, if any, is needed.

The current scenario is a challenging one for several reasons: the number of meta-analyses applying OIS estimates or using the trial sequential analysis methodology has increased in recent years; the Cochrane Collaboration has included a mention of the OIS in the Cochrane Handbook (11); and new methods based on semi-Bayesian or Bayesian approaches are being developed (12). The information compiled in this thesis would indicate that further research is required into OIS and the use of sequential methods for meta-analyses, but also into approaches, given that the problem with

frequentist approaches is how to obtain reliable values for CER, ES and heterogeneity. The availability of prior distributions of ES and heterogeneity values, for instance, would foster the development and extension of semi-Bayesian and Bayesian approaches.

Another challenge related to OIS estimation is the 'moving target phenomenon' (13): since the CER, ES and heterogeneity values vary each time a new RCT is included in a meta-analysis, any previously calculated OIS has to be modified. Further research is therefore required into how the underlying statistical parameters should be calculated (14). In relation to heterogeneity, a recent study has reported that around a third of between-trial heterogeneity might be explained by trial design characteristics (15). The use of prediction intervals based on outcome type could be a good option for estimating heterogeneity that could be applied to OIS calculations. It would also be of interest to analyse whether the quality of RCTs as well as outcome type affect heterogeneity calculations. Regarding ES, it has been observed that estimates vary due to temporal trends (8,16). This thesis has shown that setting a standard ES of 5% is inappropriate (at least for the mortality outcome), as it would over- or underestimate the real ES. Finally, regarding CER, it would be interesting to further investigate how this parameter behaves according to sample size, in order to establish how baseline risk influences estimates of OIS. The most prudent approach to determining CER has been to use median CER for all RCTs included in a meta-analysis — as proposed by Hayden (17) — but it could be of interest to investigate and compare CER for high-quality vs poor-quality RCTs.

Other parameters that also need further investigation are the alpha and beta values governing type I and II errors. In much clinical research, values of $\alpha = 0.05$ and $\beta = 0.2$ are standard (18,19). However, some studies question these values and consider that they should be adapted to take the particular characteristics of each RCT into account (20).

In light of the results reported in this thesis, it would seem that a prudent approach to OIS estimation would be to estimate a higher OIS as a more conservative value, and a lower OIS as a less conservative value.

An interesting alternative that has emerged in recent years is the application of a non-frequentist — semi-Bayesian or Bayesian — approach to sequential meta-analysis (6,12). This alternative may be especially suitable for meta-analyses with fewer than 300 events.

Another issue that needs to be debated is whether OIS calculation would be appropriate or useful in the case of meta-analyses that report negative results (no effect) for the intervention group. It would also be necessary to understand why a meta-analysis failed to report positive results in favour of an intervention: was this due to a lack of power or was the intervention genuinely ineffective? To rule out the possibility of a lack of power, it would be useful to analyse the power of the meta-analysis — as described by Imberger et al. (21,22) — using TSA software, so as to highlight possible futility as proposed by Wetterslev et al. (23).

7.3.2. Acceptability of OIS estimation

One of the main reasons for not using OIS estimation and for its non-inclusion in the PRISMA Statement appears to be non-acceptance of the methodology, as deduced from responses to the survey of the CRGs and the survey addressed to PRISMA Statement authors and systematic review authors. Yet it is interesting to note a clear upward trend from 2010 — and most especially from 2015 — in the number of published meta-analyses that include OIS estimates. If this trend is maintained, it would seem logical and necessary to establish some guidelines or recommendations on how to estimate and report OIS values.

A recent Cochrane Collaboration development (in June 2017, i.e. after the data for this study was compiled) has been the inclusion of a recommendation regarding OIS in the latest version of the

Cochrane Handbook (version 5.2.0). The recommendation — similar to that by GRADE (Grading of Recommendations Assessment, Development and Evaluation) — is phrased as follows: “Authors can use the optimal information size (OIS) to make judgments about imprecision. The OIS is calculated on the basis of the number of participants required for an adequately powered individual study”. More specific guidelines are as follows: “If the 95% confidence interval (CI) excludes a relative risk (RR) of 1.0, and the total number of events exceeds the OIS criterion, precision is adequate. If the 95% confidence interval includes appreciable benefit or harm (an RR of 0.75 or over 1.25 is often suggested as a rough guide), downgrading for imprecision may be appropriate even if OIS criteria are met.” (Chapter 11.2.2). However, no explicit guidance is provided regarding how to calculate heterogeneity, ES or CER.

7.4. Conclusions

The use of OIS in systematic reviews with meta-analyses conceptually seems to be necessary as a measure of when sufficient information regarding an intervention has been obtained. Not only would this information ensure greater confidence in results, it would also contribute to a more rational use of resources and avoid unnecessary interventions in patients by reducing RCT durations.

7.5. References

1. Turner T, Green S, Tovey D, McDonald S, Soares-Weiser K, Pestridge C, Elliott J; Project Transform Team; IKMD developers. Producing Cochrane systematic reviews-a qualitative study of current approaches and opportunities for innovation and improvement. Syst Rev. 2017;6(1):147. doi: 10.1186/s13643-017-0542-3.
2. Page MJ, Shamseer L, Altman DG, Tetzlaff J, Sampson M, Tricco AC, Catalá-López F, Li L, Reid EK, Sarkis-Onofre R, Moher D. Epidemiology and Reporting Characteristics of Systematic Reviews of Biomedical Research: A Cross-Sectional Study. PLoS Med.;13(5): e1002028.
3. Thorlund K, Mills EJ. Sample size and power considerations in network meta-analysis. Syst Rev. 2012;1:41. doi: 10.1186/2046-4053-1-41.
4. Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. Trials. 2015;16:495.
5. Sideri S, Papageorgiou SN, Eliades T. Registration in the international prospective register of systematic reviews (PROSPERO) of systematic review protocols was associated with increased review quality. J Clin Epidemiol. 2018 Jan 12. pii: S0895-4356(17)30888-0. doi: 10.1016/j.jclinepi.2018.01.003.
6. Spence GT, Steinsaltz D, Fanshawe TR. A Bayesian approach to sequential meta-analysis. Stat Med. 2016; 35(29):5356-5375.
7. Turner RM, Davey J, Clarke MJ, et al. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the cochrane database of systematic reviews. Int J Epidemiol 2012;41:818–27.
8. Zhang Z, Xu X, Ni H. Small studies may overestimate the effect sizes in critical care meta-analyses: a meta-epidemiological study. Crit Care. 2013;17(1): R2. doi: 10.1186/cc11919.

