Issues in the design, conduct and reporting of clinical trials that impact on the quality of decision making

Davina Ghersi
2006

PhD Thesis
School of Public Health
Faculty of Medicine
The University of Sydney

This thesis is submitted in total fulfilment of the requirements for the degree of Doctor of Philosophy, University of Sydney.

Abstract / executive summary

Background

Clinical trial investigators and research ethics committees are obliged to ensure the that clinical trials are methodologically sound, and that the results of those trials are made publicly available in an open and honest fashion that is not misleading and does not misrepresent the data. This will ensure that only the best research is conducted, and that clinical decision making is based on the best possible evidence.

Objective

The objective of this thesis was to explore issues in the design, conduct and reporting of clinical trials that impact on the quality of decision making by research ethics committees, as well as health care providers and consumers.

Methods

Three studies were conducted, the first of which was a follow-up study of 103 randomised trials submitted to a human research ethics committee in central Sydney. Information in the trial protocol was compared with that reported in the trial publication, the intention being to identify discrepancies between the two documents, particularly in relation to the primary outcomes. The second study reported is a systematic review of published studies evaluating the impact of shared scientific or ethical review of multi-centre clinical trial protocols on the quality of clinical research and the research process. The third is a prospective cohort study of trials submitted to a central committee for scientific review, and the impact that review had on the functioning and decision-making of human research ethics committees.

Results

Selective reporting of primary outcomes encompasses selection of which outcomes are reported (discrepancy in identity), how the outcome is defined (discrepancy in definition) and selection of the amount of information reported for an outcome (completeness of reporting). Selective reporting of the primary outcome existed in some form in a significant proportion of trial publications. 17% of outcomes declared as a primary in the protocol were not reported as the primary outcome in the

publication, and 15% of outcomes declared as a primary in the publication were not declared as primary outcomes in the protocol. Lack of adequate outcome definition in protocols and publications meant it was not possible to assess discrepancy in definition. 76% of trials completely reported all of their comparisons. Statistically significant comparisons were more likely to be completely reported. Trials with a completely or partially documented sample size calculation in the protocol were significantly less likely to selectively report the primary outcome, and this variable may be a proxy measure for the quality of the trial based on the protocol.

The quality of the trial based on information obtained from its protocol is a fundamental component in the decision-making process of an ethics committee to approve or reject an application. Evidence for this is reflected by the nature of the changes requested by ethics in committees as a condition of approval. There is insufficient evidence provided by the identifiable published studies to be able to determine whether centralised scientific or ethical review improves the quality of trials or the quality of decisions made by RECs.

The prospective study revealed that the 22 ethics committees affiliated with NSW Health spend a considerable amount of time each year on scientific review: an estimated 2,315 hours - equivalent to approximately 1.4 full-time equivalent positions. RECs found the information provided by the central committee useful, and believed that it reduced the overall time taken to consider the trial at their meeting and improved the level of confidence they had in their decision making.

Discussion

Reporting of clinical trials should be consistent with original trial design as outlined in the clinical trial protocol. In the follow-up study of randomised trials, incomplete reporting of the primary outcome was a particularly problem among trials with outcomes that were not conventionally significant statistically. This may compromise the ability to undertake unbiased systematic reviews of all relevant trials.

The research also identified that there can be changes in one or more aspects of a primary outcome between the protocol and the publication, such as a change in the definition of a positive test, or a change in the time frame for measurement. These

changes could potentially bias estimates of treatment effects, if based on knowledge of trial results. It was not possible in most cases to be able to judge whether these changes were appropriate due to insufficient documentation provided in either the trial protocol or the final publication.

This research has identified the importance (when considering the scientific quality of clinical trials) of having processes for ensuring that trial reporting is consistent with the original trial protocol and subsequent protocol amendments. Ethics committees may be in a good position to help with this process.

Currently ethics committees are responsible for ensuring that trials of appropriate scientific quality are undertaken. A variety of models currently exist for reviewing scientific quality either directly or indirectly for such committees. A systematic review of a centralised process of scientific or ethical review found very little evidence on whether this improved scientific quality. A pilot project in NSW has been exploring more efficient models for assessing scientific quality centrally for multicentre trials. An assessment of this scheme's value was limited in scope and further efforts to improve and assess strategies for ensuring high scientific quality of clinical trials in protocol design and reporting are warranted.

Conclusion

There would appear to be a direct relationship between the quality of clinical trials based on information available in trial protocols, and the quality of reporting of the results of the trial in a peer-reviewed publication. The quality of clinical trial protocols is an important part of the decision to allow a trial to proceed to recruiting participants. Further research is required to evaluate the effectiveness of measures taken to improve the method of evaluating the science of a trial as part of the ethical review process, including the value of centralising all or part of the process.

Table of contents

List of Tables

List of Figures

Acknowledgements

There are many to thank for their role, large or small, in the completion of the work performed as part of this thesis. Firstly, Lesley Townsend and the team in the Research Office at Central Sydney Area Health Service for access to their files and for their infinite patience. Thanks to all of the hardworking Executive Officers of Human Research Ethics Committees in NSW and all those in the Health Ethics Branch at NSW Health, particularly Carmel Edwards and Ainsley Martlew. Thanks to my supervisors: John Simes for adding to the challenge (!), Mike Clarke for having one foot on the ground and Chris Silagy for his inspiration; An-Wen Chan for your ground-breaking work (and your thesis); Megan Evans and Sue Carrick for your help in some aspects of data collection; Jordi Pardo for translating from Spanish, Adrienne Kirby for statistical advice and a friendly ear; Sally Lord and Lisa Askie for their feedback on Chapters; Jenny Chow for being so calm and organised and, importantly, everyone in the Systematic Review team at the NHMRC Clinical Trials Centre for being so fantastic (and tolerant of my absences)! Thank you Kay Dickersin for your consistent support and encouragement, Joanne Gubbins for Saturday mornings, my family for caring and assorted Riff Raff for the occasional escape opportunity.

Contribution of the author

The author was responsible for the design, conduct, and analysis and reporting of all components of all 3 projects conducted.

For the follow-up study this included:

- Initial concept and design
- Preparation and submission of the original proposal to the Central Sydney Area Health Service Human Research Ethics Committee
- Design of data collection forms
- Identification of randomised trials by examining the individual REC files kept in archive boxes in CSAHS Research Office
- Identification of publications for identified randomised trials
- Obtaining permission from investigators to use their trials
- Extraction of data from REC files and publications
- Data entry of extracted data
- Analysis
- Interpretation and writing up

For the systematic review this included:

- Initial concept and design
- The literature search
- Eligibility assessment
- Data extraction and data entry
- Analysis
- Interpretation and writing up
- Kay Dickersin repeated the data extraction and commented on the interpretation and discussion section of the review
- Megan Evans repeated the eligibility assessment, obtained articles and assisted in the management of the Reference Manager database created to manage the review

For the review of the NSW Health Shared Scientific Assessment Scheme:

- Initial concept and design with some feedback from the NSW Health Reference Group for the Shared Scientific Assessment Scheme
- Preparation and submission of the original proposal to the University of Sydney Human Research Ethics Committee
- Design of data collection forms
- Data entry of extracted data
- Analysis
- Interpretation and writing up
- As the author remained blinded to the identity of the trials and the RECs to which the trials had been submitted, the processing and blinding of data collection forms was performed by Ainsley Martlew (NSW Health).

Glossary

Term	Definition
Adverse event	An undesirable or unwanted experience (expected or unexpected) that results from an intervention, including toxicity, injury or hypersensitivity.
Allocation concealment	The most important requirement of a treatment allocation scheme is that it is not possible for the next treatment to be allocated to be identified before the participant enters the trial. In this thesis allocation concealment is classified as: **Adequate**: if the individual enrolling trial participants was kept unaware of the randomisation sequence in advance through the use of central randomisation, independent preparation of drugs in sequential unmarked containers, sealed opaque envelopes, post-enrolment randomisation such as a coin toss, or variations thereof. **Inadequate**: if the allocation sequence was predictable or known prior to patient enrolment, or was not described.
Bias	A distortion in the selection of patients, collection of data, determination of endpoints, and final analyses which might result in misleading conclusions.
Blinding (or masking)	Concealing the identity of the treatment to which the patient has been allocated to the patient, the health care practitioner, the outcome assessor, the statistician, the data monitoring committee, or any combination of these. In this thesis the blinding of trials is classified as: **Open label**: no treatment blinding is used. **Patient is blinded**: the person receiving treatment is not aware of the nature of the intervention. **Person administering treatment is blinded**: the person administering the treatment is not aware of the nature of the intervention. **Double blind:** where both the person receiving the treatment, and the person administering the treatment, are unaware of the nature of the intervention.
Clinical drug trial	In the SSAS evaluation, a clinical drug trial was any trial involving a drug requiring notification to the Therapeutic Goods Administration (TGA) through the Clinical Trials Notification (CTN) or Clinical Trials Exemption (CTX) schemes.
Clinical trial	A prospective study comparing the effect and value of interventions in human beings.

Confirmatory trial	A confirmatory trial is an adequately controlled trial in which the hypotheses are stated in advance and evaluated. As a rule, confirmatory trials are necessary to provide firm evidence of efficacy or safety. The rationale and design of confirmatory studies nearly always rests on earlier clinical work carried out in a series of exploratory studies. (International Conference on Harmonisation 1998)
Control group	Used for comparison with the investigational treatment. In cancer trials this can be the current standard treatment.
Date of actionable approval document	In the SSAS evaluation, the date of an actionable approval document is the date a trial could technically start recruiting patients at the site/s the REC covers. This is barring practical issues such as drug availability, etc.
Effectiveness	Does the intervention work in the people to whom it has been offered. (Jadad 1998) Effectiveness trials tend to be pragmatic.
Efficacy	Does the intervention work in the people who have received it. (Jadad 1998) Efficacy trials tend to be explanatory.
Experimental study	A study in which the investigator controlled one or more variables in order to monitor the effect on a process or outcome. Designed to learn more about the population under study rather than about the procedure or treatment. (Easterbrook & Matthews 1992)
Explanatory trial	Pose specific scientific hypotheses aimed at improving basic understanding of how treatments work, provide a scientific basis for modifying therapies and for introducing novel approaches to medical treatment. An explanatory trial recruits as homogeneous a population as possible to maximise the chances of demonstrating treatment effects. (Roland & Torgerson 1998;Simes 1998)
Exploratory trial	In contrast to confirmatory trials, the objectives of an exploratory trial may not always lead to simple tests of pre-defined hypotheses. Their analysis may entail data exploration; tests of hypotheses may be carried out, but the choice of hypothesis may be data dependent. Such trials cannot be the basis of the formal proof of efficacy, although they may contribute to the total body of relevant evidence. An individual trial may have both exploratory and confirmatory aspects. (International Conference on Harmonisation 1998)
REC Executive Officer	The individual responsible for managing the work of the REC. May also be known by other terms such as REC Secretary.

Interim analysis	An interim analysis of a clinical trial is conducted after a proportion of the anticipated total sample size has been randomised. The purpose is to monitor the progress of the trial and to assess whether there is a significant difference between the groups that may warrant early closure of the trial.
Intervention group	The intervention/s in a clinical trial which are being compared with the control (or standard) treatment.
Multi-centre	Refers to clinical trials involving more than one health care institution (such as hospitals).
Observational study	A study in which the investigator observed a process or disease without intending to alter it during observation. (Easterbrook & Matthews 1992)
Outcome	A variable intended to be assessed in all study participants for the purpose of comparing the effects of interventions between randomised groups. (Chan AW & Altman 2005)
Placebo	An inactive agent given to a participant as a substitute for an active agent. (Meinert 1986b) Double-dummy designs involve 2 or more arms with an active intervention where there is a placebo for each intervention.
Pragmatic trial	Primarily concerned with identifying the optimal treatment from the patient's perspective. The patient population in a pragmatic trial is likely to be more heterogeneous than an explanatory trial, reflecting variations between patients that occur in clinical practice. They aim to inform choice between treatments. Outcome measures in pragmatic trials are patient-oriented and represent the full range of health gains. (Roland & Torgerson 1998;Simes 1998)
Primary outcome	A primary outcome is clearly distinguishable if there is a clear statement that it is the primary (main) outcome. If there is no clear statement then it assumed that the primary outcome is that used to calculate the sample size, or stated in the aims or objectives (in that order).
Protocol	A written description of a clinical trial including the objectives, eligibility criteria, treatment regimens, statistical and administrative details. A blueprint for an experiment.

Randomisation	The process by which patients are randomly assigned to one of the interventions on a randomised clinical trial. The purpose of randomisation is to ensure that the types of patient in each treatment group are as similar as possible, and hence "to eliminate possible biases that may lead to systematic differences between the treatment groups". (Altman 1991)
Research Ethics Committee	Or Ethical Review Committee (ERC): entities or committees responsible for reviewing the ethical aspects of a clinical trial. Includes, but is not restricted to: institutional review boards (IRB - the term commonly used, for example, in the USA), human research ethics committees (REC - used in Australia), institutional ethics committees (IEC) and research ethics boards (REB: used in Canada).
Safety trial	The main aim of the study is to evaluate whether the intervention is safe to use in the target population, in what dose / schedule and at what cost (in terms of toxicity).
Sample size	The total number of participants to be recruited to a clinical trial.
Scientific review	Refers to a review of the science (but not the ethics) of a clinical trial. The "science" includes the methodological quality of the design of the clinical trial (eg randomisation, sample size calculation), and/or safety issues (including toxicology), and/or the quality of the clinical aspects of the trial (relevance of the question, appropriateness of outcomes, etc).
Sequence generation	The method used to generate the order in which participants are allocated to treatment. In this thesis sequence generation is classified as: **Adequate**: if a truly random method was described, including a random number table, computer-generated random sequence, coin toss, draw of numbers from a container, or variation thereof. **Inadequate**: if the method was described and is not truly random, or was not described.
Sponsor	The organisation (eg pharmaceutical company or collaborative group), institution or individual (eg Principal Investigator) responsible for the initiation, management and / or financing of a clinical trial.

Acronyms

AHEC	Australian Health Ethics Committee
CDAI	Crohns Disease Activity Index
CONSORT	Consolidated Standards for the Reporting of Randomised Trials
CRA	Clinical Research Associate
CSAHS	Central Sydney Area Health Service
CTN	Clinical Trial Notification
CTS	Clinical Trials Sub-committee (Scheme)
CTX	Clinical Trial Exemption (Scheme)
EF	Evaluation Form
FDA	Food and Drug Administration
FEV	Forced Expiratory Volume
FTE	Full Time Equivalent
FVC	Forced Vital Capacity
GCP	Good Clinical Practice
HCEC	Hospital Clinic Ethics Committee
HCP	Health Care Practitioner
REC	Human Research Ethics Committee
ICH	International Conference on Harmonisation
ICMJE	International Committee of Medical Journal Editors
IEC	Institutional Ethics Committee
IRB	Institutional Review Board
LREC	Local Research Ethics Committee
MREC	Multi-Centre Research Ethics Committee
NHMRC	National Health and Medical Research Council
NSW	New South Wales
QOL	Quality of Life
RCT	Randomised Clinical Trial
REC	Research Ethics Committee
SSAC	Shared Scientific Assessment Committee
SSAS	Shared Scientific Assessment Scheme
SPSS	Statistical Package for the Social Sciences
SSRI	Selective Serotonin Reuptake Inhibitors
TGA	Therapeutic Goods Administration
UK	United Kingdom
US	United States of America
WHO	World Health Organisation
WMA	World Medical Association

Chapter 1: Background to this Thesis

This chapter will document background information that is relevant to the thesis as a whole. The issues addressed in this chapter include:

- Misleading the reader
- Trial quality
- Publication bias
- Multiplicity
- The impact these issues have on the work of research ethics committees

Introduction

> *"Both authors and publishers have ethical obligations. In publication of the results of research, the investigators are obliged to preserve the accuracy of the results. Negative as well as positive results should be published or otherwise publicly available."*
>
> (World Medical Association 2000)

Clinical trial investigators and RECs are obliged to ensure that clinical trials approved to recruit patients are methodologically sound, and that the results of those trials are made publicly available in an openly honest fashion that is not misleading. This will ensure that only the best research is conducted, and that clinical decision making is based on the best possible evidence.