9. Fanshawe TR, Shaw LF, Spence GT. A large-scale assessment of temporal trends in meta-analyses using systematic review reports from the Cochrane Library. Res Synth Methods. 2017;8(4):404-415. doi: 10.1002/jrsm.1238.
10. Clarke L, Clarke M, Clarke T. How useful are Cochrane reviews in identifying research needs? J Health Serv Res Policy. 2007;12(2):101-3.
11. Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011.
12. Higgins JP, Whitehead A, Simmonds M. Sequential methods for random-effects meta-analysis. Stat Med. 2011;30(9):903-21. doi: 10.1002/sim.4088.
13. Wetterslev J, Jakobsen JC, Gluud C. Trial Sequential Analysis in systematic reviews with meta-analysis. BMC Med Res Methodol. 2017;17(1):39. doi: 10.1186/s12874-017-0315-7.
14. Garcia-Alamino JM, Bankhead C, Heneghan C, Pidduck N, Perera R. Impact of heterogeneity and effect size on the estimation of the optimal information size: analysis of recently published meta-analyses. BMJ Open. 2017;7(11): e015888. doi: 10.1136/bmjopen-2017-015888.
15. Rhodes KM, Turner RM, Savović J, Jones HE, Mawdsley D, Higgins JPT. Between-trial heterogeneity in meta-analyses may be partially explained by reported design characteristics. J Clin Epidemiol. 2018;95:45-54. doi: 10.1016/j.jclinepi.2017.11.025.
16. Hoffmann TC, Thomas ST, Shin PN, Glasziou PP. Cross-sectional analysis of the reporting of continuous outcome measures and clinical significance of results in randomized trials of non-pharmacological interventions. Trials. 2014;15:362. doi: 10.1186/1745-6215-15-362.
17. Hayden JA, Tomlinson G, Ni A, et al. Approaches to estimate and present baseline risks considerations for Cochrane reviews evidence-informed practice integration of: 19th Cochrane Colloquium Madrid, 2011.
18. Piantadosi S. Clinical trials: a methodologic perspective. 2nd ed. Hoboken, NJ: Wiley; 2005.

19. Meinert CL. Clinical trials: design, conduct, and analysis. Oxford, UK: Oxford University Press; 1986).

20. Ioannidis JP, Hozo I, Djulbegovic B. Optimal type I and type II error pairs when the available sample size is fixed. J Clin Epidemiol. 2013;66(8):903-910.e2. doi: 10.1016/j.jclinepi.2013.03.002.

21. Imberger G, Thorlund K, Gluud C, Wetterslev J. False-positive findings in Cochrane meta-analyses with and without application of trial sequential analysis: an empirical review. BMJ Open. 2016;6(8):e011890. doi: 10.1136/bmjopen-2016-011890.

22. Imberger G, Gluud C, Boylan J, Wetterslev J. Systematic Reviews of Anesthesiologic Interventions Reported as Statistically Significant: Problems with Power, Precision, and Type 1 Error Protection. Anesth Analg. 2015;121(6):1611-22.

23. Jakobsen JC, Wetterslev J, Winkel P, Lange T, Gluud C. Thresholds for statistical and clinical significance in systematic reviews with meta-analytic methods. BMC Med Res Methodol. 2014;14:120. doi: 10.1186/1471-2288-14-120.

PUBLICATIONS

BMJ Open

Impact of heterogeneity and effect size on the estimation of the optimal information size: analysis of recently published meta-analyses

Josep M Garcia-Alamino, Clare Bankhead, Carl Heneghan, Nicola Pidduck, Rafael Perera

To cite: Garcia-Alamino JM, Bankhead C, Heneghan C, *et al*. Impact of heterogeneity and effect size on the estimation of the optimal information size: analysis of recently published meta-analyses. *BMJ Open* 2017;**7**:e015888. doi:10.1136/bmjopen-2017-015888

► Prepublication history for this paper is available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2017-015888).

Received 17 January 2017
Revised 11 July 2017
Accepted 14 July 2017

Nuffield Department of Primary Care Health Sciences, University of Oxford, Oxford, UK

Correspondence to
Dr Rafael Perera;
rafael.perera@phc.ox.ac.uk

ABSTRACT

Objective To estimate the proportion of systematic reviews that meet the optimal information size (OIS) and assess the impact heterogeneity and effect size have on the OIS estimate by type of outcome (eg, mortality, semiobjective or subjective).

Methods We carried out searches of Medline and Cochrane to retrieve meta-analyses published in systematic reviews from 2010 to 2012. We estimated the OIS using *Trial Sequential Analysis* software (TSA V.0.9) and based on several heterogeneity and effect size scenarios, stratifying by type of outcome (mortality/semiobjective/subjective) and by Cochrane/non-Cochrane reviews.

Results We included 137 meta-analyses out of 218 (63%) potential systematic reviews (one meta-analysis from each systematic review). Of these reviews, 83 (61%) were Cochrane and 54 (39%) non-Cochrane. The Cochrane reviews included a mean of 6.5 (SD 6.1) studies and the non-Cochrane included a mean of 13.2 (SD 10.2) studies. The mean number of patients was 2619.1 (SD 6245.8 or median 586.0) for the Cochrane and 19 888.5 (SD 32 925.7 or median 6566.5) patients for the non-Cochrane reviews. The percentage of systematic reviews that achieved the OIS for all-cause mortality outcome were 0% Cochrane and 25% for non-Cochrane reviews; for semiobjective outcome 17% for Cochrane and 46% for non-Cochrane reviews and for subjective outcome 45% for Cochrane and 72% for non-Cochrane reviews.

Conclusions The number of systematic reviews that meet an optimal information size is low and varies depending on the type of outcome and the type of publication. Less than half of primary outcomes synthesised in systematic reviews achieve the OIS, and therefore the conclusions are subject to substantial uncertainty.

Strengths and limitations of this study

- ► To our knowledge, this is the first analysis to estimate the optimal information size by type of outcome.
- ► This study includes only systematic reviews from the Cochrane library and the top five general medical journals; therefore, our results may not be generalisable to systematic reviews published in other journals.

INTRODUCTION

The concept of optimum information size (OIS) was first proposed in 1998 by Pogue *et al*[1 2] as 'the minimum amount of information required in the collective literature for reliable conclusions about an intervention to be reached'. This OIS estimate is based on standard sample size calculations. For example, the required number of participants (information size) for a meta-analysis should match those required in an adequately powered single trial.[3] Other measures of information size have been proposed[4 5]; however, the OIS involves a relatively simple calculation, which under some scenarios will underestimate the information required to define whether firm evidence has been reached to draw robust conclusions.[6] Brok *et al*[3] demonstrated, in a subset of Cochrane reviews, that many meta-analyses have false-positive results due to insufficient information, and Turner *et al* showed that most meta-analysis do not have sufficient power to identify even moderate effects.[7 8]

Sample size calculation and the OIS are influenced by several variables such as the control event rate (CER) (baseline risk), effect size, the power and the alpha value. Deciding on which values to use can be difficult and is typically based on values observed or estimated from the meta-analysis or one of the included studies. In addition, increased variation can also effect the estimate of the OIS, and there is currently no consensus about which value of heterogeneity should be used to calculate the OIS.

The OIS can help determine the stability of an effect and whether treatment effect estimates are likely to differ based on further information. However, they are difficult to

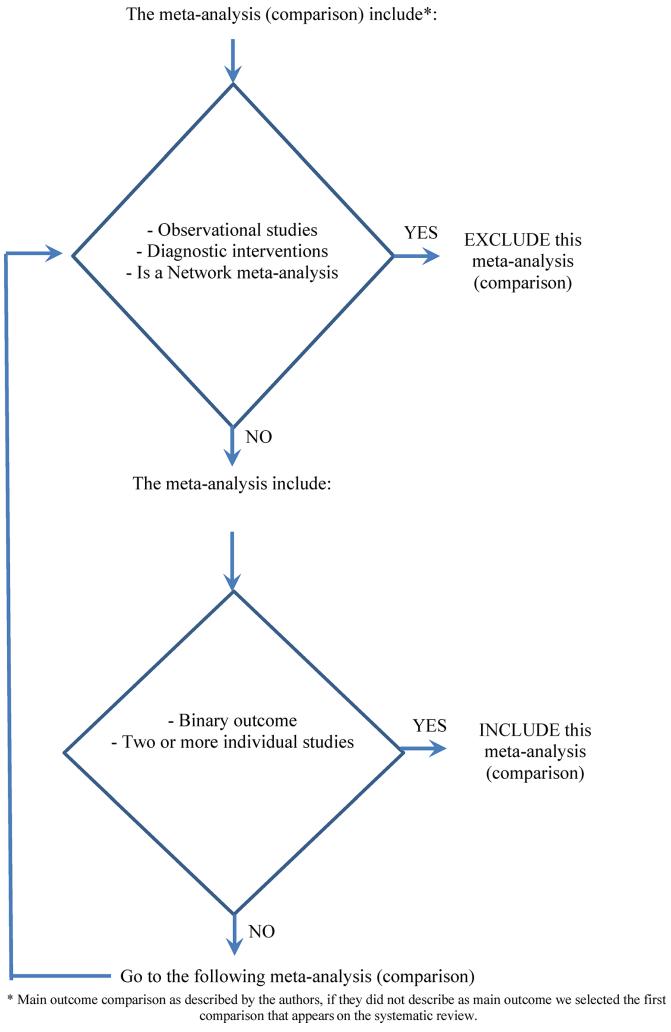


Figure 1 Algorithm for the selection of the meta-analysis (main comparison) in the systematic review.

define in advance, and there is no consensus regarding the alpha (significance) or power value used at the outset.

It is therefore not currently known if evidence accumulation and its associated OIS depend on the type of outcome studied and if this varies by publication type (Cochrane or non-Cochrane review). Therefore, we set out to quantify this by studying systematic reviews (SRs) published in the Cochrane Library and the top five general medical journals and in the process describe the impact that observed variation in heterogeneity and effect size (relative risk reduction (RRR)) have on the OIS estimation.

METHODS

We defined two sets of SRs to evaluate: Cochrane and non-Cochrane. We identified all Cochrane SRs published during 2010–2012 through the Archie Database (http://archie.cochrane.org), which contains all Cochrane published reviews and allows electronic searching. We randomly selected a total of 120 of these based on random numbers generated using Microsoft Excel for inclusion.

To search for non-Cochrane reviews, we identified all SRs with meta-analyses published in the top five general medical journals (*The New England Journal of Medicine*, *Lancet*, *The Journal of the American Medical Association*, *Internal Medicine*, *Annals of Internal Medicine* and *British Medical Journal*) using the following search strategy in Medline (PubMed): 'BMJ'[Journal] OR 'Ann Intern Med'[Journal] OR 'JAMA'[Journal] OR 'Lancet'[Journal] OR 'N Engl J Med'[Journal] AND (systematic review [ti] OR meta-analysis [pt] OR meta-analysis [ti]) restricted to SRs published during 2010–2012.

Inclusion/exclusion criteria

From all the selected Cochrane and non-Cochrane reviews, we included one meta-analysis from each. Based on the order the outcomes were reported (eg, outcome 1.1 for Cochrane SRs), we selected the first outcome presented in the meta-analysis that was based on: binary data from two or more individual studies (clinical trials or randomised controlled trials). If the first outcome did not meet this inclusion criteria, we continued through the listed outcomes until one was identified or we had exhausted the list of outcomes reported (figure 1). Meta-analyses that included observational studies, of diagnostic interventions or that were based on network meta-analysis were excluded. Meta-analyses showing no effect (pooled effect=1) or meta-analyses with no events in all included trials were also excluded.

Data extraction

Full texts were obtained for those abstracts that met the inclusion criteria and assessed for eligibility. One reviewer JMG-A extracted the data, and a second reviewer (RP or NP) checked the data. We developed customised Excel spreadsheets for the data extraction process. From each included meta-analysis, we extracted and calculated the following items: outcome type as defined by Turner[9] ('all cause mortality', 'semi-objective' (cause-specific mortality, major morbidity event) and 'subjective' (pain, mental health outcomes)), comparison, number of included patients, number of trials, number of events in each arm, CER, effect size and heterogeneity.