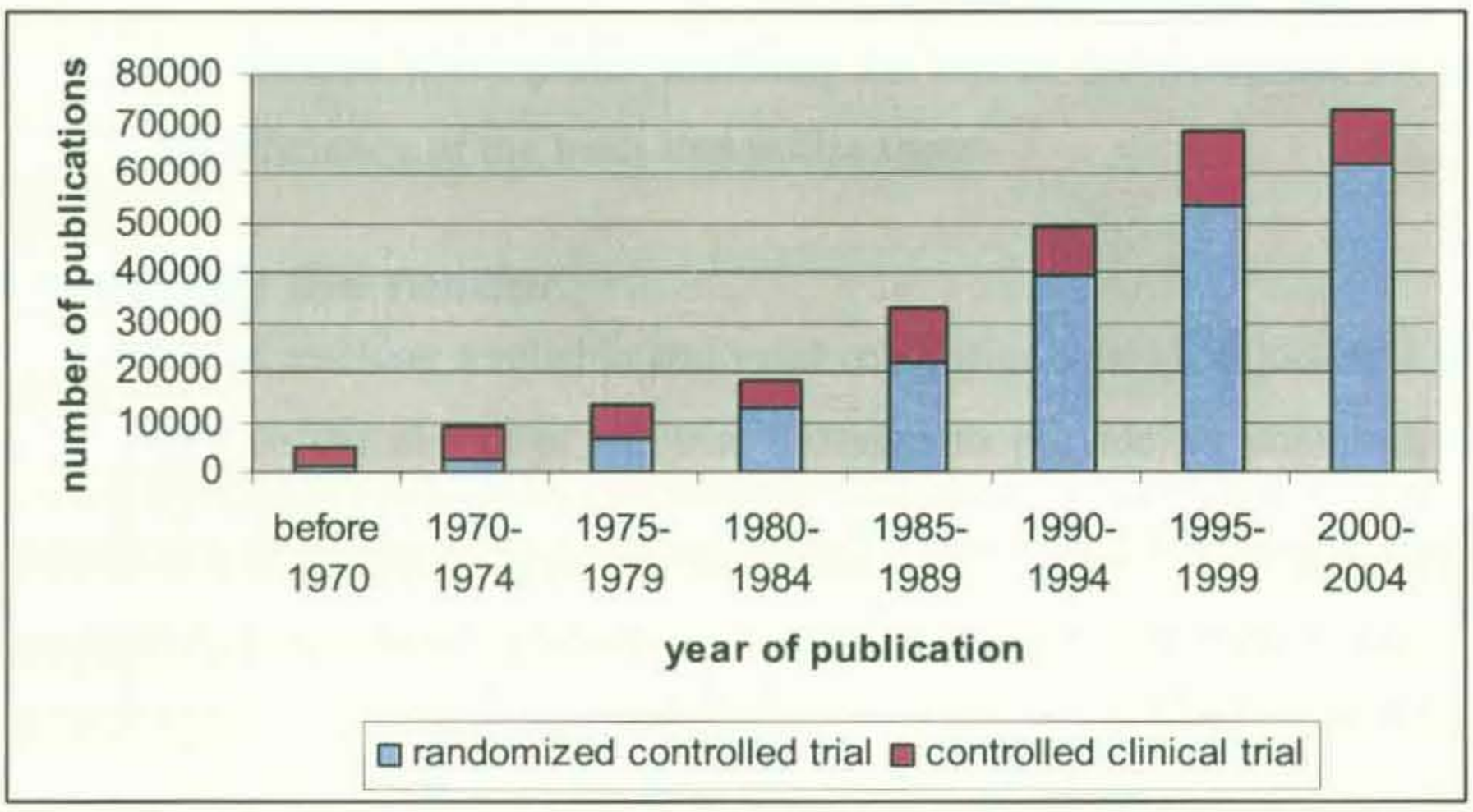
The objective of this thesis was to explore issues in the design, conduct and reporting of clinical trials that impact on the quality of decision making by research ethics committees, as well as health care providers and consumers.

This thesis will investigate ways in which the quality of randomised controlled trials might be improved at the design stage (before participants have been recruited) to ensure that decisions made based on the results of those trials are reliable. Of particular interest is the role ethics committees could potentially play in improving the quality of clinical trials through improving the quality of clinical trial protocols.

The volume of clinical trials research

The practice of medicine over the last two decades has emphasized the importance of basing treatment decisions on evidence that has demonstrated those treatments to be safe and effective. It is now generally accepted that the highest level of evidence for assessing the effects of interventions is provided by systematic reviews of all relevant randomised controlled trials or at least one properly designed randomised clinical trial. (Quality of Care & Health Outcomes Committee 1995) A clinical trial, as defined by the International Committee of Medical Journal Editors, is "any research project that prospectively assigns human subjects to intervention or comparison groups to study the cause-and-effect relationship between a medical intervention and a health outcome." (De Angelis *et al.* 2004) Randomised controlled trials (that is, when allocation to an intervention is by a chance process) are the closest thing to a true experiment available to those involved in the evaluation and provision of health care.

Figure 1: Publication type on Medline



The recognition of the importance of high level evidence in clinical decision-making has resulted in a dramatic increase in the volume of clinical trials research in the last 15 years. Based on references indexed on the Medline database of the National Library of Medicine, for example, there was a three-fold increase in the number of publications reporting the results of randomized clinical trials from almost 35,000 in the 1980s, to 93,000 in the 1990s, and over 60,000 between 2000 and 2005 (Figure 1). One USA-based source (CenterWatch) estimated that 80,000 trials were under way in

that country alone in 2001 (Lemonick & Goldstein 2002). As Australia did not have a prospective clinical trials register until June 2005 it is not yet possible to obtain accurate figures on the number of clinical trials that have been (or are being) conducted that involve Australians. During the first 5 months of the Australian Clinical Trials Register (July to November inclusive) almost 800 trials were submitted for registration, although registration during that time was voluntary.

As well as increasing in number, clinical trials are increasing in size and complexity. It is not uncommon for a trial to enrol thousands of patients across multiple centres in multiple countries. These large sample sizes are required to reliably detect the relatively small effects that can now be expected of most new treatments. (Peto, Collins, & Gray 1995) As clinical trials are both time and resource consuming, and considering the finite number of potentially eligible participants, it is important that they adopt design features that maximise their methodological efficiency. Features such as factorial designs, crossover designs, cluster designs, multiple treatment arms, and placebo run-in periods to identify potential non-compliers, eligibility criteria to identify and exclude participants who may be lost to follow-up are all intended to improve the efficiency of the trials that utilise them.

Misleading the reader

The chances of reaching a reliable and valid conclusion when conducting a clinical trial is dependent on the ability of the trial's design to provide an unbiased, meaningful answer to the question being addressed. The results of a trial which does not take measures to minimise bias, or is not statistically capable of answering the question, will mislead and confuse those attempting to interpret the results and incorporate them into clinical practice. The report of a clinical trial could also mislead if it is not an accurate reflection of the trial as it was originally designed. That is, there are unreported differences between the protocol and the publication that could influence the reader's interpretation of the results of the trial. Some examples of situations where this may be problematic are outlined in Table 1. Even an inconclusive ("null") result may mislead if the lack of a conclusive result is misinterpreted as the treatment being ineffective, when there may in fact be an effect which the trial was simply incapable of detecting. The fault in many cases will not lie with the treatment strategy under investigation, but will be due to inadequacies in the trial's design and interpretation.

(Fisher 1989) When calculating the sample size, for example, the investigators may have overestimated the expected benefit of the investigational treatment resulting in an underestimate of the number of patients required to answer the question. Or it may have been possible for the investigator to identify the next treatment to be allocated before randomisation took place due to the way in which randomisation was achieved, thereby enabling investigators to selectively exclude patients.

Table 1: Ways in which elements of a trial protocol may change between the initial protocol and the publication

Element	Example of how it might change	Why could a change be misleading
Eligibility criteria	Addition of criteria that may result in the post-hoc exclusion of patients previously considered to be eligible	If a per-protocol analysis is performed the eligibility criteria might be changed by investigators after the data has been looked at. It is not always possible to determine from a trial publication if analysis was per protocol or intention to treat and, if the latter, investigators do not always use this term correctly.
The intervention or comparator	Dose or schedule of a drug; timing of a procedure; therapist; etc	The intervention may be deliberately changed (eg dose increased) as a result of something that happened in the trial (eg adverse events). Interventions may end up being confounded but the confounding factor is not always evident to the investigator or the reader of the trial report.
Outcomes	Change in (or lack of) the definition	A problem if the definition changed after looking at the data. Lack of definition includes the non-specification of a time-point of interest (when applicable)
Outcome evaluation	Change in the way the outcome is measured	For example, the use of a more (or less) accurate test or instrument which increases (or decreases) the event rate
Outcome evaluation	Change in the frequency with which it is measured	Although follow-up may be planned at regular intervals using the same schedule for all trial arms, subjects on one arm may end up being followed more intensively
Design	Crossover design becomes parallel; The termination / non-reporting / merging of trial arms	Has implications for the power calculation and hence the ability of the trial to answer the question posed
Method of randomisation	The integrity of randomisation may be challenged (eg blinding is not being maintained) and hence the method changed.	Investigators should report problems encountered with achieving randomisation and measures taken to resolve these problems.
Sample size	Change in the outcome used, estimated size of effect, total number to be recruited, etc	These could be as the result of interim analyses or other looks at the data.

A poor quality clinical trial may mislead by:

1. Providing a biased estimate of the treatment effect. That is, there is a distortion in some aspect of the trial (such as the selection of patients, collection of data, determination of endpoints, the final analyses) which might result in misleading conclusions.
2. Providing an imprecise estimate of the treatment effect. That is the accuracy of the parameter estimates and their differences and is a result of inadequate power.
3. Underestimating the treatment effect due to partial treatment (for example, through poor compliance or high drop-out (or drop-in) rates).

The reader could also be misled through overemphasis of positive results, or underplaying of negative results, particularly when the reader does not have the skills necessary to interpret the information presented accurately, or to determine if the authors have interpreted it appropriately. (Pocock 2002) In addition, the authors of publications reporting the results of trials often fail to put their trial into an appropriate context, with many not including reference to relevant systematic reviews addressing the same or similar questions. (Clarke, Alderson, & Chalmers 2002)

It is important to note that a change made to any aspect of a clinical trial (to design, outcomes, treatment, etc) is not necessarily a problem as long as the reader of a resulting publication is provided with adequate information so they can evaluate the potential impact of the change. When a change is made to a protocol the trial (or publication) "user" needs to consider if the change was made as a consequence of knowledge of the results of the trial (including interim results). If the trial is still active then the user will also need to consider if the change is such that its existence in the original protocol would have impacted on the decisions made by ethics committees, funding agencies, etc at the time the trial was originally considered.

The impact of misleading results

Misleading results can have a devastating impact on the health outcomes of individuals treated as a result of this misleading research. In the 1950s, as an example, acetazolamide (alone or with furosemide) became standard practice in the management of post-haemorrhagic hydrocephalus in preterm infants, even though it was evident that there were side effects of cause for concern. In 1992 an adequately powered, randomised trial in preterm infants was conducted, the results of which indicated that the drug was associated with significantly poorer health outcomes including a higher rate of shunt placement and increased neurological morbidity. (Silverman 1999) During the 40 year period in which this drug was standard practice many babies will have suffered as the result of a change in practice not based on sound evidence.

> *"When the first few tries indicate that a new procedure 'works', all too often the exciting results are published and enthusiastic investigators become so convinced of the value of the intervention, it is baptised and named 'Standard Practice'. Moreover, the innovators are now unwilling to conduct a trial with concurrent controls."* (Silverman 1999)

A recent example is the use of selective serotonin reuptake inhibitors (SSRIs) in the treatment of depression, which are now known to be associated with an increased risk of suicide and suicidal ideation. At the heart of the heated and very public debate surrounding these drugs is the selective interpretation and reporting of data from clinical trials on suicidal behaviour. Cynics have suggested that it is only as the result of threatened legal action that one manufacturer, GlaxoSmithKline, made the unpublished results of the relevant trials available. (Tonkin & Jureidini 2005;Wessely & Kerwin 2004) The unpublished trials in children revealed an increased suicide risk and little evidence of efficacy. (Tonkin & Jureidini 2005)

Exhibits submitted in the case of Fentress v Lilly demonstrate that there was debate within another company manufacturing SSRIs (Lilly) as to whether some of the safety outcomes should be reclassified. Such reclassification would have lead to the under-interpretation of the significance of the results pertaining to those outcomes, including

a request that "suicide attempt" be change to "overdose", and "suicidal ideation" to "depression".(Bouchy 1990)

> *"the physician has reported a suicide attempt. Do we have a right to change it to some terminology which we may consider to be more specific, e.g. overdose, but which is not free from ambiguity and could be regarded as inaccurate or misleading. The term overdose is not free from ambiguity because there are clearly forms of overdose which are not related to suicide attempts, for instance wrong dose prescribed or dispensed, error on the part of the patient, etc."*

Further, the massaging of clinical trials to show a desired result can start at the very early stages of the study – for example, with the design of data collection forms. It has been suggested in the case of the SSRIs that the trials did not record suicidal behaviour *"owing to a lack of boxes corresponding to the side effect in question. This lack of recorded data has then been used against claimants as evidence that the supposed problem doesn't happen." (Healy 2002)*

The probability of reporting a false positive

A result will mislead if it turns out to be falsely positive: that is, a treatment is reported to be effective when in reality it is not. Tannock suggests that there are at least three factors that increase the probability that the results in a report of a clinical trial will be false positive: (Tannock 1996)

- The low probability that new treatments will lead to therapeutic advances, implying a low prevalence of true-positive trials
- Publication bias – the selective reporting of trials
- The performance of multiple significance tests, only some of which may be reported in an article

Low probability of new therapeutic advances

Most medical advances come in small steps rather than giant leaps. Large treatment effects are rare and the smaller, more realistic, treatment effects require very large sample sizes if they are to be detected with any degree of accuracy. (Meinert 1986a;Peto, Collins, & Gray 1993;Peto, Collins, & Gray 1995) A recognised problem

with many clinical trials is the tendency to underestimate the required sample size due largely to investigators overestimating the expected benefit of the investigational treatment. (Peto, Collins, & Gray 1995) Lack of precision can lead to type II error (i.e. false negative) if a null result is interpreted as the treatment having no effect, when such a result could in fact be due to the trial not having sufficient power rather than a genuine lack of treatment effect. The precision could be improved by increasing the number of events in each arm of a trial, thereby increasing the trial's ability to detect differences between the treatments.

Table 2: Type I and Type II error

Is there really a difference between treatments (The "truth")	Statistical significance	
	Significant	Not significant
Yes	True positive	False negative[2]
No	False positive[1]	True negative

Note 1: the significance level (α) = the probability of Type I error

Note 2: the power ($1-\beta$) = 1 – the probability of a type II error

A number of studies have demonstrated the inadequacy of trial sample sizes and their calculation. An early survey by Freiman *et al* examined 71 trials which did not reach a statistical significance level of 0.05 (referred to by the authors as "negative" trials) published in a selection of 20 journals between 1960 and 1977. (Freiman *et al.* 1978) The survey found that only 15% of trials had sample sizes large enough to detect a 25% improvement in response (32% were large enough to detect a 50% improvement) with 90% power. The authors concluded that sample sizes are often too small to "offer a reasonable chance of successfully rejecting the null hypothesis in favor of the treatment". (Freiman, Chalmers, Smith, Jr., & Kuebler 1978)

A second study by Meinert *et al* investigated a random sample of 180 papers published in 1980. Of 113 reports indexed as a clinical trial, only 2 showed any evidence of having calculated a sample size prior to the commencement of the trial. (Meinert 1986b) In a similar study, Pocock *et al* evaluated 45 clinical trial reports published in the latter half

of 1985 in the British Medical Journal, the Lancet and the New England Journal of Medicine. (Pocock, Hughes, & Lee 1987a) Only 5 reports mentioned the number of patients it was originally intended to recruit, and supported this with a statement of statistical power. In most cases it was impossible to tell whether there had actually been a sample size calculation before the trial started, whether trials had reached their target accrual, whether accrual had been deliberately exceeded to increase power, or whether the trial had stopped early due to a statistically significant difference demonstrated at interim analysis.

In 1995 Moher *et al* reviewed all 383 trials published in JAMA, the Lancet and the New England Journal of Medicine during 1975, 1980, 1985 and 1990. (Moher et al. 1995) A statistically non-significant result was reported by 102 trials, only 33 of which reported a sample size calculation, although the number of articles reporting such a calculation improved over time. This study also revealed various other problems with reporting, including lack of reporting of the statistical test on which the calculation was based, the event rate in the control group, and the alternative treatment hypothesis. Only 30% of trial reports provided sufficient detail to enable the sample size calculation to be duplicated.

Components of the sample size calculation

The components of the sample size calculation are important markers of the quality of a clinical trial in that they are a clear statement of the intent of the investigators conducting the trial. The key components are the outcome chosen and the expected effect of treatment on this outcome, and the level of error protection.