Analyses

We extracted data from each trial and repeated the meta-analysis using random-effects models (DerSimonian and Laird) to account for potential heterogeneity of effects. Estimates for trials with only one group reporting zero events were adjusted with a constant continuity adjustment of 0.5 in each arm (default adjustment in Revman). The obtained estimates for the pooled effect (eg, RR) and I^2 were compared with the published results to detect any relevant disagreement, and if required, the analyses were repeated to identify the source of the difference. Meta-analyses and calculation of the OIS were done using *Trial Sequential Analysis* software (TSA V.0.9)[10] freely downloadable at www.ctu.dk/tsa. The TSA software allows meta-analysis of dichotomous or continuous data under

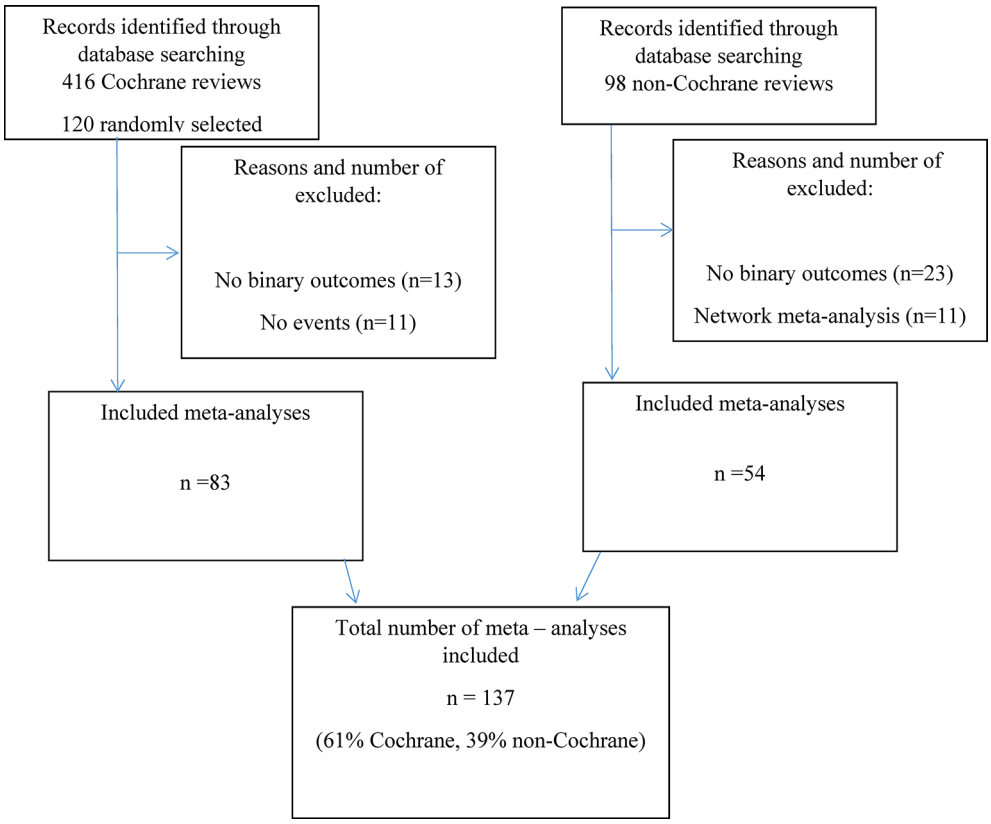


Figure 2 Flow chart identification of Cochrane and non-Cochrane systematic reviews.

fixed or random-effects models and has the option to estimate an information size and the stopping boundary. This estimation of the OIS is based on the alpha spending method (O'Brien Fleming and Lan-DeMets).

To evaluate the impact of changes in heterogeneity and effect size (RRR), we estimated the OIS under different scenarios. For heterogeneity, we analysed three values of heterogeneity: 'heterogeneity=rep' as that reported in the meta-analysis using a random-effects model (or obtained from fitting a random effects model if a fixed effect model was used originally), 'heterogeneity=0' and 'heterogeneity=Q3' (upper quartile or 75th percentile), which was determined based on estimates of predictive distributions published by Rhodes *et al*.[11] These two estimates of 'heterogeneity=Q3' and 'heterogeneity=0' were chosen as extreme scenarios to evaluate the impact that this parameter has on the OIS. Consistent with Rhodes *et al*, the estimation of the OIS took into account the outcome type: 'all cause mortality', 'semiobjective' (cause-specific mortality and major morbidity event) and 'subjective' (pain and mental health outcomes) and, for simplicity, was based on assuming an average mean study size between 50 and 200 participants.

To evaluate the impact of effect size on the OIS, we used two different estimates of the effect size for the meta-analyses with mortality outcome: the RRR obtained in each meta-analysis as well as an a priori conservative value of 5% for the RRR as reported by Djulbegovic *et al*.[12] For the transformation of relative risk (RR) measure to RRR, we used the following formula RRR=1 RR. If the RR was greater than 1, we used the RRR as a negative value. We did not determine an alternative estimate for the effect size for the other two outcomes (semiobjective and subjective) as the distribution of possible effects makes the choice of 'average effect' difficult to justify. We used only one value, per meta-analysis, for the baseline risk or CER. This was taken to be the median of the proportion of events in the included trials in each meta-analysis, following the method proposed by Hayden *et al*.[13]

We used descriptive statistics and plots to quantify differences in CER, effect size and heterogeneity between Cochrane and non-Cochrane reviews, stratified by type of outcome (six groups in total). We also determined the proportion of reviews that have achieved the OIS based on reported results and our two extreme scenarios 'heterogeneity=0' and 'heterogeneity=Q3' comparing between Cochrane and non-Cochrane and again stratifying by type of outcome ('all-cause mortality', 'semiobjective' and 'subjective' outcomes).

The descriptive analysis of the characteristics of included meta-analyses was carried out using SPSS V.22 software.

RESULTS

Search results

Figure 2 presents a flow chart of the results. We excluded 11 Cochrane SRs due to no events reported in the included trials or due to only one study being included in the review. We included a total of 137 meta-analyses out of 218 (63%)

Table 1 Descriptive results of included meta-analyses by type of outcome and intervention

	Cochrane (n=83)	Non-Cochrane (n=54)	All reviews (n=137)
	% (n/N)	% (n/N)	% (n/N)
Type of outcome			
All cause mortality	16.8 (14/83)	22.2 (12/54)	19 (26/137)
Semiobjective	21.7 (18/83)	44.4 (24/54)	30.7 (42/137)
Subjective	61.4 (51/83)	33.3 (18/54)	50.4 (69/137)
Type of intervention			
Pharmacological	56.6 (47/83)	62.9 (34/54)	59.1 (81/137)
Non-pharmacological	44.6 (36/83)	35.2 (20/54)	40.9 (56/137)

Table 2 Descriptive results for the statistical assumptions in the included meta-analyses