The outcome

The outcome used in the sample size calculation is critical as the rate of occurrence of the outcome event will affect the power of the study and the length of time it is required to run. (Meinert 1986b) The outcome used in the sample size calculation should be the primary outcome, and an accepted measure of the effectiveness of treatment both in the context of the trial as well as in existing routine clinical practice. Higher priority should be given to serious morbid events, such as death, than to softer non-clinical outcomes (or surrogate end points) such as a change in a laboratory value. The latter may not be directly relevant to clinical practice or accurately capture the effect on the true outcome.

(Grimes & Schulz 2005;Meinert 1986b) Some examples of inappropriate surrogate endpoints include breakage and slippage instead of pregnancy to evaluate condoms, bone mineral density instead of fracture to assess the safety of depo-medroxyprogesterone acetate, and ventricular arrhythmia instead of death to evaluate anti-arrhythmic drugs.(Grimes & Schulz 2005)

> *"Although many surrogate markers correlate with an outcome, few have been shown to capture the effect of a treatment (for example, oral contraceptives) on the outcome (venous thrombosis). As a result, thousands of useless and misleading reports on surrogate endpoints litter the medical literature. New drugs have been shown to benefit a surrogate marker, but, paradoxically, triple the risk of death. Thousands of patients have died needlessly because of reliance on invalid surrogate markers. ...Clinical research should focus on outcomes that matter."* (Meinert 1986b)

A single, clinically relevant outcome used to calculate sample size will make the trial easier to interpret and understand. The potential problem with a single outcome is the number of expected events which, if relatively small, will increase the sample size. One way to reduce sample size is to increase the number of events through the use of a composite outcome (that is, a combination of 2 or more outcomes into a single outcome) in its calculation. When a composite outcome is used to determine sample size it will increase the number of events and reduce the number of patients required, although it can be difficult to interpret in clinical practice. (Meinert 1986b) A composite outcome can also pose problems if investigators are tempted during the analysis of the trial to separate this grouped outcome into the individual outcomes, for which it is most unlikely there will be adequate statistical power to investigate owing to the small numbers in each group.

Using the chosen outcome, investigators need to determine the minimum treatment difference they wish to detect under the alternative hypothesis. (Meinert 1986b) The estimated size of the expected treatment effect is a critical component in the calculation of sample size, although it is usually not selected for statistical but for clinical, economic, ethical and pragmatic reasons. (Raju, Langenberg, & Sen 1993) This may explain why some investigators tend to overestimate the size of the expected effect. This tendency

was demonstrated by Raju *et al* in their study of 21 trials which were being considered for inclusion in a meta-analysis of artificial surfactant trials for neonatal respiratory distress syndrome. (Raju, Langenberg, & Sen 1993) The expected and the observed effect sizes were compared and every one of the 21 trials investigated had overestimated the potential benefit of treatment.

When designing a trial it is also important that investigators consider whether the expected size of the effect is the smallest considered clinically worthwhile and if, when the trial is completed, such a difference would be sufficient to change clinical practice. If the evidence supporting the benefit of treatment is sufficiently strong it may in such cases obviate the need to conduct a randomised trial. (Raju, Langenberg, & Sen 1993) As discussed previously, the expected benefit of any new treatment is most likely to be relatively small, although even small benefits have the potential to have a major public health impact, particularly if the intervention is of low financial cost and simple to administer, and the disease is common.

Error Protection

The significance level (α) is the probability of a type I error (i.e. false positive). The smaller the value of α, the larger the sample size required. If at the time of the analysis of the trial results the p value is less than or equal to α, then the null hypothesis of no difference between the treatments will be rejected. That is, the difference between the treatment groups is unlikely to be due to chance.

The choice of a p value is arbitrary, although it has become accepted to use cut-offs of 0.01 or 0.05. (Friedman, Furberg, & DeMets 1985b) There is, however, very little difference between a p value of 0.04 and 0.06 and any value from 0.01 to 0.10 could be conceivably justified. The founders of statistical inference certainly did not intend for the p value to be dichotomised into significant and non-significant. (Sterne & Davey Smith 2001) It is therefore "preferable to think of the significance test probability as an index of the strength of evidence against the null hypothesis". (Bland et al. 1985)

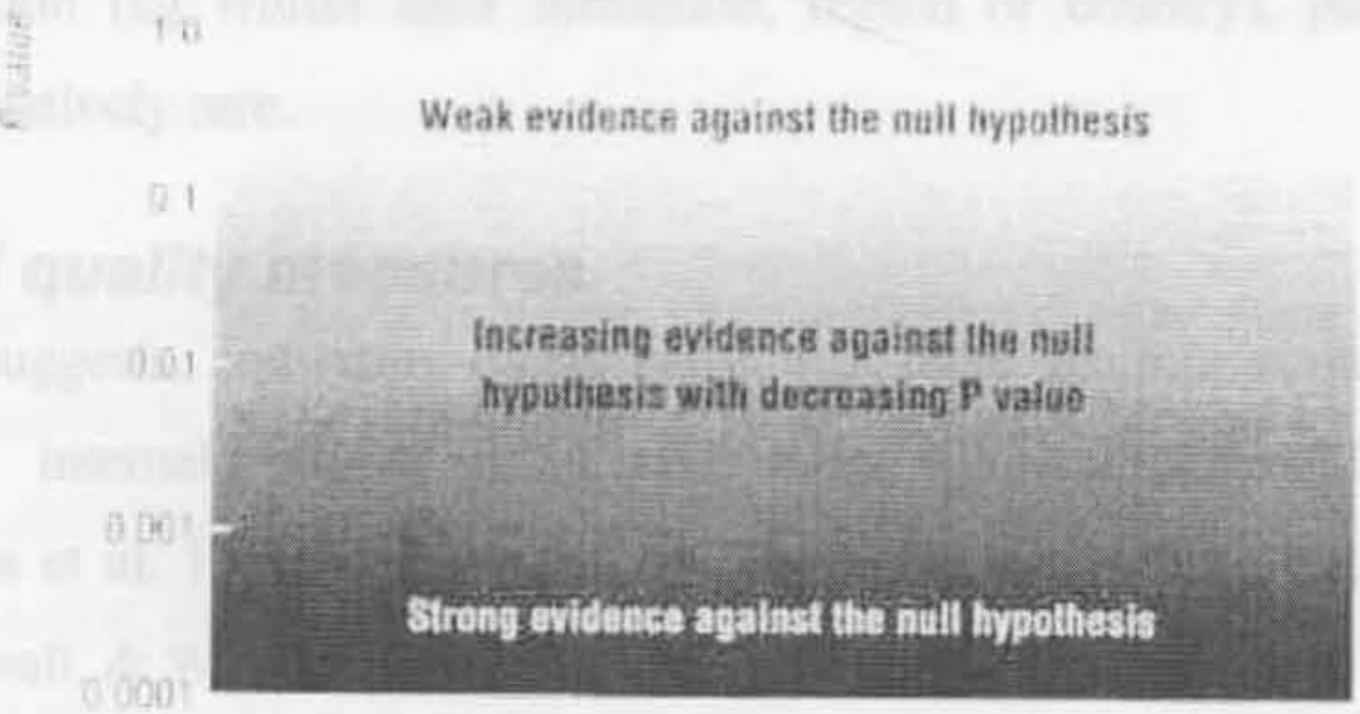


Figure 2: Suggested interpretation of p values from published medical research

From (Sterne & Davey Smith 2001)

The power of the study is the probability of correctly rejecting the null hypothesis. That is, 1 minus the probability of a Type II error (1 - β; i.e. false negative) occurring. The larger the expected treatment effect, the greater the power of the trial is to detect it. Traditionally the power of a study is set between 0.80 or 0.95, that is, an 80% to 95% chance of finding a statistically significant difference between the event rates, given that a difference actually exists. (Friedman, Furberg, & DeMets 1985a) Evidence suggests that randomised trials tend not to have adequate power to detect the differences in treatment anticipated by the investigators. In the previously cited study by Moher *et al*, for example, there were 70 trials with a simple two-group parallel design of which 52 had dichotomous and 18 had continuous primary outcomes. (Moher, Jadad, Nichol, Penman, Tugwell, & Walsh 1995) Only 16% of the former and 36% of the latter had a sample size large enough to detect, respectively, a 25% or a 50% treatment difference with at least 80% power.

The significance level and power the investigators elect to use to calculate sample size is a decision based largely on the need for the trial results to ultimately be accepted by the scientific community. Being arbitrary, there is a temptation to adjust the power (and the other components of the sample size calculation such as the estimated size of the treatment effect) in order to obtain the "desired" result. This is not a desirable practice and is considered by some to be unethical as it will result in the conduct of a trial that is incapable of addressing the research question. (Halpern, Karlawish, & Berlin 2002) Such manipulation may be somewhat inevitable when investigators are faced with a limited

target population (eg within their institution, region or country), particularly when diseases are relatively rare.

Other trial quality measures

Moher *et al* suggests, and many others agree, that there are four content areas which influence the internal validity of a randomised clinical trial: (Chalmers et al. 1990;Chalmers et al. 1981;Detsky et al. 1992;Jadad et al. 1996;Moher, Jadad, Nichol, Penman, Tugwell, & Walsh 1995)

- Patient assignment (randomisation)
- Masking (blinding)
- Patient follow-up
- Statistical analysis

The way in which patients are assigned to treatment (that is, randomised), and the degree of masking (or blinding) are two areas which should be discussed in the trial protocol. Both of these issues have been addressed in further detail below. It is unlikely that a protocol would document the measures to be taken to ensure complete patient follow-up as this would usually be considered to be a matter of good data management practice and not something that would vary across trials. We would expect to see information on follow-up reported in the publication. Whether an analysis plan should be documented in the protocol is debatable, and this issue will be addressed in further detail in Chapter 5.

Randomisation (patient assignment)

The purpose of randomisation is to ensure that the types of patient in each treatment group are as similar as possible. It can then justifiably be concluded at the end of the trial that any differences demonstrated between the treatment groups are not due to the selection of patients or other known (and unknown) underlying differences between the groups (although it should be emphasised that the play of chance alone could still lead to a false positive result in even the best conducted randomised trial). (Chalmers, Smith, Jr., Blackburn, Silverman, Schroeder, Reitman, & Ambroz 1981;Fisher 1989) The purpose of randomisation therefore is "to eliminate possible biases that may lead to systematic differences between the treatment groups" and is perhaps the most important measure of the quality of a clinical trial. (Altman 1991) To ensure the benefits of randomisation are

protected it is necessary to ascertain the method of randomisation to confirm that the trial really is randomised, and that the allocation schedule is concealed.

The most important requirement of a treatment allocation scheme is that it is not possible for the treatment to be allocated to be identified before entry into the trial. In a study conducted by Schulz *et al* it was demonstrated that inadequate concealment of the treatment allocation schedule is associated with larger estimates of treatment effect. (Schulz et al. 1994;Schulz et al. 1995) The study involved an assessment of the quality of 250 reports of randomised trials included in the Cochrane Pregnancy and Childbirth Database. Only 79 trials used a method of allocation which successfully concealed the identity of the next treatment to be allocated. The exaggerated estimates of treatment effect were also seen when it was unclear whether treatment identity had been adequately concealed. The odds ratios for inadequately concealed trials were on average, 30% lower than those for adequately concealed trials, resulting in larger estimates of treatment effect. In these trials the differences in treatment effect may have resulted from the bias caused by knowledge of the treatment allocation prior to randomisation influencing a practitioner's decision to include a patient in a trial.

The term "random allocation" may also be misinterpreted as meaning "haphazard" which it is not. (Altman 1991) Systematic methods of allocating treatment, unit number or date of birth for example, are not considered to be truly random as it is possible to identify the treatment before entry to the trial and patients may therefore be systematically excluded, thereby introducing selection bias. Pre-randomised envelopes are also open to abuse due to the opportunity to open the envelope and identify the treatment prior to randomisation. If the randomisation method involves the use of pre-randomised (and perhaps pre-stratified) blocks, but these are too small, it may be possible to accurately guess the next treatment to be allocated by recalling the result of previous randomisations.

In a historical review of randomised controlled trials published in the British Journal of General Practice, Silagy and Jewell found that, of 90 trials identified, allocation was stated to be by randomisation for 74%. Only 19 of these described the method of randomisation, thereby making it impossible to tell whether or not the trial was actually

randomised, and thus whether it was possible to determine the identity of the next treatment to be allocated before study entry. (Silagy & Jewell 1994)

Randomisation should occur at a time as close to the commencement of treatment as possible. If a trial involves the recruitment of the seriously ill, for example, a treatment delay could result in patients dying before treatment is received. This could be particularly problematic if the delay is differential, that is, there is a delay on one arm of the trial and not on other/s.

Masking (Blinding)

Knowledge of treatment assignment after the allocation has been made may also influence the assessment of treatment effectiveness. One important way of protecting against this treatment related bias is masking (also known as blinding). (Fisher 1989;Meinert 1986b) In a masked trial the identity of the treatment assignment is withheld to improve the objectivity of the treatment administration and assessment, data collection, reporting and analysis processes. (Meinert 1986b) The identity of the allocated treatment may be withheld from the patient, the health care practitioner, from those involved in the assessment and analysis of outcomes, or a combination of the four. It should be possible to unblind the treatment should it be necessary to do so during the course of the trial (due to an unforeseen adverse effect, for example).

In their study of the methodological quality of the 250 trials in the Cochrane Pregnancy and Childbirth Database (referred to previously in 2.1), Schulz *et al* demonstrated that the lack of double blinding was associated with treatment effects which were larger. In their case the odds ratios for non-blinded trials were, on average, 17% larger than blinded trials. (Meinert 1986b;Schulz, Chalmers, Grimes, & Altman 1994)

Double blind trials which are comparing an investigational treatment with no treatment will use a placebo - a "pharmacologically inactive agent given to a patient as a substitute for an active agent". (Meinert 1986b) The purpose of a placebo in a controlled clinical trial is to attempt to remove the bias that exists when a patient is aware of the treatment they are receiving and their expectations of its effects and side effects.

Publication bias

> *"All policymakers must be vigilant to the possibility of research data being manipulated by corporate bodies and of scientific colleagues being seduced by the material charms of industry. Trust is no defence against an aggressively deceptive corporate sector."*
>
> (Lancet Editorial 2000)

There is now good evidence to demonstrate the specific problem of publication bias. That is, when the decision to publish (or to delay publication) is influenced by the direction or strength of the results of the trial. (Dickersin 1990;Easterbrook et al. 1991;Simes 1986;Stern & Simes 1997) It is of particular importance in the conduct and interpretation of systematic reviews of clinical trials, where the exclusion of unpublished results will introduce bias if the unpublished results differ from those which have been published.

There is evidence that this selective reporting extends to progress reports submitted to organisations such as regulatory agencies. A cohort of 274 clinical drug trials submitted to the Finnish National Agency for Medicines in 1987 were followed to December 1993, by which time final reports had been received for 68, and 24 had been suspended. (Bardy 1998) The current status of each trial was obtained from the Sponsors of all but one trial. Each sponsor was asked to classify the trial as positive (defined as "investigational drug better than comparator"), inconclusive (defined as "investigational drug not clinically significantly different to comparator") or negative (defined as "objective of the study confirmed"). Based on the 188 trials with a classifiable outcome, the authors concluded that there was substantial evidence of selective reporting with a statistically significant association ($p=0.023$) between trial outcome and submission of the final report to the regulatory agency.

In addition to publication bias in the classic form, there may bias in the selection of data for inclusion in the publication. In 2003, Chan *et al* were able to provide evidence of the selective reporting of the outcomes of clinical trials. (Chan AW 2003;Chan AW & Altman 2003) The authors conducted a cohort study of 102 trials submitted to a

Danish ethics committee and demonstrated that 50% of efficacy and 65% of harm outcomes were incompletely reported. They also demonstrated that statistically significant outcomes were more likely to be completely reported.

Another issue with the reporting of outcomes are changes in the outcome that may occur at some point between the protocol and the publication. In 62% of trials in the study by Chan *et al*, there was at least 1 primary outcome that had been changed, introduced or omitted. (Chan AW 2003;Chan AW & Altman 2003)

As discussed in more detail in Chapter 2, composite outcomes combine 2 or more outcomes into a single outcome. Potential for bias arises when one or more of the components of the composite are altered in some way. This could involve the addition or removal of a component from the composite, or a change in the definition of one or more of the components. Freemantle *et al* reviewed the use of composite primary outcomes incorporating all-cause mortality in 167 clinical trials published in 9 major journals. (Freemantle et al. 2003;Meinert 1986b) The findings suggested that the reporting of composite outcomes is generally inadequate, with the specific problem of the authors implying that the results apply to the individual components of the composite outcome rather than only to the overall composite.

It has been suggested that the source of funding of a clinical trial has a direct impact on the reporting of trial results. The most comprehensive analysis of this issue was a systematic review conducted by Lexchin and Bero. . (Lexchin et al. 2003) The authors searched Medline from 1966 to December 2002 looking for studies that had analysed research sponsored by the pharmaceutical industry and had compared methodological quality or outcomes with studies with other sources of funding. Based on the 30 studies identified the authors concluded that research sponsored by pharmaceutical companies was more likely to have outcomes favouring the sponsor than were studies with other sponsors.

It could be argued that another aspect to publication bias is the interpretation made by the authors around a positive or negative finding by either the authors or the readers of a manuscript. Inadequate pre-specification of the primary hypotheses, or post-hoc emphasis on the most positive findings, can result in "unduly assertive claims of

treatment benefit" (Pocock 2002) It has been suggested that the effect size is exaggerated when based on small trials and that the authors of these manuscripts have a "predilection for overly optimistic conclusions", particularly when the nature of the interventions being investigated cannot be masked. (Silverman 1999)

Multiple tests for statistical significance

Considering the cost of conducting clinical trials it is perhaps not surprising that investigators are tempted to collect as much data as possible, and to perform multiple tests on that data. Even with the best of intentions, the quest to ensure that a potentially important difference between treatments is not missed, can lead to misleading significance values and incorrect conclusions.

The problem with performing multiple tests of comparisons in the same study is how to interpret the results of those tests given that, by chance alone, 5% will reject the null hypothesis at the "conventional" 5% significance level (referred to as Type I error). The probability of at least one false-positive finding can be written as $1-(1-\alpha)^n$ where α is the p value and n is the number of independent comparisons. (Smith et al. 1987) Based on this formula, if 10 independent tests are performed at a significance level of .05 then the probability of rejecting the null hypothesis in at least one test when the null hypothesis is true in all cases is $1-(1-0.05)^{10} = 0.4$. i.e. a 40% chance of declaring a difference when none exists (Fieller 2003)

There are numerous sources of multiplicity in clinical trials, including:

- the conduct of multiple comparisons
- the use of multiple outcomes
- multiple looks at the data
- repeated measures

Multiple comparisons

Multiple comparisons result from the conduct of several treatment comparisons all involving the same outcome measure and all at the same time point. There are many general settings where it occurs. One is the investigation of subgroups which should be clinically justifiable and clearly defined in the trial protocol. The results of

subgroup analyses are more likely to differ from the main treatment effect when the latter is small. Subgroups have the additional problem of poor statistical power. Most trials will aim to recruit sufficient participants to be able to address the primary research question, and statistical tests on subgroups will only have power to detect larger treatment effects than those anticipated for the primary comparison.

Assmann *et al* examined the use of baseline data in the publications of 50 consecutive clinical trials published in 4 major journals over a 3 month period in 1997. (Assmann et al. 2000) It was found that two-thirds of the publications presented subgroup findings. Less than half of these tested for statistical interaction, it was difficult to determine if the subgroups were declared *a priori* and most trials only had sufficient power to detect very large subgroup effects. Most of the trials reporting subgroup analyses claimed a treatment difference dependent on the subgroup, and most of these were included in the trial summary or conclusions.

Another setting for multiple comparisons is when there are more than 2 treatment arms, resulting in multiple pair-wise comparisons. A 3-armed trial (arms A, B and C) investigating a single outcome, for example, could compare AvB, BvC, AvC, AvB+C, BvA+C, CvA+B. The primary comparison of interest may not be clearly evident.

The CONSORT authors suggest that multiple comparisons should not be applied unless an overall single statistical test (if possible) is significant. (CONSORT Group 2005) It should also be clear in a trial report if the subgroup was declared *a priori* or was the result of post-hoc looks at the data.

Multiple outcomes / endpoints

Most trials are interested in the effect of the investigational treatment on more than one outcome. Although it is possible to calculate sample size to ensure sufficient statistical power to investigate multiple primary outcomes, it is not always clear which outcomes are primary or were used to calculate the sample size in a trial publication. The problem of multiple outcomes is further complicated by the likely interdependence of primary and secondary outcomes. (Meinert 1986a) The relationship between cause-

specific mortality and overall mortality in cancer or cardiovascular disease, for example. (Pocock 1997)

Multiple outcomes increase the risk of Type I error as they result in multiple analyses. For example, there is approximately a 20% chance of a trial with a genuine difference between treatments, and five not strongly correlated endpoints, of detecting at least one treatment difference with a p value of less than 0.05. (Pocock, Hughes, & Lee 1987a)

Multiple looks

During the course of a clinical trial it may be necessary to conduct of a number of interim analyses, usually for the purposes of safety or other data monitoring. Looking at the data multiple times during the course of a trial increases the probability of obtaining a significant result. Problems arise when those using the results of interim analyses interpret each look as if it is the only analysis that has been performed. Stopping rules that specify pre-determined conditions for the significance level at each time point that would result in termination of a trial should those rules be met are one way to deal with the problem of multiple looks. A major problem is the assumption that all of the contingencies that may arise during the course of a trial may be difficult to predict. (Meinert 1986a)

Repeated measures

An outcome may be measured at more than one time point during the course of a trial. It is not uncommon, for example, for routine tests (eg blood) and procedures (eg application of instruments such as quality of life scales) to be performed at each follow-up visit. The temptation for investigators is to report the results at each time point, and to plot these results on a diagram, which can trick the eye into seeing an effect that may not exist statistically. The results at each time point are likely to be highly correlated. Multivariate analysis techniques can account for the observations being obtained in a sequence and incorporate the correlation across time points.

Going fishing

The practice of conducting ad hoc analyses on data sets, usually without a pre-specified hypothesis, as a means of identifying comparisons (specifically subgroups) of interest is referred to as "exploratory analyses", "data dredging" or "fishing expeditions" - depending on the degree of cynicism of the individual in relation to the

data being explored. It is a technique commonly used in epidemiological research to identify aetiological disease factors, and in clinical trials to identify subgroups that may benefit or be harmed by the intervention. (Meinert 1986a)

The problem with data dredging is its post-hoc nature and issues around how a statistically significant difference should be interpreted. Meinert proposes ground rules for data dredging via subgroup analyses, including making a distinction between *a priori* and post hoc subgroups, avoiding conventional interpretation of significance tests and caution in making conclusions based on post hoc subgroups. (Meinert 1986a)

The danger is that investigators can "try out" various analyses before deciding which method they use as it offers "obvious opportunities for selecting a favourable analysis strategy". (The European Agency for the Evaluation of Medicinal Products 2002) Poor reporting of clinical trials can make it difficult to assess when such opportunities have been taken advantage of.

Methods to correct for multiple testing

"If you torture the data often enough it will eventually confess"

(Fieller 2003)

The best overall solution to multiple testing is, when possible and practicable, to focus on a single, primary comparison which is clearly declared *a priori* in the trial protocol. The best statistical solution is debatable and opinions differ as to whether or not adjustments should be made to analyses to account for multiple testing in a clinical trial. Proschan and Waclawiw suggest guidelines for multiplicity adjustment based on how related the questions are (interim analyses, for example, address the same question at multiple time points), the number of comparisons, the degree of controversy (have there been conflicting results in previous studies), who stands to benefit and the nature of the alternative hypothesis. (Proschan & Waclawiw 2000)

Bonferroni described criteria by which it should be possible to judge whether statistical significance is impaired by the analysis of multiple comparisons. (Smith, Clemens, Crede, Harvey, & Gracely 1987) When n multiple independent comparisons are performed, and a p value of 0.05 is considered to be statistically significant, then the difference should be significant at $p = 0.05/n$ to reduce the number of falsely significant findings. (Smith, Clemens, Crede, Harvey, & Gracely 1987)

Smith *et al* attempted to describe the level of attention given to the issue of multiple comparisons by evaluating all randomised trials published in 4 major journals over a 6 month period in 1982. (Smith, Clemens, Crede, Harvey, & Gracely 1987) Using a threshold of 0.01 (rather than 0.05) they determined whether the statistical significance of therapeutic comparisons were "impaired", basing their judgement on an adaptation of the Bonferroni criteria. Of the 67 trials identified 50 had a least one comparison whose claim of statistical significance was impaired, and in none of the trials publications reporting an impaired comparison was the potential impact of multiplicity discussed.

The traditional approach to deal with false positive findings is to control the familywise error rate (FEW); that is, the probability of at least 1 Type I error (Proschan & Waclawiw 2000) or, "the probability of rejecting at least one true null hypothesis in the given family of the hypotheses". (Hsueh, Chen, & Kodell 2003) The two main methods to statistically correct for multiple testing are:

1. Adjusting the nominal significance levels to allow for the multiplicity

The best known method to adjust significance levels is the Bonferroni correction, where the observed p value is multiplied by the number of tests performed. This is a simple but very conservative method of adjustment, and problems arise if a large number of tests are to be performed as the p value required will be extremely small. It does not take the fact that outcomes are usually correlated into consideration, assumes that all outcomes are equally important and reduces the power to detect real treatment effects. (Pocock 1997)

2. Multivariate statistical techniques

Multivariate techniques allow for correlated observations but can be complex and are not commonly used. (Pocock, Hughes, & Lee 1987a) Their advantage is that they deal with all measurements simultaneously and return a single p value. (Fieller 2003) It can, however, be difficult to interpret the nature of any difference that may be detected.

Impact of clinical trials on modern Research Ethics Committees

> *"Among the essential values for research is that of the integrity of researchers. This includes the commitment to research questions that are designed to contribute to knowledge, a commitment to the pursuit and protection of truth, a commitment to reliance on research methods appropriate to the discipline and honesty."* (NHMRC 2001)

The increased attention given to clinical trials has had a direct impact on the workload of modern Research Ethics Committees (RECs). A review of US institutional review boards (IRBs) by the US Department of Health and Human Services reported that the amount of work conducted by most IRBs had increased by 42% since 1974–75, when the average number of proposals reviewed by an IRB was 43 per year. (Lemonick & Goldstein 2002;US Dept Health and Human Services 1998) By 1998, some IRBs had as many as 2000 proposals per year to review.

Pich *et al* reviewed the trials submitted to the Hospital Clinic Ethics Committee (HCEC), the REC to which most clinical-trial protocols are submitted for approval in Spain, as the Committee had become concerned about the effect lack of resources and workload had had on its ability to meet its ethical obligations: specifically its inability to adequately follow-up approved trials. (Pich et al. 2003) In 2001 the HCEC surveyed the investigators of all 158 clinical trials approved by the committee in 1997. By 2001 only 29 of the 123 trials that had closed to recruitment had published results in peer-reviewed journals. The committee reported that they were worried by these findings, believing that "public dissemination of clinical-research results is an important ethical requirement." (Pich, Carne, Arnaiz, Gomez, Trilla, & Rodes 2003)

As clearly stated in the Declaration of Helsinki, all clinical research should have a protocol which is submitted for evaluation by a Research Ethics Committee. (World Medical Association 2000) Although specific expectations vary across countries, RECs are expected to consider the relevance of each clinical trial (to clinical practice) and the appropriateness of the design. (European Commission. 2001) (See Table 3

and Appendix 1: Expectations of a Human Research Ethics Committee regarding the science of a clinical trial)

Table 3: Australian RECs and clinical trials

An REC must consider all aspects of the design of a clinical trial and be satisfied that:
(a) the trial is directed to answering a specific question or questions;
(b) there is a scientifically valid hypothesis being tested which offers a realistic possibility that the interventions being studied will be at least as effective as standard treatment;
(c) where the research is therapeutic, and is therefore intended and likely to be of direct benefit to participants, there is an acceptable balance between the risks and benefits of the trial;
(d) the methodology provides:
(i) a rationale for the selection of appropriate participants;
(ii) an appropriate method of recruitment;
(iii) adequate, understandable information for the purpose of obtaining participant consent;
(iv) a clear description of the intervention and observation to be conducted; and
(v) a sample size adequate to demonstrate clinically and statistically significant effects

National Statement on Ethical Conduct in Research Involving Humans; Section 12.2 (NHMRC 2001)

In addition to being responsible for the ethical aspects and considering the relevance and design of each trial, are also expected to assume responsibility for monitoring the trial (primarily for adverse events) and to review amendments to research protocols. They are also expected to assume other responsibilities, the nature of which varies across countries and regions. In Australia they include regulatory, legal and insurance responsibilities. The burden on RECs is further complicated by external pressure to make decisions quickly. (European Commission. 2001)

RECs are established in accordance with the laws and regulations applicable in each country which, in turn, comply with international requirements. They are composed primarily of volunteers who, as a group, should be capable of ensuring the competent, unbiased review of research projects submitted to them. (NHMRC 2001;NHMRC 2002) Membership is expected to include lay people but such individuals are required to have sufficient scientific knowledge to meet their ethical obligations. They are generally a group made up of "conscientious, sincere, and disinterested" amateurs who are "reviewing too much, too quickly, with too little expertise". (Pierce 1997)

In many cases, RECs themselves recognise that they do not have the necessary expertise to fully assess scientific issues. The members of RECs from 6 hospitals in

the Netherlands, for example, were asked to rate their perceived level of competence in the evaluation of scientific issues relating to phase 2 clinical trial protocols. (Van Luijn et al. 2002) Although most of the respondents were medical or para-medical professionals, they reported that they found it very difficult to evaluate the feasibility of these trials (34%), the scientific methodology (30%) and how the data were to be analysed (42%).

RECs must also deal with trial investigators who are themselves often not properly trained in research methods. Poor training has been suggested as one reason for investigators not adequately evaluating the currently available evidence before initiating new trials, and for new trials not being designed to address some of the issues raised in the previous research, including the choice of appropriate outcomes. (Halpern, Karlawish, & Berlin 2002)

Relieving the burden

In the case of multi-centre clinical trials, it is common practice for the research protocol to be approved by the REC at every institution participating in the trial. Thus, if participants in a trial are to be recruited in 20 hospitals, there are usually 20 associated RECs, often with 20 different application forms, all with different requirements. If the trial involves more than one country, cross-cultural issues can add further complications. As a result, there is a perception among clinical researchers that there is unnecessary duplication of effort across multiple ethics committees for the same multi-centre clinical trial. (Burman et al. 2001;Wolf, Croughan, & Lo 2002)

Several countries have implemented centralised systems that aim to improve the assessment process, mainly in terms of the time taken for trials to receive ethical approval and thus for recruitment to begin. (Christian et al. 2002) There is potential for these centralised systems to relieve some of the burden carried by RECs by reducing duplication, particularly in those areas where they consider themselves to be less capable (such as scientific methodology).

These centralised systems vary in detail, and might involve centralisation of the entire process of ethical review (eg, in the UK), or parts of the process of ethical review - especially scientific review. (Ghersi & Dickersin 2004) Critics of centralised systems

of review suggest that these systems will increase the burden on researchers and ethics committees by the addition of another level of bureaucracy, and also maintain that their effectiveness has yet to be ascertained. (Alberti 2000)

Project Objectives

The overall aim of the projects included in this thesis is to explore issues in the design, conduct and reporting of clinical trials that impact on the quality of decision making by health care providers, health care consumers and RECs.

This thesis has 3 major components:

1. A follow-up study of submissions to an institutional REC
    - Aim: To identify and quantify discrepancies between the trial protocol and the trial publication, particularly regarding the reporting of trial primary outcomes.
    - Aim: To determine if these discrepancies are influenced by the statistical significance (and direction) of individual trial results.
2. A systematic review of published research:
    - Aim: To evaluate the impact of central (or "shared") scientific and/or ethical review of multi-centre clinical trial protocols on the quality of clinical research and the clinical research process.
3. A prospective cohort study of trials submitted to a central scientific committee
    - Aim: to assess the influence of a central (shared) scientific committee on the functioning and decision-making of Human Research Ethics Committees, and on multi-centre clinical trials.

Chapter 2: Background and Methods for follow-up study

This chapter is the first of 4 chapters documenting the background, methods, results and discussion of a follow-up study of trials submitted over a 5 year period to a research ethics committee. Chapter 2 documents the background and methods specifically for the follow-up study including:

- Background to the follow-up study
- Background information on the rationale for clinical trial protocols
- Specification of the study's objectives, design, eligibility criteria, data collection methods and endpoints
- Issues in the identification and definition of primary outcomes
- Other issues pertinent to the conduct of the follow-up study

Background to follow-up study

As discussed in Chapter 1, a number of studies have now been conducted and published that demonstrate the existence and impact of publication bias (Dickersin 1990;Dickersin & Min 1993;Easterbrook, Berlin, Gopalan, & Matthews 1991;Simes 1986;Stern & Simes 1997). Research into the selective reporting of clinical trials is, however, relatively sparse. In addition, while the adequacy of reported sample size calculations has been investigated (Freiman, Chalmers, Smith, Jr., & Kuebler 1978;Moher, Dulberg, & Wells 1994;Pocock, Hughes, & Lee 1987b), studies examining the sample size calculations as proposed in the trial protocol have not.