	CER	RRR*	Heterogeneity (I^2)	Included patients	OIS estimated†
All reviews (n=137)					
Mean (SD)	26.9 (26.1)	28.2 31.5)	20.4 (26.1)	9426.0 (22 753.9)	386 441.1 (1 645 397.1)
Cochrane reviews (n=83)					
Mean (SD)	24.0 (27.9)	21.0 36.6)	0.0 (25.5)	586 (6245.8)	2301.0 (1 422 086.6)
Non-Cochrane reviews (n=54)					
Mean (SD)	10.0 (21.7)	20.0 20.6)	14.5 (26.9)	6566.5 (32 925.7)	7299.5 (1 946 750.2)
All cause mortality (n=26)					
Mean (SD)	12.7 (16.5)	20.3 22.1)	10.9 (17.6)	14 314.6 (25 880.1)	499 090.7 (1 966 940.8)
Cochrane reviews (n=14)					
Mean (SD)	10.0 (11.2)	25.5 27.3)	6.7 (13.7)	6902.5 (12 971.1)	813 301.7 (2 678 677.7)
Non-Cochrane reviews (n=12)					
Mean (SD)	15.8 (21.4)	14.3 12.6)	15.7 (20.8)	22 962.1 (34 232.8)	132 511.2 (201 708.8)
Semiobjective (n=42)					
Mean (SD)	17.2 (20.9)	19.0 18.7)	18.8 (25.9)	18 683.5 (33 125.7)	828 450.9 (2 444 468.1)
Cochrane reviews (n=18)					
Mean (SD)	18.5 (10.5)	21.7 24.0)	12.8 (22.7)	3384.8 (4762.7)	536 616.8 (1 803 091.2)
Non-Cochrane reviews (n=24)					
Mean (SD)	16.3 20.0)	16.9 13.7)	23.3 (27.5)	30 157.0 (40 233.9)	1 047 326.5 (2 851 698.9)
Subjective (n=69)					
Mean (SD)	38.2 (27.1)	36.8 (37.9)	24.9 (28.1)	1948.9 (2971.0)	74 944.0 (406 792.0)
Cochrane reviews (n=51)					
Mean (SD)	41.6 (27.9)	36.5 (41.6)	23.5 (27.6)	1173.0 (2244.3)	78 688.6 (449 189.7)
Non-Cochrane reviews (n=18)					
Mean (SD)	28.5 (22.9)	37.6 (25.3)	28.9 (29.7)	4147.3 (3683.7)	64 334.2 (261 366.6)

*For the calculation of this descriptive variable all the values were considered as positive.
†This estimation of the OIS was done under the conditions of the 'scenario 2' (heterogeneity=0, alpha 5%).
CER, control event rate; OIS, optimum information size; RRR, relative risk reduction.

potential SRs (figure 2): 83 (61%) were Cochrane SRs and 54 (39%) non-Cochrane SRs. The Cochrane reviews included a mean of 6.5 (SD 6.1) studies and the non-Cochrane included a mean of 13.2 (SD 10.2) studies. The number of patients was 2619.1 (SD 6245.8 or median 586.0) for the Cochrane and 19 888.5 (SD 32925.7 or median 6566.5) patients for the non-Cochrane reviews.

Scenarios under different parameter estimates

Table 1 provides results on the types of outcomes and type of intervention studied for the included meta-analyses by publication type.

Of the included meta-analyses, 26 (19%) used 'all cause mortality' as an outcome; 42 (31%) were based on 'semiobjective' outcomes and 69 (50%) on a 'subjective'

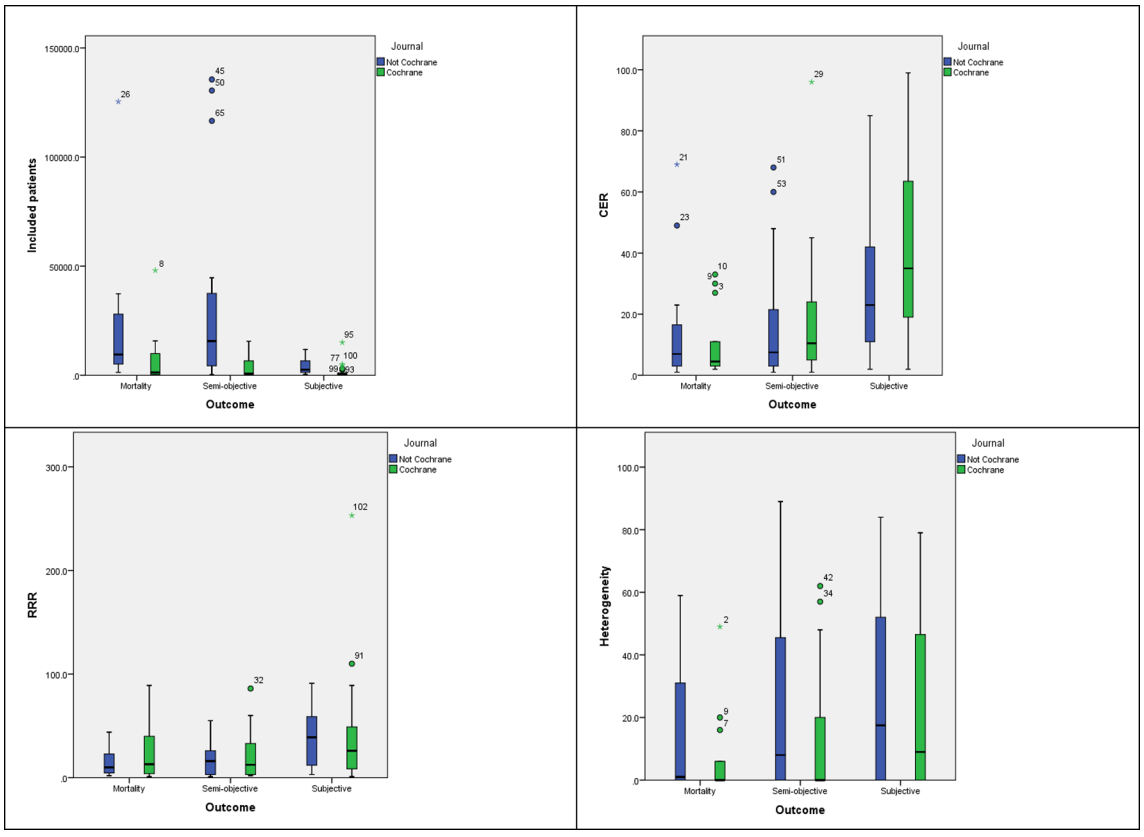


Figure 3 Systematic review characteristics stratified by source (Cochrane vs non-Cochrane) and type of outcome. CER, control event rate; RRR, relative risk reduction.

outcome. The type of intervention was pharmacological in 59% of the meta-analyses. There were significant differences in the type of outcome reported by publication type (χ^2; 2df=11.15, p=0.004) but not in the type of intervention reported (χ^2; 1df=0.54, p=0.46).