The impact of selective reporting of analyses was investigated by Melander *et al* who examined the application documentation for 42 placebo controlled studies of 5 drugs submitted to a regulatory agency and compared each application with the related publications. (Melander et al. 2003) The authors found evidence of selective reporting of analyses, with many publications preferring the "more favourable" per-protocol analyses (that is, including only those patients treated according to the protocol and excluding non-compliers) to the results of intention to treat analyses (that is, the inclusion of all patients randomised in the arm they were randomised to).

A pilot study conducted in the UK by Hahn *et al* attempted to examine within-study selective reporting by comparing the original study protocol with the subsequent study report for applications approved by a single REC in the UK. (Hahn, Williamson, & Hutton 2002) The 15 studies included in their project were not restricted to randomised trials and the authors were unable to achieve their aims due to lack of information in the protocols available to them.

The authors of a cohort study of 102 published randomised trials submitted to ethics committees in Denmark were more successful. (Chan et al. 2004a) The aim of this study was to examine the extent and nature of outcome reporting bias by comparing information in protocols and protocol amendments with that in the relevant trial publication. A "fully reported" outcome was defined in this study as one with sufficient data to enable the results to be included in a meta-analysis. Of the 99 trials measuring efficacy 91 (92%) had at least 1 incompletely reported efficacy outcome, and of the 72 trials measuring harms 58 (81%) had at least 1 incompletely reported harm outcome. They were able to demonstrate that completeness of reporting of an outcome was related to the statistical significance of the results, with outcomes with positive results ($p<0.05$) having greater odds of being fully reported. The study also reported a large number of discrepancies in the reporting of primary outcomes. In this case, a discrepancy was a difference in the identification of a primary outcome between the protocol and the publication. For example, a primary outcome in the protocol reported as a secondary or unspecified outcome in the publication.

In a sub-study of the Danish cohort described above, Pildal *et al* compared the descriptions of allocation concealment in the protocol and the resulting publications. (Pildal J et al. 2005) Using strict criteria (based on the criteria required for Cochrane systematic reviews, method of allocation concealment was unclear in 96 of the 102 trials based on the publication, and in 80 trials based on the protocol. It was concluded that, when using the protocol or the publication, most randomised trials have unclear allocation concealment.

Using the same definition of reporting completeness as the Danish cohort, Chan *et al* examined the protocols and resulting publications for 105 randomised trials funded by the Canadian Institutes of Health Research (a government funding agency) between

1990 and 1998 (Chan et al. 2004b). Again there was evidence of incomplete reporting, with efficacy outcomes with positive results ($p<0.05$) having greater odds of being fully reported. The discrepancies in the identification of the primary outcome noted in the Danish study were also noted in this cohort, with 19% of trials having the same discrepancies relating to the primary outcome.

In a third study, Chan and Altman attempted to determine the prevalence of incomplete outcome reporting by surveying the authors of all trials published in December 2000 and indexed on PubMed by August 2002. (Chan AW & Altman 2005) They found that 75% of the 505 trials reporting efficacy outcomes did not fully report all of their outcomes. 232 trials defined primary outcomes in the publication and 36% incompletely reported at least one. Again, statistically significant outcomes had greater odds of being fully reported. Reasons given by survey respondents for not reporting outcomes included journal space constraints, a result that was not clinically important or statistically significant, or not yet submitted or analysed.

Using the ability to utilise the outcome data as reported in the publication in a meta-analysis as a definition of complete reporting is useful for those who conduct those meta-analyses, however, a trial may still report sufficient information on the outcome to enable inferences to be made by other users of trial publications. It is therefore useful to (at least conceptually) distinguish between selective reporting that may be misleading, and incomplete reporting.

What is a clinical trial protocol

> ***Protocol*** *n - Specifications, rules and procedures for performing some activity or function.*
>
> (Meinert 1986b)

As the follow-up study will compared clinical trial protocols with their resulting publications, this section will describe what a protocol consists of, and what it is for.

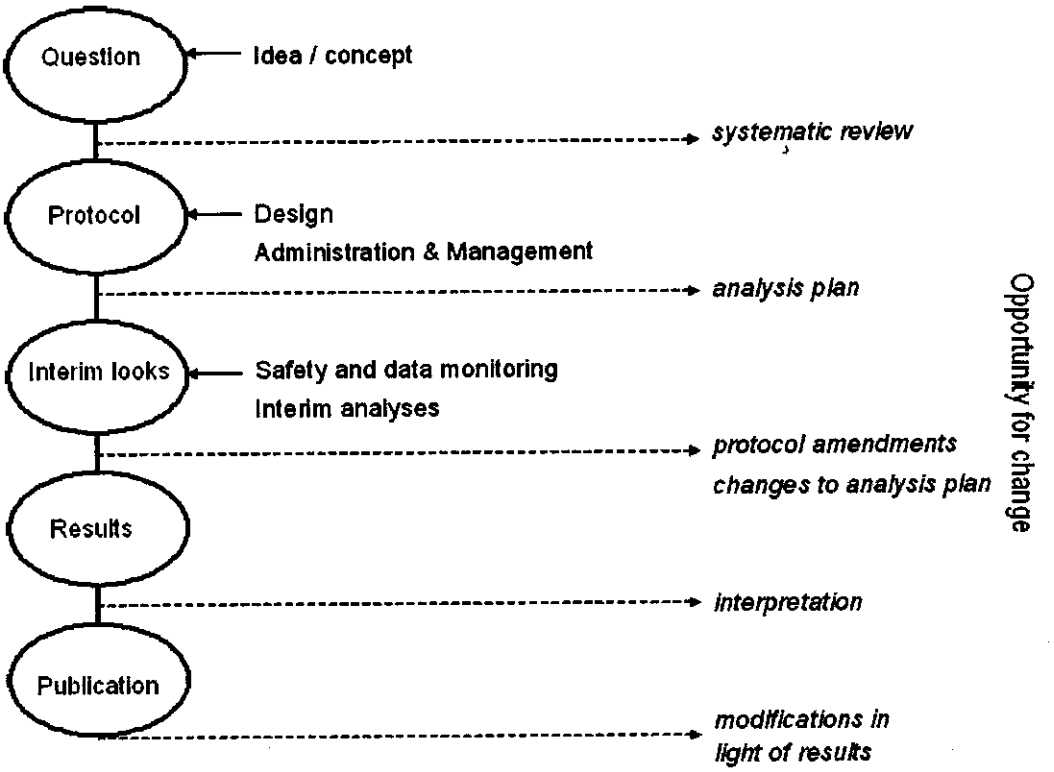
A clinical trial protocol is a written description of the trial including the objectives, eligibility criteria, and treatment regimens, statistical and administrative details. It is the blueprint for the planned experiment and ensures consistency of management of each participant in the trial.

There are multiple purposes for a clinical trial protocol and the contents required will vary depending on that purpose and the point the trial has reached in its progress (Figure 3). They are often required, for example, as part of an application to funding agencies, regulatory agencies as well as ethics committees along with accompanying documentation such as product brochures, indemnity and insurance information, patient information sheets and consent forms, etc. Most recently, there have been suggestions that details from the protocol and associated documentation (such as product brochures, consent forms, participant information sheets, contracts and financial arrangements, etc) should be submitted to and made publicly available on prospective clinical trials registries (The Ottawa Group 2006).

The Guideline on Good Clinical Practice (GCP) produced by the International Conference on Harmonisation (ICH) lists its requirements of a trial protocol, stating that is should include information on trial design, the selection and withdrawal of subjects, assessment of efficacy, direct access to source data documents and data handling and record keeping. (International Conference on Harmonisation 1996) The CONSORT statement describes various aspects of the protocol that need to be included in the trial publication when reporting results, including the planned study population, the planned interventions and their timing, primary and secondary

outcome measures, the minimum important difference and how sample size was projected. (CONSORT Group 1996)

Figure 3: Framework for undertaking a clinical trial



A clinical trial protocol submitted to an Australian REC must address all aspects of the design of a clinical trial (including the specific question the trial is designed to address). (NHMRC 2001)

> *"An REC must ensure that it is sufficiently informed on all aspects of a research protocol, including its scientific and statistical validity, that are relevant to deciding whether the protocol is both acceptable on ethical grounds and conforms with this Statement."* (NHMRC 2001)

Trial investigators must also demonstrate to each REC that the conduct of the trial will conform to the National Statement (and hence the Declaration of Helsinki), and comply with the relevant regulatory and legal requirements. (NHMRC 2001) Please refer to Appendix 1: Expectations of a Human Research Ethics Committee regarding

the science of a clinical trial, the guidance provided to ethics committees in Australia regarding the review of scientific and statistical validity.

Can a protocol contain too much information?

There may be an argument in favour of deliberately excluding information from a trial protocol. A trial using a permuted block randomisation scheme, for example, should not reveal the block size in the trial. Similarly, a trial using a minimisation scheme to randomise should not specify which stratification factors will be used to perform the minimisation in the protocol. In both cases documenting such information could potentially allow unblinding of the randomisation process, resulting in an increased ability to predict the next treatment to be allocated.

Discrepancies between a protocol and a publication

A clinical trial protocol is not a static document and may be amended numerous times during the course of the trial. It is therefore inevitable that there will be changes made to the trial leading to differences between the original protocol and the resulting publication. While the National Statement requires researchers to conduct a trial in accordance with the protocol, and to seek approval for amendments, the exact circumstances under which approval should be sought are somewhat vague and are simply stated as including those that "significantly affect the conduct of the trial" (NHMRC 2001). (See Appendix 2: Australian RECs and monitoring responsibilities) The Human Research Ethics Handbook indicates that the need to seek approval for protocol amendments is at the discretion of each REC, and encourages them to "establish procedures to assess these changes and determine whether their approval requires a full meeting of the REC or, in cases where they can be regarded as of minimal risk, by some other arrangement" (NHMRC 2002). When reviewing protocol amendments RECs are primarily concerned with changes that have a direct impact on trial participants (NHMRC 2002).

From a methodological perspective changes made to a trial protocol may be relatively minor (such as a change in an administrative process) or may have a significant impact on the trial (a change in the eligibility criteria, a reduction in the dose of a drug, an increase in the target sample size, etc) (Figure 3). It is also possible that changes will be made to a trial that are methodologically relevant but not of traditional interest to an

ethics committee and hence may not require a protocol amendment. Should, for example, a trial with a higher event rate than expected notify an REC of the increased power the trial will have if the target number of patients are recruited?

There are other agencies that use trial protocols that may have access to additional information not required by ethics committees. These include funding agencies, clinical trial registers and regulatory agencies such as the Australian Therepeutic Goods Administration (TGA) and the US Food and Drug Administration (FDA). A specific document of interest, for example, is the analysis plan which may not be available until after the trial has closed and data collection is complete. The key issue is the need to confirm that decisions regarding the analysis were made without knowledge of the results of the trial. A change made to the definition of an outcome, for example, may be appropriate if the change was made before analysis by allocated treatment of that outcome had been performed, and the same criteria for that outcome were applied to every participant in the study. It is reasonable to expect that these types of changes would be documented in the trial publication.

Objectives

Primary objective

To investigate the selective reporting of primary outcomes.

a. *Do randomised clinical trials start to address multiplicity early in the history of the clinical trial through the clear declaration of a primary outcome in the protocol?*
b. *Are there discrepancies between the trial protocol and the trial publication in relation to the identity of the primary outcome (or outcomes)?*
c. *Are there discrepancies between the trial protocol and the trial publication in relation to the definition of the primary outcome (or outcomes)?*
d. *Are primary outcomes "fully" reported*
e. *What factors in the protocol influence the selective reporting of clinical trials*
f. *What other (non-protocol) factors influence the selective reporting of clinical trials*

Secondary objectives

1: To explore issues relating to the sample size, its calculation and reporting

a. *Do randomised clinical trials provide adequate details in the protocol of the target sample size?*
b. *What is the completeness of documentation of the sample size calculation in the protocol?*
c. *What factors in the protocol are associated with the completeness of documentation of the sample size calculation in the protocol?*
d. *Is the completeness of documentation of the sample size calculation in the protocol related to the adequacy of reporting of the power calculation in the publication?*
e. *What other trial-related factors are associated with the adequacy of reporting of the power calculation in the publication?*

2: To explore other relationships between the protocol and the publication

a. *Allocation concealment and sequence generation*
b. *The use of blinding and placebo*
c. *Adverse events*

d. Journal type

e. Exclusions

3: To explore the impact of the availability of commercial funding on the trial protocol and publication

Design

This is a follow-up study of all randomised controlled trials considered by the Central Sydney Area Health Service (CSAHS) Ethics Review Committee between 1st January 1992 and 31st December 1996 that were subsequently published. Including trials submitted to the REC in between these dates allows for a minimum of 9 years (a maximum of 13 years) follow-up (to 2005) on eligible trials.

Note: the term "protocol" is used throughout this follow-up study as a collective term for the protocol as well as any other documentation submitted to the REC, including protocol amendments.

Eligibility

Studies were included in the follow-up study if they met the following inclusion and exclusion criteria.

Inclusion criteria

- Randomised controlled trials
    - The trial investigator (or sponsor) indicated in their application to the REC that the intention was to prospectively allocate participants (or groups of participants) to an intervention using a random method. It may or may not have been possible to verify using the REC files (or in any resulting publication) that the study was actually randomised.
    - While some randomised trials included nested case-control studies, or other sub-studies, the comparison of interest in this study is the randomised comparison.
- Trials reported as full publications in peer-reviewed journals.

- Records were kept of other publications including short reports, letters and conference abstracts.

- Submitted to the REC for the first time between 1992 and 1996.
  - Note that approval to conduct a study is not given for an indefinite period and it must submitted for re-approval if there is a major change such as extension of the trial beyond the time period initially requested. Previously approved trials re-submitted during this time period were not considered eligible.

Exclusion criteria

- Abandoned trials
  - During the time period covered by this study many of the major funding agencies (including the NHMRC) required investigators to obtain ethics approval from at least one institution before submitting an application for funding. If a trial was successful in obtaining ethics approval, but was unsuccessful in obtaining funding, it was usual for it to subsequently be abandoned.
  - It is usually clearly evident in the REC file (based on the last annual report submitted by the trial investigator) if a trial was abandoned.
  - This includes trials ongoing elsewhere but abandoned at the site/s covered by the REC. Abandoned studies lacked complete documentation as the REC would not have been privy to protocol amendments and other relevant documentation.
  - Multi-centre trials that did not recruit patients at the site were not considered abandoned unless it had been reported as such to the REC.
- Trials published in short form only, including conference proceedings and letters, as there is usually too little information to assess completeness of reporting
- Trials reported in report form only (eg internal reports for the funding agency or sponsor, postgraduate theses, etc) and not published in a peer-reviewed journal.

Data collection

As files were not to be removed from the REC offices, all data were extracted on-site directly from the REC records. All data were extracted by the author. A second opinion was sought where necessary.

The data set was determined based on work previously conducted by the author and the need to address the research questions posed, and was also informed by the work of Chan and Altman (Chan AW 2003) and the requirements of the CONSORT statement (CONSORT Group 1996).

Definitions were determined for each data item to ensure consistency of definition and data extraction (see "Glossary").

Data were first extracted from the trial publication (see Appendix 5: Data collection forms (follow-up study). The REC file was then re-accessed and the remaining data on the trial extracted. The data from the publication was cross-checked with the data in the REC file and inconsistencies identified and coded. Data extracted from the protocol was kept distinct from the data extracted from the publication. The method of allocation concealment, for example, was collected both as documented in the protocol and as it was reported in the publication.

Study Endpoints

To avoid confusion the outcomes of the follow-up study will be referred to throughout this thesis as "endpoints", and the outcomes in trials as "outcomes". The details of each endpoint are addressed below.

As defined by Chan *et al*, an outcome is "a variable intended to be assessed in all study participants for the purpose of comparing the effects of interventions between randomised groups". (Chan AW & Altman 2005) Trial protocols and publications should declare at least one primary outcome, which should be the same in both documents. The primary outcome should be clearly declared however, if it is not, it is reasonable to infer that an outcome is a primary outcome if it is used to calculate the trial sample size, or is the outcome included in the trial's statement of main objectives or aims.

1. Is there a primary outcome in the protocol?

A trial was considered to have primary outcome in the protocol if between 1 and 4 outcomes could be clearly identified (or reasonably inferred) as being a primary. A primary outcome was considered to be clearly distinguishable if there is a clear statement that it is the primary (main) outcome. If there was no clear statement then was inferred that the primary outcome is that used to calculate the sample size, or stated in the aims or objectives. This was coded as:

0: No

1: Yes

If a primary outcome was reasonably inferred this was coded as:

0: outcome used to calculate sample size

1: outcome referred to in aims or objectives

2. Discrepancy in the identity of the primary outcome

This was coded as:

0: the outcome was identified in both documents as being a primary outcome

1: the outcome was identified in the protocol but not in the publication as being a primary outcome

2: the outcome was identified in the publication but not in the protocol as being a primary outcome

3. Discrepancy in the definition of the primary outcome

Much as a clinical research question is composed of a number of parts (patient / intervention / comparator / outcome) an outcome in a clinical trial can be composed of the name of the outcome, the time frame in which it will be measured and the instrument used to measure the outcome. For example:

- quality of life measured every 3 months using the SF36 (and how it will be reported: for example, as a single, global measure)
- airway responsiveness as indicated by a fall in FEV before and after treatment measured by spirometer

- nausea and vomiting over a 24 hour post-operative period measured by keeping a log of episodes (meeting pre-defined criteria)

An outcome definition was considered to be discrepant if one or more of the elements of the outcome had changed between the protocol and the publication. This was coded as:

0: the definition was the same in both documents

1: unable to judge if the definition was the same, either because the outcome was not recorded in one of the documents, or a definition was not provided

2: definitions were provided in both documents and they were different

4. Completeness of reporting of the primary outcome

Completeness of reporting of each comparison was classified based on the criteria described by in Appendix 4: Completeness of reporting. Each comparison was classified as:

Fully reported

Partially reported

Qualitatively reported

Not reported

In binary logistic regression analyses this outcome was coded as:

0: not fully reported (included partially reported, qualitatively reported and not reported)

1: fully reported

5. Is there a target sample size in the protocol?

Defined as any mention of a target number of participants to be recruited, with or without mention of a sample size calculation. This was coded as:

0: Yes, but no evidence of an appropriate calculation

1: Yes, with evidence of an appropriate calculation

Note: "No" was a third option however there were no studies without a target sample size included. Potential trial investigators are asked to provide the target sample size on the application form that is submitted to the REC.

6. Completeness of documentation of the sample size calculation in the protocol. This was coded as:
   0: Incomplete: there was no evidence of a sample size calculation OR there was some evidence of a sample size calculation but the outcome used was not determinable
   1: Partial: there was evidence of a sample size calculation and as a minimum the outcome used was determinable.
   2: Complete: there was evidence of a sample size calculation and the outcome, effect size, power and significance level were all provided.

7. Adequacy of reporting of the power calculation in the publication

This was coded as:
0: Inadequate: if no power calculation was mentioned in the report
1: Adequate: if a power calculation was mentioned with any amount of data in the report

8. Exclusions from analysis

Regardless of a claim by the authors of a publication that an intention-to-treat analysis was conducted, each trial was classified according to whether or not there were any participants excluded from the analysis. This was coded as:
   0: no exclusions reported (either explicitly reported or reasonably inferred that there were no exclusions. Eg denominators reported in analyses the same as the number randomised)
   1: there were exclusions (either explicitly reported that there were exclusions or reasonably inferred that there were exclusions)

9. Journal type

Coded as:
   0: specialty journal (eg the Journal of Cardiac Failure, the Journal of Clinical Oncology, etc)
   1: general journal (eg the Lancet, JAMA, etc)

The context for this follow-up study

This study will examine randomised trials submitted to an Australian REC (see The CSAHS Human Research Ethics Committee on page 43) for the first time between 1992 and 1996 (inclusive). The year 1992 was chosen as the starting year owing to the important changes in the process of ethical review resulting from the deregulation of the clinical trials industry the previous year. The Australian Therapeutic Goods Act shifted some of the responsibilities from the Commonwealth Government to RECs (also referred to as institutional ethics committees (IECs)). As described in Table 4, the main impact of deregulation on the work of RECs was the new responsibility to assess toxicology and safety. (Australian Government 1984;Commonwealth Department of Health and Ageing 2001b)). Although the primary role of RECs is to protect participants in research by "refusing approval to research projects which do not conform to acceptable ethical standards" (NHMRC 2001), deregulation had a substantial impact on the nature and volume of work conducted by those committees.

Table 4: The Commonwealth Therapeutic Goods Regulations (NHMRC 1995)

Under these Regulations, the institution has responsibility for:

- Conducting the trial
- Taking advice from the IEC on the conduct of the trial
- Giving approval to the trial (the institution may be responsible for more than one site)
- setting terms of approval for the trial which are no less restrictive than the ethics committee's advice; and
- *withdrawing approval for the trial if the ethics committee advises that continuation of the trial is not appropriate.*

The main impact of the deregulation of clinical trials, from the point of view of IECs, has been an expansion of their tasks and responsibilities to include assessment of toxicological and safety data for trials submitted under the Clinical Trials Notification (CTN) scheme.

If adequate expertise is not available amongst the members of an IEC to properly assess the scientific validity of a research protocol, or the data to CTX or CTN application, or for any other reason, the IEC should seek such expertise from outside its institution.

Under the Therapeutic Goods Act it is an offence to supply a therapeutic good unless it is listed on the Australian Register of Therapeutic Goods, which specifies the circumstances under which the good may be supplied. The exception are goods supplied for the purpose of research in humans (that is, for use in clinical trials), in which case the supplier of the good (that is, the trial Sponsor) must obtain approval from at least one AHEC-registered REC that has undertaken to monitor the trial.

AHEC-registered RECs must formally undertake to comply with the National Statement. (NHMRC 2001)

Another factor impacting on the work of RECs during the years examined in this follow-up study was the requirement of the National Health and Medical Research Council (NHMRC), Australia's major public funder of health and medical research at the time, that all grant applications must obtain ethics approval before submission to the NHMRC. As a result, a large number of applications receiving ethics approval were ultimately abandoned as the result of inability to obtain funding.

The CSAHS Human Research Ethics Committee

The Central Sydney Area Health Service (CSAHS) Human Research Ethics Committee (REC) is a registered REC with the NHMRC. It reviews proposals for research in humans to be conducted in 10 institutions (including 4 hospitals and 4 research institutes) in the central suburbs of Sydney. (See Appendix 3: About CSAHS REC) The Committee meets once each month and in the year 2004 reviewed more than 300 new protocols. It also monitors the progress and compliance of all ongoing and previously approved studies.

A Clinical Trials Sub-Committee (CTS) reviews all proposals for clinical trials in drugs or devices before the proposal is considered by the REC. The CTS includes individuals with appropriate scientific and clinical trial expertise and is responsible for reviewing all of the scientific data (including toxicology and pharmacology). External experts are also consulted if and when required. The recommendations of the CTS are forwarded to the REC who then take them into consideration when reviewing the remaining ethical requirements of the trial and respond to the individuals who submitted the application. The CTS was first established in 1991 in response to the additional expectations placed on the committee as the result of the deregulation of clinical trials that year.

The files kept by the REC included all documentation provided to the committee by the applicant including protocol amendments and annual reports. The files also included copies of all outgoing correspondence sent from the REC to the applicant.

Data was extracted from the most up-to-date version of the trial protocol, which often required sorting through protocol amendments to ensure the correct data was being obtained.

Identification of eligible randomised trials

As there was no paper summary or computerised record of the study type of each submission to the REC, potentially eligible trials were identified by systematically searching the file containing all correspondence (including the trial protocol) kept for each submission by the Research Office (RO) at Central Sydney Area Health Service (CSAHS). All records for each eligible year were accessed and a notation made of the study type (or types in the case of multiple studies in a single submission). All submissions to the REC are allocated a sequential identifying number, starting with the year of submission. The records for the years accessed are stored by year in archive boxes, and filed in the order of the REC identifying number.

When an RCT was identified a note was made of the status of the trial at the time of the last annual report, the date of the last annual report, and the details of any publications or presentations that had been notified to the REC.

The Sponsor (or Principal Investigator, as named in the REC file) of each trial was asked to give permission for their trial to be used in the follow-up study, and to provide additional information on the status of their trial and any resulting publications (see Appendices 5 and 6: *Investigator Form* and *Letter of Invitation*).

Identification of publications

Efforts were made to identify all publications reporting the results of each potentially eligible trial.

When investigators were contacted they were asked to give citation details for any publications resulting from the research. In addition, the MEDLINE database was searched using a combination of:

- the named principal investigators or collaborative group
- the intervention/s being investigated
- the patient population

In addition to Medline, specialty databases were also searched depending on the clinical area in which the research was being conducted: PSYCHINFO (for psychology) and CINAHL (for nursing), for example.

If the REC had been notified of any abstracts / conference proceedings (eg in an annual report) then a search of Medline (and/or other databases) was conducted using the authors and keywords identifiable in the abstract title.

A publication was deemed related to a particular trial if it could be reasonably ascertained that:

- the patient population was the same (allowing for possible unreported changes in the eligibility criteria)
- the intervention/s were the same or similar (allowing for possible unreported changes in the intervention such as the dose or schedule of a drug)
- the sample size calculation (if reported) was the same or similar
- the sample size achieved was the same as or similar to the sample size proposed in the REC file (allowing for possible problems in achieving the target sample size)
- the date of publication, and other dates mentioned in the publication (such as accrual dates) were aligned with the dates mentioned in the REC file
- the investigator provided citation details (although in two cases the incorrect citation was provided)

In the case of collaborative group studies, which are usually published under a group name, the appendix of the publication was searched for the name of the institution or the investigators at that institution, or the name of the collaborative group.

The date of the last search of Medline for possible publications was July 2005. In September 2005 the REC files for each RCT were revisited and updated information noted for each study regarding published abstracts or manuscripts. The relevant databases were searched only for those trials with new information.

Feasibility

As mentioned previously, Hahn *et al* attempted to conduct a similar study comparing 15 submissions (not restricted to randomised trials) made to an REC with their subsequent publications. The authors found that the lack of information in the protocols available to them made it difficult to assess selective reporting in this way. (Hahn, Williamson, & Hutton 2002) The current study was similar in design to that conducted by Hahn *et al*, but involved a much larger number of trials, and the sample was limited to randomised controlled trials, with the expectation that the protocols for randomised trials would be more comprehensive than applications made for non-randomised studies.

The data collection process was evaluated after data from the first 20 eligible trials submitted in 1992 had been extracted. The intention was to assess the feasibility of continuing the study and make improvements to the design and the data collection instruments if necessary. Alterations were made to the data collection instruments as a result of this pilot. All data collected on the first 20 eligible trials was re-collected using the updated instruments. As none of these trials documented the analysis plan it was decided to remove this item from the data collection forms.

Issues in the Identification and definition of primary outcomes

The identity of the primary outcome was not always clear in either the protocol or the publication. In many cases it was necessary to infer that an outcome was a primary outcome based on other information, such as the sample size calculation or the trial aims and objectives. A distinction was therefore made between primary outcomes that were clearly stated and those that needed to be inferred. A record was also kept of the way in which an outcome was inferred.

Outcome complexity

The complexity of outcome definition and construct made determining a mechanism for simply classifying the type of outcome in a way that would render it quantifiable a major challenge. The key issue is that outcomes are essentially multi-dimensional. A single outcome, for example, may be measured using multiple instruments over multiple time points. Each instrument may have multiple "dimensions" and each dimension may have multiple components or questions.

Each primary outcome was classified according to its complexity:

- Single outcome: where it is possible to identify a single comparison resulting in a single statistical test.
    - Single outcomes measured with multiple instruments were handled by treating each instrument as measuring a separate outcome.
- Composite outcome: combines 2 or more outcomes into a single outcome.
- Global outcome: is measured with an instrument composed of multiple dimensions (see Global Outcomes below).
- Multiple time points: where a single, composite or global outcome is measured over more than two time points (see Multiple time points on page 48)

Global Outcomes

The term "global outcome" is used to describe those outcomes where either a single or summary score could be calculated based on several components (eg quality of life), or is a multiple-item outcome with a method of analysis that includes all components in a single statistical test. The particular problem with these outcomes is that it is rarely clear if the outcome is an overall summary "score", or if each of the dimensions is to be reported separately. Some examples of global outcomes are:

- Quality of life: Quality of life instrument tend to be composed of multiple dimensions measuring the various aspects of quality of life as they pertain to the condition being investigated. An overall quality of life score may or may not be calculated and reported, and some/all of the components may be reported separately. In addition, each component may be further divided into sub-components eg specific questions. The protocol may or may not specify how quality of life will be reported (overall or component results).
- Multiple tests measuring a single outcome: eg lung function. It is likely that the multiple component assessments are reported with an assessment of overall function based on the test results. A specific example is lung function which may be measured using FEV1, FVC, FEV1/FVC. This type of outcome is common in exploratory trials.
- Calculated outcomes: An outcome calculated using two or more single outcomes where the component parts may or may not be reported separately. Eg infant

growth, which could be 3 simple measures of length, head circumference and weight, or as one or more calculations based on these components, eg length x weight, head circumference x length.

- Compliance: In a trial comparing two lipid-lowering diets compliance was measured using various blood levels including saturate fat, mono-unsaturated fat, complex carbohydrate, etc. No single measure of compliance was specified.

Multiple time points

Outcomes measured over multiple time points may or may not have specified a particular time point as the primary outcome. If each time point is counted as a separate outcome then this may place undue emphasis on outcomes with multiple time points. Outcomes with multiple time points were classified as:

- Multiple time points measured but clear single time point is identifiable as the primary timepoint
- 2 time points specified
- More than 2 time points specified

Composite outcomes

Composite outcomes are usually used to increase the event rate and hence increase the power of the study to detect a difference. Issues in the reporting of composite outcomes are:

- The addition or removal of one or more outcomes from (components of) a composite outcome between the protocol and the publication
- The separate reporting of the composite outcome and each of the elements of the composite. It may or may not have been intended to report both the composite and one or more elements as primary outcomes. This issue may appear in trials where the ideal situation would be an available target population that would enable the single primary outcome of interest to be addressed, but where the available target population is limited.
- The elements of the composite might be reported in various combinations and it may be difficult to ascertain which one, if any, is the "primary" combination.

Multiple calculations based on a single instrument

There were occasions when the results of a single test or instrument were applied and used in multiple ways. Trial investigators may be interested in the actual value, the number of participants with a value over or under a specific value (threshold), or the change in a value. Some examples include:

- Crohn's Disease Activity Index (CDAI), used to calculate an overall score at multiple time points which was in turn used to determine change in CDAI score, and relapse (defined as a particular CDAI score).
- sperm count, used to calculate: total sperm concentration, azoospermia and oligozoospermia (both defined as a particular levels of sperm concentration)

In these situations, each use of the test or instrument was considered to be a different outcome unless it was explicitly declared otherwise in either the protocol or the publication.

Multiple measures

Some outcomes can be measured using more than one instrument or test. There are multiple instruments that could be used to measure the outcome "depression", for example. It is not uncommon for a single outcome (such as depression) to be clearly stated as being the primary outcome, with a number of instruments used to measure this single outcome. Problems arise when a single instrument is not declared as the primary instrument of interest. In this follow-up study, each instrument was considered to be a separate primary outcome.

Trials with more than 4 primary outcomes

Trials with more than 4 primary outcomes in either the protocol or the publication were not considered to have a clearly identifiable primary outcome.

In cases where 4 or less outcomes were identified in the protocol, and 4 or less outcomes were identified in the publication, but 1 or more was a new outcome (ie resulting in more than 4 outcomes overall) then all of the outcomes were included.

Process issues

A number of issues related to the conduct of the trial may impact on the way in which data on a primary outcome is collected or interpreted. It was not possible to address these issues in this study as they are unlikely to be detected in the trial protocol or the publication. Process issues include:

- Outcomes that are adjudicated (by a central process such as a committee) versus the same outcome as reported by investigator
- The way in which an instrument is administered. One example is the participant who completes the instrument themselves versus the investigator or a study nurse asking the questions on the instrument and completing it on behalf of the participant. Another example might be when the participant is asked to complete the instrument in the clinic while another may be invited to take the instrument home. The environment may influence the responses.

Notes re reported p values

When a p value for a comparison was not reported, but the publication states "there was no significant difference" or similar, then the comparison was classified as $p>0.05$. If there was no mention of the significance level then this was recorded as "not reported".