The descriptive analysis of the different parameter estimates (CER, RRR and I^2) used in the calculation of the OIS showed considerable variation depending on the type of outcome (table 2 and figure 3).

The number of included patients was higher in non-Cochrane reviews for all outcomes analysed (table 2). The CER for 'all cause mortality' had the lowest mean value and the distribution differed between outcome types. For RRR, the highest mean value and heterogeneity was observed for 'subjective' outcomes (table 2 and figure 3).

Meta-analyses that reached the OIS

Figure 4A presents the estimated OIS for each meta-analysis in the extreme scenario of no heterogeneity. All-cause mortality required the highest OIS for both types of reviews. However, this was only marginally higher than 'semiobjective' outcomes. For 'subjective' outcomes, OIS estimates are considerably smaller due to higher CERs and RRR. Figure 4B shows the number of meta-analyses that have achieved sample sizes equal or higher to the estimated OIS with more non-Cochrane reviews achieving this estimate (see figure 4C).

Estimation of the OIS based on reported heterogeneity shows that the necessary sample was only reduced for Cochrane SR reporting subjective outcomes (table 3). Further increasing the level of heterogeneity (worst-case scenario: heterogeneity=Q3) did not substantially change the proportion of meta-analyses achieving the OIS.

When using a more stringent estimate for the effect size (5% RRR) for 'all cause mortality', none of the identified meta-analyses had achieved the necessary sample size to meet the OIS (0/14 Cochrane and 0/12 non-Cochrane). Box presents five examples of meta-analyses reporting 'all cause mortality' as an illustration of SRs where the OIS has been reached and where it has not.

DISCUSSION

Our results show that there is wide variability in the range of values that impact on the OIS calculation: effect size (RRR), heterogeneity (I^2) and CER, regardless of source (Cochrane or non-Cochrane). This variability is partially explained by the type of outcome ('all cause mortality', 'semiobjective' or 'subjective') evaluated.

OIS estimates could therefore be obtained from different types of outcomes, as previously proposed by Turner *et al*[9] and Rhodes *et al*.[11] To our knowledge, this is the first time that accounting for the type of outcome in the estimation of the OIS has been proposed. We also found that the type of outcome impacts on the range of heterogeneity observed and was particularly high for 'subjective' outcomes. One possible explanation for this

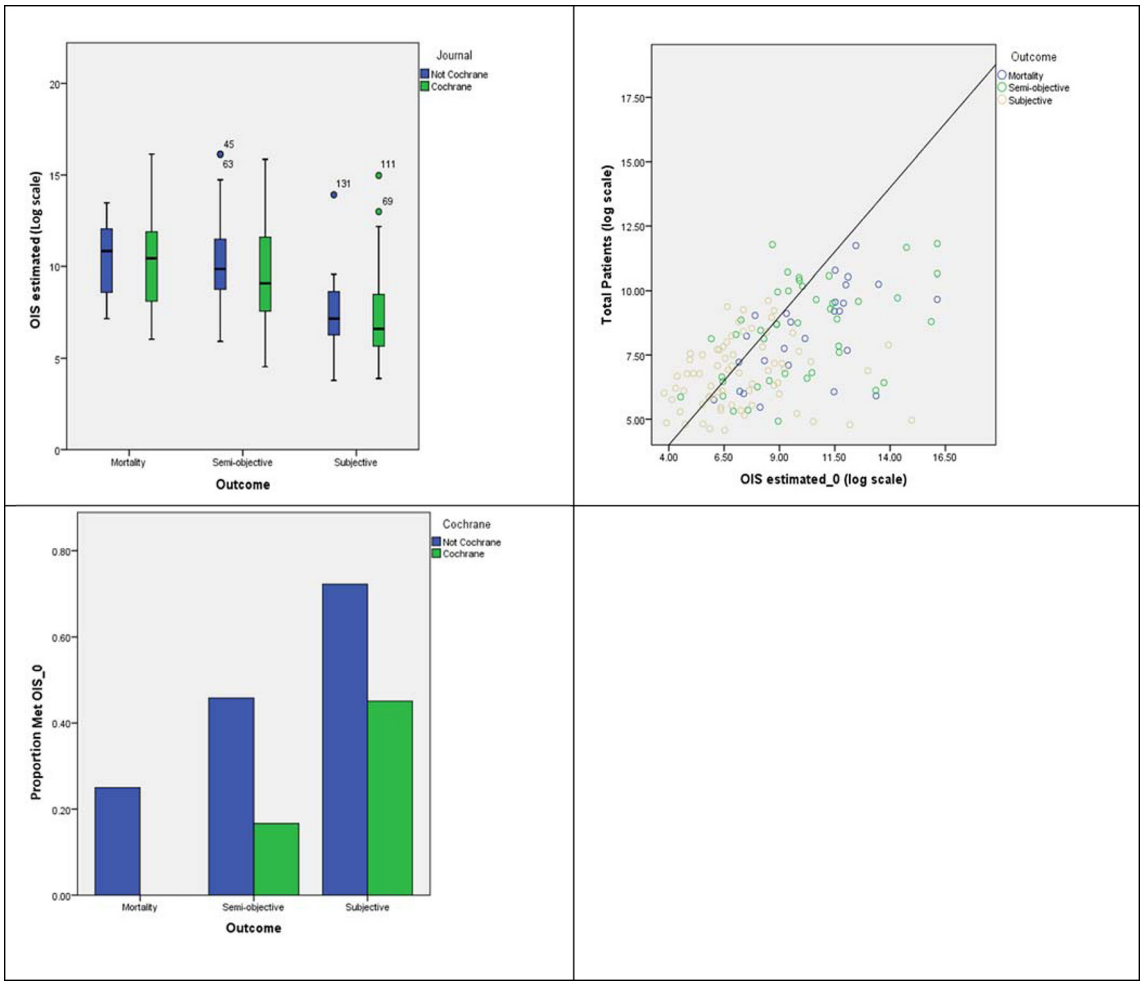


Figure 4 OIS estimated considering best-case scenario (heterogeneity=0) by (A) type of outcome and source; y-axis is on the log scale, (B) related to the total number of patients included in each SR and (C) proportion of SRs achieving the OIS by type of outcome and source. OIS, optimum information size; SRs, systematic reviews.