The outcomes from one trial (a small (28 participants), 7-period crossover, dose-response, exploratory study in a healthy population) had multiple pair-wise comparisons. Although no p value was reported the 95%CIs indicate that there was a significant difference so the p values for the relevant comparisons in this trial were entered as <0.05.

Other considerations

During the course of conducting the study and extracting the data from the REC files it became evident that there were a number of factors that needed to be taken into consideration in order to make conducting the study feasible and meaningful.

Multi-arm trials

A trial with 3 treatment arms could potentially contain up to 6 discrete treatment comparisons: AvB, AvC, BvC, AvBorC, BvAorC, CvAorB; each of which might be subject to an analysis. Data on primary outcomes for trials with more than two arms were entered according to the comparisons reported in the publication. The way in which comparisons were handled in the analysis is described in more detail in Chapter 3 ("Creating trial-based measures" on page 57).

- Data for all comparisons reported were collected regardless as to whether or not they were specified in the protocol.
- There were trials that randomised to 1 of 3 or more arms where the primary comparisons were explicitly reported. Eg one included trial specified that the primary comparisons were: AvBorC, AvB and AvC. The BvC comparison was also reported although it was clearly stated that this was not a primary comparison. Data from this comparison was therefore not collected in this case.
- The "control" arm may change from comparison to comparison in trials with more than 2 arms. A decision regarding the identity of the control arm (or arms) was determined based on data reported in the publication.
- If there were 3 or more arms and a simple statement along the lines of "no significant difference was detected" then it was assumed that all possible pair-wise comparisons were tested.

Safety and adverse events

> *"While there is an ethical obligation to monitor for serious unexpected adverse events, one suspects the process of reporting and categorizing is often inconsistent. Usually the lack of any prespecified hypotheses or priorities leaves one at risk of data dredging"*. (Pocock 1997)

Pocock describes reporting of adverse events as multiple outcomes "gone crazy". (Pocock 1997) In many clinical trials the protocol will state that safety (also referred to as toxicity or tolerability) is an outcome, or that a number of tests may be performed routinely to monitor safety, without describing a specific safety outcome of interest. A general comparison will be made between the protocol (that is, was there evidence that the intention was to monitor safety) and the publication (were any safety outcomes reported).

Choosing a comparator

It is not always clear which arm of a trial is the comparator. For example, when both arms of a 2-armed trial are considered to be "standard treatment". In such cases, even if there was a statistically significant difference, the direction of the treatment effect was coded as "not applicable". Trials with more than 2 arms, where it is not possible to determine which arm (or arms) are the comparator arms, were handled in the same way.

Explanatory versus pragmatic trials

The purpose of a trial, ie whether it is explanatory or pragmatic, is a trial characteristic that may be associated with selective reporting, and hence is a variable included in the regression modelling described on page 56. Although a distinction is often made between explanatory (also referred to as exploratory or efficacy) and pragmatic trials (or effectiveness, sometimes referred to as confirmatory), there is no universally agreed definition, and many trials fit somewhere between the two.

It is generally agreed that explanatory research asks whether an intervention works under ideal or selected conditions. It is more concerned with how and why an intervention works. Explanatory trials:

- generally measure efficacy
- recruit as homogeneous a population as possible
- often use intermediate (or surrogate) outcomes
- are more relevant to examining biological effects
- endeavour to discover whether a treatment effect exists
- are highly controlled and idealised in the spirit of a laboratory experiment

Pragmatic research asks whether an intervention works under "real-life" conditions and in terms that matter to the patient. It is concerned with whether the intervention works, not how or why.

Explanatory trials may have smaller sample sizes so one way to distinguish between explanatory and pragmatic trials might be to use a threshold for sample size. A problem with this approach is that in some situations, smaller sample sizes may

simply reflect the nature of the target population (eg a rare condition) rather than the purpose of the trial.

Explanatory versus pragmatic designs are considered to be a potential prognostic factor for reporting of outcomes (eg latter more likely to conduct exploratory analyses). Trials were therefore categorised as explanatory or pragmatic based on the judgement of the author using the above definitions.

Issues in the conduct of methodological research in this area

In this series of trials it has been necessary to make judgements about clinical trials based on the information documented in the protocol and other records kept by the REC. This may or may not be an accurate reflection of the trial as actually planned. Although there are no doubt problems with relying on the quality of the recorded documentation, it is the best source of information available. Absence of the detail required for the purposes of this follow-up study does not mean each trial did not perform particular tasks – it may just mean that they did not write them down. Lack of information in the protocol on allocation concealment and sequence generation, for example, does not mean they did not adequately manage the randomisation process.

Following initial approval, if significant changes are made to the protocol, it is usual for a protocol amendment (or some other form of communication) to be submitted to the REC. It is likely that not all changes made to a protocol during the course of a study will be captured.

Another issue to keep in mind is that the documentation from which data was extracted was created for a specific purpose: that is, to obtain ethics approval. Although the information required for the purpose of ethical review is aligned with that required for this follow-up study it is possible that the ethics committee was privy to additional information as part of its decision-making process. It is not uncommon, for example, for investigators to attend the ethics committee meeting at which their proposal is being discussed. This will give them the opportunity to deal with specific questions that may not be documented in the REC record. In the case of CSAHS REC issues raised during the course of an REC meeting are usually documented in formal

minutes, and those minutes form the basis of written communication following the meeting with the investigator. A copy of all such communication is kept in the REC record, and the complete file was accessed for this follow-up study.

Some issues are time-dependant and difficult to assess retrospectively. The appropriateness of the comparator, for example, or the size of the treatment effect, are very much dependant on our understanding of the condition and the appropriate standards of care at the time the trial was submitted to the REC.

Chapter 3: Description of trials included in follow-up study

This chapter includes:

- information on the analyses conducted, including
    - factors that may be associated with selective reporting (used as covariates in logistic regression)
    - how trial-based measures were created
- the identification of eligible trials
- descriptive information about the identified trials
- descriptive details extracted from the trial protocol
- descriptive details extracted from the trial publication

Data analysis

All analyses were conducted using the statistical package SPSS 13.0 for Windows (SPSS Inc., 233 S. Wacker Drive, Chicago, Illinois 60606).

Features of the protocol and publication are described as frequencies in tables, with 2x2 tables constructed to investigate relationships within protocols, and within publications, and tested using Chi square (Fisher's exact test) when appropriate.

Trial characteristics reported in the protocol (listed on page 56) that may be associated with the endpoints relating to selective reporting were investigated using logistic regression. The endpoints relating to selective reporting are:

- Discrepancy in the identity of the primary outcome
- Completeness of reporting

These outcomes were investigated in the logistic regression models as trial-based endpoints, rather than outcome or comparison-based endpoints. The method used to create trial-based endpoints is described on page 57).

All variables were investigated univariately, as well as adjusted for the number of outcomes (for the trial-based endpoint for "Discrepancy in the identity of the primary

outcome") or the number of comparisons (for trial-based endpoint for "Completeness of reporting"). Independent variables that were significant either on univariate analysis, or adjusted for the single covariate number of outcomes or comparisons, were included in a multivariate model.

Trial characteristics that may be associated with selective reporting

Trial characteristics in the protocol that may be associated with the reporting of the trial were selected *a priori* for investigation using logistic regression. The coding of each of these variables for inclusion in the regression model is described in Chapter 4 (Table 28).

- Design (crossover or parallel)
- Purpose (exploratory or pragmatic)
- Administration (single centre, multi-centre national, multi-centre international)
- Commercial funding available at the time of submission to the REC
- Proposed sample size
- Completeness of documentation of the sample size calculation in the protocol
- Allocation concealment (adequate or inadequate)
- Sequence generation (adequate or inadequate)
- Blinding
- Number of outcomes in a single trial for the endpoint "Discrepancy in the identity of the primary outcome" OR number of comparisons in a single trial for the endpoint "Completeness of reporting"

Note: see Glossary for definitions used for each variable.

Creating trial-based measures

A single trial can have more than one primary outcome, and each outcome can involve more than one comparison. Take, for example, a trial with 3 treatment arms (A, B and C) and 2 primary outcomes (X and Y). Such a trial could potentially have 6 treatment comparisons:

- Comparison 1: AvB for outcome X
- Comparison 2: AvC for outcome X
- Comparison 3: BvC for outcome X
- Comparison 4: AvB for outcome Y
- Comparison 5: AvC for outcome Y
- Comparison 6: BvC for outcome Y

Most of the trial characteristics considered to be potentially associated with the reporting of trials are trial-based. They include, for example, information on trial design, purpose, sample size, etc. A number of the follow-up study endpoint variables are, however, either outcome or comparison-dependant. That is, the value of those variables will not be the same for each outcome or comparison within each trial (Table 5).

Table 5: Outcome-dependant and comparison-dependant variables

The outcome –dependant endpoint is: • Discrepancy in the identity of the primary outcome The comparison–dependant endpoint is: • Completeness of reporting The comparison-dependant covariate is: • Level of statistical significance

Comparisons within a trial are unlikely to be independent, and it would not be appropriate to treat each comparison as a separate event when investigating the above variables. For this reason a trial-based measure was created for each of the events described in Table 5 by determining the frequency of values for each variable for each trial. Building on the above example, and using the variable "statistical significance", the result could be:

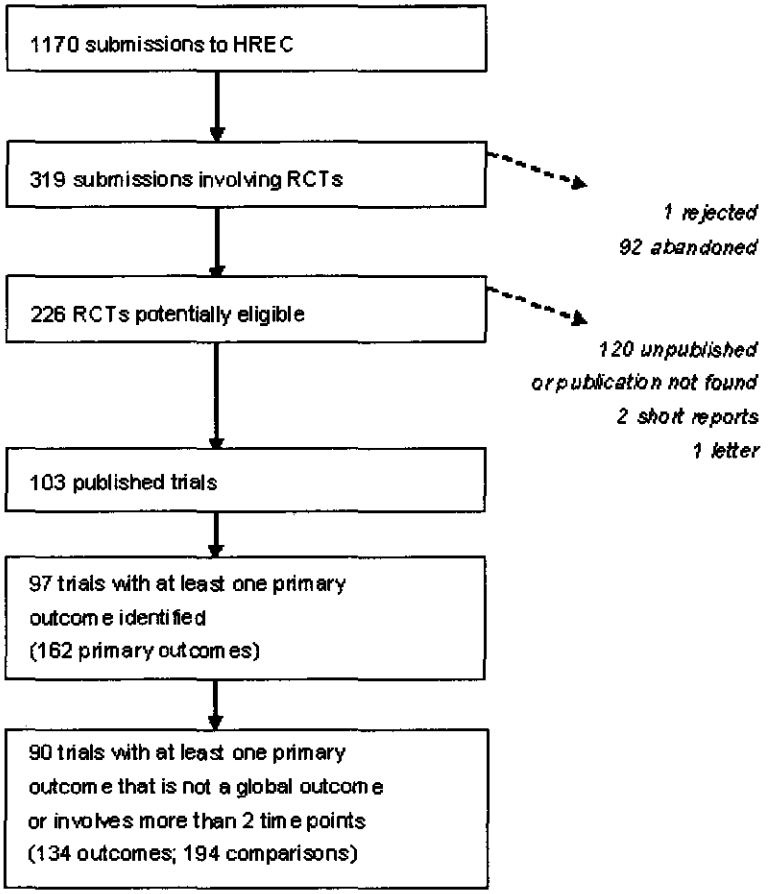
- Comparison 1: AvB for outcome X: p>0.05
- Comparison 2: AvC for outcome X: p>0.05
- Comparison 3: BvC for outcome X: p<=0.05
- Comparison 4: AvB for outcome Y: p>0.05
- Comparison 5: AvC for outcome Y: p>0.05
- Comparison 6: BvC for outcome Y: p value not reported

The frequency of statistically significant comparisons for this trial is 17% (1 in 6). That is, 17% of comparisons in this trial met conventional statistical significance (p<=0.05). A threshold for a "positive" trial can then be set based on the expected acceptable frequency of a statistically significant result. It is possible to raise or lower the threshold depending on the level considered acceptable for the specific variable under investigation.

Identification of trials

There were 1170 submissions to the REC between 1/1/92 and 31/12/96. (See Figure 4 and Table 6) Randomised trials were involved in 319 (27%) submissions of which 92 (29%) were ultimately abandoned, and 1 was rejected. The search strategy described in Chapter 2 was applied to each of the remaining 226 submissions, and a publication was identified for 106 (46%): two publications were short reports and 1 was a letter hence full publications were available for 103 trials. One submission involved 2 related trials published in a single publication.

Figure 4: Identification of studies



As it was a requirement of the REC that permission be obtained to use each trial in the follow-up study, the site-specific Principal Investigator for each trial (that is, the individual responsible for the submission at the site) was approached to inform them of the objectives of the study, to obtain permission to include their trial in the follow-up study, and to request information on publications resulting from the trial. If the Principal Investigator was not contactable (that is, was now deceased, or was no

longer employed at the hospital and not locatable) then one or more of the remaining investigators was contacted. Permission was obtained from at least one investigator for each trial for which a publication was identified.

In one submission the trial could be considered to be either one or two trials, with 2 sample size calculations for 2 similar patient populations (differing only by stage of disease). This trial has been treated as a single trial for the purposes of describing and comparing trial details, but the outcome details for each trial have been considered separately.

Most trials were in cardiovascular disease (20%) or cancer (19%) followed by respiratory medicine (8%) and gastroenterology (7%). (See Table 7). A total of 28% were placebo controlled and most (70%) had at least one comparator arm that involved one or more drugs (See Table 8). Most trials in cardiovascular disease and cancer (90% and 84% respectively) had at least one comparator arm that involved one or more drugs (See Table 9).

The majority of trials had either two (72) or three (24) treatment arms and were multi-centre (70), 76% of which (53) were international (See Table 10). Most were parallel, pragmatic trials by design (See Table 11).

Table 6: Identification of trials

Submissions	**Year of submission**					
	1992	**1993**	**1994**	**1995**	**1996**	**Totals**
Number made [See Note 1]	225	213	218	261	253	1170
Number involving RCTs	63	61	53	74	68	319
Number involving abandoned RCTs (% of RCTs identified)	24 (38%)	15 (25%)	16 (30%)	21 (28%)	17 (25%)	92 (29%)
Number rejected	0	0	0	0	1	1
Number not abandoned	39	46	37	53	50	226
Number RCTs with publications identified (% of those not abandoned)	21 (54%)	17 (37%)	14 (38%)	24 (45%)	30 (60%)	106 (47%)

Note 1: each submission could involve more than one study

Table 7: Patient population

Patient population	Number of trials	Patient population	Number of trials
Cardiovascular disease	20	Nephrology	1
Critical Care	1	Neurology	4
Dermatology	3	Obstetrics and gynaecology	4
Drug and alcohol	3	Oncology	19
Endocrinology	3	Paediatrics	4
Fertility	3	Physiology	1
Gastroenterology	7	Psychology	2
Geriatric care	2	Respiratory medicine	8
HIV / AIDS	6	Rheumatology	2
Haematology	2	Surgery	1
Immunology	2	healthy/normal	5

Table 8: Interventions

Nature of intervention	arm 1 (control)	arm 2	arm 3	arm 4
Placebo	29	0	0	0
Drugs	46	72	26	6
Surgery/Procedure	5	4	0	0
Device	0	1	1	0
Lifestyle	2	2	0	0
Counselling	3	3	1	0
Rehabilitation	1	1	0	0
Other intervention	17	20	4	0
Total	103	103	32	6

Table 9: Intervention by patient population

	Drug v drug	Placebo v drug	Placebo v device	Surgery v surgery	Lifestyle	Couns-elling	other
Cardiovascular	10	8	0	1	1		
Oncology	15	1	0			1	2
Respiratory	1	5	1				1
Gastroenterology	5	1		1			
HIV/AIDS	4	1					1
Healthy/normal		2			1		2

Note: For areas with 5 or more trials, based on arm1 x arm2

Table 10: Trial details (n = 103)

Characteristic	Classification	Number
Number of trial arms	2	72
	3	24
	4	4
	5	1
	6	1
	7	1
Single or multi-centre	Multi centre	70
	Single centre	33
National or international	International	53
	National	17
	not applicable	33
Design	Crossover	12
	Parallel	91
Purpose	Explanatory	23
	Pragmatic	80