Table 3 Percentage of meta-analyses that achieve the OIS by heterogeneity (I^2) level assumed

	All cause mortality			Semiobjective			Subjective		
% (n/N)	Coch	Non-Coch	95% CI difference	Coch	Non-Coch	95% CI difference	Coch	Non-Coch	95% CI difference
OIS			(–0.074 to 0.571)			(–0.043 to 0.499)			(0.111 to 0.616)
Achieved	0	25		11.1	37.5		31.4	72.2	
I^2=reported	0/14	3/12		2/18	9/24		16/51	13/18	
OIS			(–0.074 to 0.571)			(–0.031 to 0.534)			(–0.024 to 0.490)
Achieved	0	25		16.6	45.8		45.1	72.2	
I^2=0	0/14	3/12		3/18	11/24		23/51	13/18	
OIS			(–0.134 to 0.491)			(–0.079 to 0.460)			(0.027 to 0.553)
Achieved	0	16.6		11.1	33.3		29.4	61.1	
I^2=Q3	0/14	2/12		2/18	8/24		15/51	11/18	

coch, Cochrane meta-analyses; non-Coch, non-Cochrane meta-analyses; OIS, optimum information size.

Box Example of five meta-analyses with the 'all cause mortality' as the main outcome that do, or do not, achieve the optimum information size (OIS)

Meta-analysis that meet the OIS

Weng et al (Annals of Internal Medicine)[25]
This meta-analysis evaluated the use of a non-pharmacological intervention (non-invasive ventilation) to treat patients with acute cardiogenic pulmonary oedema including a total of 1369 patients with a control event rate (CER) of 23%, relative risk reduction (RRR) 27% and 0% heterogeneity. For this systematic review assuming a 0% heterogeneity, the OIS estimated was 1296 patients.

Gastric team (The Journal of the American Medical Association)[26]
This meta-analysis evaluated the use of adjuvant chemotherapy for resectable gastric cancer including a total of 3781 patients with a CER 69%, RRR 9% and 24% heterogeneity reported by the meta-analysis. For this systematic review assuming a 0% heterogeneity, the OIS estimated was 1828 patients.

NSCLC meta-analysis (The Lancet)[27]
This meta-analysis evaluated the use of adjuvant chemotherapy in patients with operable non-small cell lung cancer including a total of 8447 patients with a CER 49%, RRR 11% and 1% heterogeneity reported by the meta-analysis. For this systematic review assuming a 0% heterogeneity, the OIS estimated was 2686 patients.

Meta-analysis with a large number of included patients that not meet the OIS

Adam et al (Annals of Internal Medicine)[28]
This meta-analysis evaluated the use of warfarin versus new oral anticoagulants for the management of atrial fibrillation and venous thromboembolism including a total of 14 143 patients with a CER of 2%, RRR 12% and 0% heterogeneity reported. For this systematic review assuming a 0% heterogeneity, the OIS estimated was 100 562 patients.

Rizos et al (The Journal of the American Medical Association)[29]
This meta-analysis evaluated the administration of omega-3 fatty acid supplementation and risk of major cardiovascular disease events including a total of 125 410 patients with a CER of 7%, RRR 4% and 1% heterogeneity reported. For this systematic review assuming a 0% heterogeneity, the OIS estimated was 255 912 patients.

is the higher number of smaller randomised controlled trials. Nevertheless, these differences were more marked in Cochrane reviews, while non-Cochrane reviews showed more similar levels of heterogeneity across all types of outcomes. The obtained results show that globally less than half of recent published meta-analysis in high-quality journals achieved the OIS and therefore do not have appropriate statistical power to draw firm conclusions.

As expected, the estimation of the OIS assuming different levels of heterogeneity, and alpha values, showed a strong correlation. Although we used specialist software for the estimation of the OIS (TSA V.0.9), it is possible to estimate this value using any software that allows sample size estimation if the heterogeneity level is assumed to be zero. Incorporation of heterogeneity can be done using a simple adjustment proposed by Wetterslev.[5] This author proposes the use of an alternative index named the diversity (D^2) statistic as opposed to the I^2 factor. However, there is currently no consensus on what measure of heterogeneity to adopt for the OIS.[4 14]

Published meta-analyses that estimate optimal information size often use one or more statistical assumptions, such as a RRR of 10%, or the median RRR of trials with low risk of bias.[15–17] Our analysis shows that the median of the RRR is 20% for all pooled reviews. However, because the distribution of RRR varies by outcome type, in some cases optimal information size is underestimated, while in others it is overestimated.

Limitations

There are several proposed statistics to define a 'desirable sample size in terms of numbers of participants across all studies'.[4] The OIS as described in this paper involves a relatively simple calculation, which if anything is likely to underestimate the information required to define whether firm evidence has been reached to draw robust conclusions[4 14] Therefore, we used this definition of OIS as a measure to estimate what proportion of SRs meet this minimum requirement.

We have focused exclusively on the calculation of a single threshold to define when/if a minimum level of evidence has been collected. However, retrospective analyses of meta-analytical results are more commonly used to inform prospective studies. For example, to determine the size of a new trial to answer definitively a question around efficacy. The use of trial sequential methods has been proposed to identify early signals of effect with monitoring boundaries being defined by frequentist, semi-Bayesian and fully Bayesian methods.[4 18 19] Although there is still considerable uncertainty about the estimates and the best method to use, empirical studies have provided examples to suggest these methods could help detect signals early (benefit, harm or futility).[8 20] Of note, the identification of the sample size required in a new study or studies will depend on the method used in the meta-analysis.[14]

Reviews conducted by the Cochrane collaboration are considered to be higher quality[21 22] and of greater methodological rigour than meta-analyses published in paper-based journals. Our study only included meta-analyses from the top five medical journals, and therefore our results may not be applicable to other meta-analyses published in other journals. Nevertheless, this would bias our results towards better evidence being evaluated to what is currently being generated. Also, our results do not generalise to network meta-analyses, which is an area of evidence synthesis that has grown rapidly.[23] A recently published study demonstrated that substantial variation exists in such network-based meta-analysis,[24] and the statistical methodology to estimate the OIS in these meta-analyses is less developed than for traditional meta-analysis, hence our exclusion of these studies.

Implications for researchers and methodologists

This study has shown that the type of outcome when estimating the OIS can be used as a proxy for defining the basic parameters (CER, RRR and I^2) required to perform the calculation. Systematic reviewers can use these results to calculate an OIS value for their primary outcome independently of the confidence they have on the specific parameters obtained from their review. Therefore, we encourage reviewers to use the estimation of a sample size as a measure of the likely confidence in their results. Particularly as >50% of the primary outcomes in recent SRs appear to fall below this minimum requirement, pointing out the need for further evidence to reduce uncertainty.

CONCLUSIONS

Heterogeneity and effect size impact on the estimation of the OIS. It is however possible to estimate the OIS using traditional sample size estimation software and if necessary adjust for heterogeneity. Our results demonstrate that the type of outcome is relevant to the estimation of the OIS, as well as the heterogeneity and the CER and RRR. Currently less than half of published meta-analysis in high-quality journals have achieved the OIS, and therefore conclusions based on such results are subject to substantial uncertainty.

Acknowledgements The research was supported by the NIHR Oxford Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. We wish to thank Pol Escarpenter for his support to adapt the figures to the requested format of the journal.

Contributors JMG-A and RP conceived and designed the study. JMG-A and NP extracted the data. JMG-A and RP analysed the data. JMG-A, RP, CB and CH interpreted the data. JMG-A and RP wrote the first draft. All authors contributed to the writing of the manuscript, revised the intellectual content and approved the version to be published.

Competing interests None declared.

Patient consent None.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement All of the data used in this research are provided within this publication, its appendices and the publications referenced in the online supplementary appendices.

REFERENCES

1. Pogue J, Yusuf S. Overcoming the limitations of current meta-analysis of randomised controlled trials. *Lancet* 1998;351:47–52.
2. Pogue JM, Yusuf S. Cumulating evidence from randomized trials: utilizing sequential monitoring boundaries for cumulative meta-analysis. *Control Clin Trials* 1997;18:580–93.
3. Brok J, Thorlund K, Gluud C, *et al*. Trial sequential analysis reveals insufficient information size and potentially false positive results in many meta-analyses. *J Clin Epidemiol* 2008;61:763–9.
4. Higgins JP, Whitehead A, Simmonds M. Sequential methods for random-effects meta-analysis. *Stat Med* 2011;30:903–21.
5. Wetterslev J, Thorlund K, Brok J, *et al*. Estimating required information size by quantifying diversity in random-effects model meta-analyses. *BMC Med Res Methodol* 2009;9:86.
6. Wetterslev J, Thorlund K, Brok J, *et al*. Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. *J Clin Epidemiol* 2008;61:64–75.
7. Turner RM, Bird SM, Higgins JP. The impact of study size on meta-analyses: examination of underpowered studies in Cochrane reviews. *PLoS One* 2013;8:e59202.
8. Imberger G, Thorlund K, Gluud C, *et al*. False-positive findings in Cochrane meta-analyses with and without application of trial sequential analysis: an empirical review. *BMJ Open* 2016;6:e011890.
9. Turner RM, Davey J, Clarke MJ, *et al*. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the cochrane database of systematic reviews. *Int J Epidemiol* 2012;41:818–27.
10. Thorlund K, Engstrøm J, Wetterslev J. User manual for trial sequential analysis (TSA). *Trial Unit Available Internet* 2011 (accessed 11 Jun 2014)
11. Rhodes KM, Turner RM, Higgins JP. Empirical evidence about inconsistency among studies in a pair-wise meta-analysis. *Res Synth Methods* 2016;7:346–70.
12. Djulbegovic B, Kumar A, Glasziou PP, *et al*. New treatments compared to established treatments in randomized trials. *Cochrane Database Syst Rev* 2012;10:MR000024.
13. Hayden JA, Tomlinson G, Ni A, *et al*. *Approaches to estimate and present baseline risks considerations for cochrane reviews evidence-informed practice integration of : 19th cochrane colloquium madrid*, 2011.
14. Kulinskaya E, Wood J. Trial sequential methods for meta-analysis. *Res Synth Methods* 2014;5:212–20.
15. Imberger G, Orr A, Thorlund K, *et al*. Does anaesthesia with nitrous oxide affect mortality or cardiovascular morbidity? A systematic review with meta-analysis and trial sequential analysis. *Br J Anaesth* 2014;112:410–26.
16. Bangalore S, Kumar S, Wetterslev J, *et al*. Angiotensin receptor blockers and risk of myocardial infarction: meta-analyses and trial sequential analyses of 147 020 patients from randomised trials. *BMJ* 2011;342:d2234.
17. Afshari A, Wetterslev J, Brok J, *et al*. Antithrombin III in critically ill patients: systematic review with meta-analysis and trial sequential analysis. *BMJ* 2007;335:1248–51.
18. Wetterslev J, Jakobsen JC, Gluud C. Trial sequential analysis in systematic reviews with meta-analysis. *BMC Med Res Methodol* 2017;17:39.
19. Spence GT, Steinsaltz D, Fanshawe TR. A Bayesian approach to sequential meta-analysis. *Stat Med* 2016;35:5356–75.
20. AlBalawi Z, McAlister FA, Thorlund K, *et al*. Random error in cardiovascular meta-analyses: how common are false positive and false negative results? *Int J Cardiol* 2013;168:1102–7.
21. Remschmidt C, Wichmann O, Harder T. Methodological quality of systematic reviews on influenza vaccination. *Vaccine* 2014;32:1678–84.
22. Windsor B, Popovich I, Jordan V, *et al*. Methodological quality of systematic reviews in subfertility: a comparison of cochrane and non-cochrane systematic reviews in assisted reproductive technologies. *Hum Reprod* 2012;27:3460–6.
23. Hutton B, Salanti G, Chaimani A, *et al*. The quality of reporting methods and results in network meta-analyses: an overview of reviews and suggestions for improvement. *PLoS One* 2014;9:e92508.
24. Chambers JD, Naci H, Wouters OJ, *et al*. An assessment of the methodological quality of published network meta-analyses: a systematic review. *PLoS One* 2015;10:e0121715.
25. Weng CL, Zhao YT, Liu QH, *et al*. Meta-analysis: noninvasive ventilation in acute cardiogenic pulmonary edema. *Ann Intern Med* 2010;152:590–600.
26. Paoletti X, Oba K, Burzykowski T, *et al*. Benefit of adjuvant chemotherapy for resectable gastric cancer: a meta-analysis. *JAMA* 2010;303:1729–37.
27. Arriagada R, Auperin A, Burdett S, *et al*. Adjuvant chemotherapy, with or without postoperative radiotherapy, in operable non-small-cell lung cancer: two meta-analyses of individual patient data. *Lancet* 2010;375:1267–77.
28. Adam SS, McDuffie JR, Ortel TL, *et al*. Comparative effectiveness of warfarin and new oral anticoagulants for the management of atrial fibrillation and venous thromboembolism: a systematic review. *Ann Intern Med* 2012;157:796–807.
29. Rizos EC, Ntzani EE, Bika E, *et al*. Association between omega-3 fatty acid supplementation and risk of major cardiovascular disease events: a systematic review and meta-analysis. *JAMA* 2012;308:1024–33.