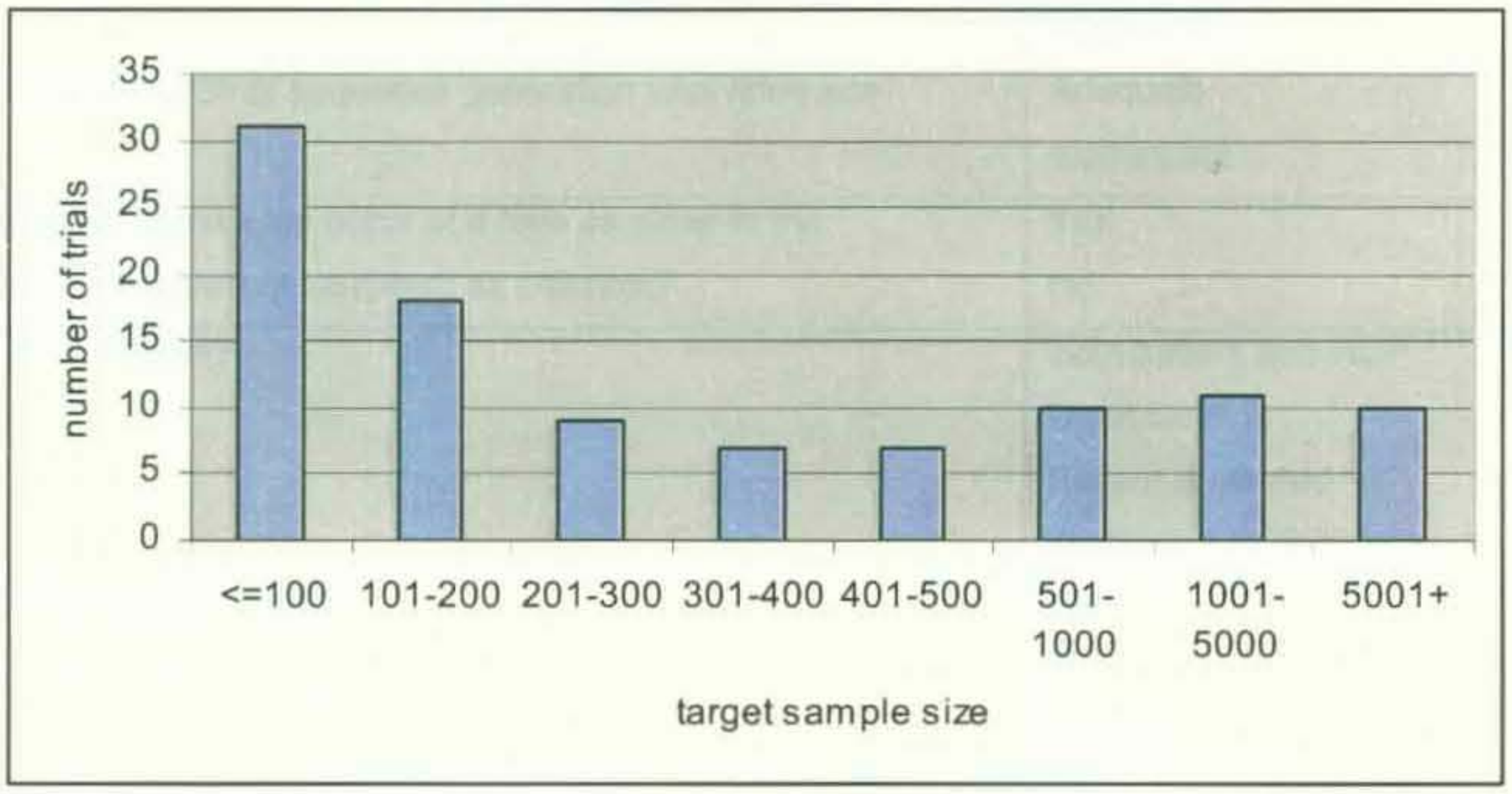
Table 11: Trial design by purpose

Design	Trial Purpose		Total
	Explanatory	Pragmatic	
Crossover	11	1	12
Parallel	12	79	91
Total	23	80	103

From the trial protocol

The median target sample size for trials as a whole was 240 (range 20-40,000), with 30% of trials having a target sample size of 100 or fewer (Figure 5).

Figure 5: Target sample size



At the time of submission to the REC most trials (65) reported that commercial funding was available to support the conduct of the trial (See Table 12).

Information on allocation concealment and sequence generation was poorly documented in the protocols evaluated. A total of 60 trials used some form of masking of the interventions being compared, and 48 used a placebo of which 31 described the placebo but only 17 adequately.

Table 12: Descriptive details in trial protocol

Characteristic	Classification	Number
Was any commercial funding available to support the conduct of the trials at the time of submission to the REC?	No	38
	Yes	65
Is there at least one clearly distinguishable primary outcome in the protocol	No	9
	Yes	94
The description of allocation concealment was adequate	Adequate	44
	Inadequate	59
The description of sequence generation was adequate	Adequate	31
	Inadequate	72
Does randomisation occur at a time as close to the commencement of treatment as possible?	Yes	21
	No	82
Use of blinding	Both patient and HCP	55
	Open label	42
	Patient is blinded	4
	Treatment HCP blinded	2
Was a placebo used?	Yes	48
	No	54
	Unclear	1
Was there a description of the placebo?	Yes	31
	No	17
	Not applicable	55
Was the description of the placebo adequate?	Yes	17
	No	14
	Not applicable	72
Did the protocol state that adverse events would be monitored?	Yes	82
	No	21
Did the protocol include a section on the handling of withdrawals?	Yes	57
	No	46

Note: the wording for some of the questions presented above differs slightly from the wording on the data collection forms. The data collection forms used shorthand terms and were applied in combination with a set of standard operating procedures developed specifically for the purpose. All tables are an accurate reflection of these SOPs.

From the trial publication

Most trials were published in speciality journals (Table 13). The median sample size achieved (by trials as a whole) was 230 (range 6-58,000). There was mention of a power calculation in 69 of the publications resulting from the 103 trials. Use of a placebo was reported in 48 trials, only 7 of which provided a description that allowed the reader to determine that the placebo was adequate (Table 14).

In the publication, allocation concealment was clearly described as using a centralised telephone, fax or computerised system in 15 trials. In a further 9 trials it was stated that randomisation was conducted "centrally" but it was unclear what this actually entailed (Table 15). Envelopes systems were used in 9 trials but only 1 provided clear evidence that the envelopes were free from the potential biases inherent in such systems. Three trials described methods of allocation concealment that were inadequate.

The methods of sequence generation reported in the publication included schemes generated by a computer (16), dynamic balancing (4) or minimisation (5) techniques, random number tables (6) and permuted blocks (7). In one trial the patient appeared to select their treatment (clearly not adequate). The remaining 64 trials did not report the method of sequence generation.

Table 13: About the publication

Characteristic	Classification	Number
Journal type	General	26
	Specialty	77
Type of report	full publication	103
	short report	2
	letter	1
Was any commercial funding mentioned in the publication?	No	39
	Yes	64
Number of subjects randomised	≤100	33
	101-500	37
	501-1000	12
	>1000	21
Was there at least one clearly distinguishable primary outcome in the publication?	No	13
	Yes	90
Were any safety outcomes reported?	No	15
	Not applicable	11
	Yes	77
The reporting of the power calculation was:	Adequate	69
	Inadequate	34

Table 14: Blinding and placebo in publication

Characteristic	Classification	Number
Was a placebo used?	No	54
	Yes	48
	Unclear	1
If a placebo, the description of the placebo was:	adequate	7
	inadequate	42
	not applicable	54
If no placebo, is masking used?	Yes	9
	not applicable	94
If masked, the description of masking was:	adequate	3
	inadequate	2
	unclear	5
	not applicable	93

Table 15: Allocation concealment and sequence generation in publication

Characteristic	Classification	Number
The method of allocation concealment	Central phone, fax or online	15
	Central system: method unclear	9
	Sealed envelopes: acceptable	1
	Sealed envelopes: not acceptable	8
	Other	3
	Not reported	67
Classification of allocation concealment	Adequate	16
	inadequate	87
The method of sequence generation	Computer	16
	Dynamic balancing	4
	Minimisation	5
	Permuted blocks	7
	Random number table	6
	Other	1
	Not reported	64
Classification of sequence generation	adequate	38
	inadequate	65

Chapter 4: Results

This chapter includes the main results of the follow-up study. These will be reported as follows:

- Primary objective: to investigate the selective reporting of primary outcomes
- Multiple time points and global outcomes
- Secondary objective 1: to explore issues relating to the sample size, its calculation and reporting
- Secondary objective 2: to explore other relationships between the protocol and the publication
- Secondary objective 3: to explore the impact of the availability of commercial funding on the trial protocol and publication

As outlined in Figure 4, 103 trials have been included, 97 of which had at least one primary outcome (a total of 162 outcomes). Excluding global outcomes and outcomes with more than 2 time points resulted in the inclusion of 90 trials with 134 outcomes and 194 comparisons. The data set used to address each question within each objective will therefore differ depending on the nature of the question.

Primary Objective: to investigate the selective reporting of primary outcomes

a. Do randomised clinical trials start to address multiplicity early in the history of the trial through the clear declaration of a primary outcome in the protocol?

Of the 103 trials in the study, six listed so many outcomes in the protocol and the publication that it was not possible to identify 4 or less as primary outcomes, leaving 97 trials reporting a total of 162 primary outcomes (including global outcomes and outcomes with more than 2 time points). It was possible to identify a single primary outcome in 59 trials, 2 primary outcomes in 20 trials and 3 or more primary outcomes in 18 trials (see Table 16). 60% of explanatory trials and 59% of pragmatic trials had a single outcome.

Table 16: Number of primary outcomes by trial purpose

Number of outcomes	Purpose		Total
	Explanatory	Pragmatic	
0	2	4	6
1	12	47	59
2	7	13	20
3	2	8	10
4	0	6	6
5	0	1	1
6	0	1	1
Total	23	80	103

Note: In the case of 2 included trials there were up to 4 primary outcomes identifiable in each document (ie 4 in the protocol and 4 the publication, however the identity of the outcomes differed in each document resulting in more than 4 primary outcomes for these 2 trials.

b. Are there discrepancies between the trial protocol and the trial publication in relation to the identity of the primary outcome?

Using the 97 trials with at least one primary outcome identified, there were coincidentally 97 of 162 outcomes (60%) that were clearly stated as being primary outcomes in the protocol, plus 40 (25%) that could be reasonably inferred as being primary outcomes (based on the sample size calculation or the aims and objectives)

(Table 17). In 25 cases (15%) the primary outcome subsequently identified in the publication was not documented as being a primary outcome in the protocol. Of these 25, 9 were declared as being a secondary and not a primary outcome in the protocol, 5 were mentioned in the protocol as an unspecified outcome, and 11 were not mentioned at all in the protocol.

64% of pragmatic trials clearly state the primary outcome in the protocol (62% in the publication), compared to 44% of explanatory trials (25% in the publication) (Table 18 and Table 19). 47% of explanatory trials reasonably infer the outcome is a primary in the protocol (50% in the publication) compared to 19% of pragmatic trials (24% in the publication).

There were 27 outcomes that were either reasonably inferred (12) or clearly stated (15) as being a primary outcome in the protocol, but were not reported as a primary outcome in the publication. Most (21 (78%)) were reported but the nature of the outcome was not specified, 5 were reported as secondary outcomes and 1 was not reported at all. The level of significance of a comparison did not appear to be associated with the decision to change the identity of the primary outcome in the time between the protocol and the publication (Table 20) (this was not tested statistically).

Table 17: How you can tell it is the primary outcome

How can you tell the outcome is a primary outcome in the protocol?	How can you tell the outcome is a primary outcome in the publication?			
	Not stated	Reasonably inferred	Clearly stated	Total
Not stated	0	13	12	25
Reasonably inferred	12	18	10	40
Clearly stated	15	16	66	97
Total	27	47	88	162

Table 18: How you can tell it is the primary outcome (protocol) by trial purpose

How can you tell it is a primary outcome in the protocol	Purpose		Total
	Explanatory	Pragmatic	
Not stated	3 (9%)	22 (17%)	25
Reasonably inferred	15 (47%)	25 (19%)	40
Clearly stated	14 (44%)	83 (64%)	97
Total	32	130	162

Table 19: How you can tell it is the primary outcome (publication) by trial purpose

How can you tell it is a primary outcome in the publication	Purpose		Total
	Explanatory	Pragmatic	
Not stated	8 (25%)	19 (15%)	27
Reasonably inferred	16 (50%)	31 (24%)	47
Clearly stated	8 (25%)	80 (62%)	88
Total	32	130	162

Table 20: Change in identity of primary outcome by statistical significance

Level of statistical significance	Change in identity of primary outcome			Total
	Outcome used for comparison declared as primary outcome in publication but not in protocol (% of total)	Outcome used for comparison declared as primary in protocol but not in publication (% of total)	Outcome used for comparison declared as primary in both protocol and publication (% of total)	
P value not reported	4 (13%)	8 (25%)	20 (63%)	32
$P \leq 0.05$	8 (13%)	10 (16%)	45 (71%)	63
$P > 0.05$	8 (8%)	13 (13%)	78 (79%)	99
Total	20 (10%)	31 (16%)	143 (74%)	194

c. Are there discrepancies between the trial protocol and the trial publication in relation to the definition of the primary outcome

Using all 103 trials, it was possible to reasonably deduce that the definitions for 98 (of 162, or 60%) outcomes were the same in both the protocol and the publication. It was not possible to judge in the case of 51 outcomes (31%), primarily because the outcome was not declared in either the protocol or the publication.

It was possible to ascertain that the definitions were different in the case of 13 (8%) primary outcomes (see Table 21). These differences can be summarised as:

- addition of an event (or outcome) to a composite outcome (1 outcome)
- change in the tests performed (or the way in which the tests were performed) that were used to determine the primary outcome (2 outcomes)
- change in the definition of a positive test or an event (7 outcomes)
- change in the specified time point of interest (3 outcomes)

Table 21: Nature of the difference in definition

Trial	Nature of the difference in definition
A (2 trials)	There was a change in the definition of the outcome during the course of the trial, and this change was documented in the publication, however, appear to use a different definition again for reporting
B	Added an outcome to a composite outcome
C	There was a change in the definition of the outcome during the course of the trial, and this change was documented in the publication.
D	This outcome was "safety" and the definitions in the protocol and the publication were slightly different, and the impact of the difference is unclear.
E	Time period is different, and quite important in this trial. Protocol specifies week 4 and the publication specifies week 1. In addition, the outcome measurement in the protocol uses 2 specific tests with specific values required; the publication uses one of these tests but specifically excludes the results of the second test.
F (2 outcomes)	The test used to measure the outcome is different. Participants were asked to perform a repetitive task a given number of times during a specific time period. The number of repetitions was double in the publication to the number described in the protocol.
G (2 outcomes)	The outcome was measured by the total dose of a drug administered over a given time period. The protocol states the time period as being 2 days, while the publication states it as being over 5 days. The second outcome was based on 4 criteria in the protocol, and 5 criteria in the publication with overlap of 3 criteria.
H	The outcome is "cure" with specific tests and results required. The protocol and publication both require these at week 4, but the publication states an additional test with a specific result was also required at the end of week 1.
I	The difference in the definition of the outcome could potentially result in a different event rate.
J	The outcome of this trial was willingness to perform one of 2 tasks. One of the tasks was abandoned for practical reasons. This change in the outcome was reported and unlikely to have an impact on the findings of the trial.

d. Are primary outcomes "fully" reported?

As described in Chapter 3, a single trial may have more than one primary outcome, and each outcome may involve more than one comparison if the trial has more than 2 treatment arms. As the completeness of reporting may be comparison-dependent it is necessary to divide each outcome down into its comparisons to look at whether and how each of these was reported in order to be able to investigate completeness of reporting.

Difficulties in quantifying outcomes with multiple time points and global outcomes in a meaningful way mean that these outcomes have not been included in the analysis of completeness of reporting of primary outcomes and have been discussed separately (see Multiple time points and global outcomes). A total of 13 trials did not contribute to the analysis of completeness of reporting of primary outcomes: 6 that did not have identifiable primary outcomes, and 7 where the only primary outcomes in the trial involved multiple time points or were global outcomes. The remaining 90 trials contributed 134 primary outcomes to the analysis of the selective reporting of primary outcomes. Twenty (15%) of the 134 were composite outcomes.

Breaking the outcomes with 2 identifiable time points, and trials with more than 2 treatment arms, into comparisons resulted in a total of 194 comparisons available for analysis of completeness of reporting of the primary outcome. 45 trials contributed a single comparison and 15 contributed 4 or more (Table 22). Most comparisons involved binary (39%) or continuous (36%) data (Table 23 and Table 24).

Table 22: Number of comparisons in a trial by trial purpose

Number of comparisons	Purpose		Total
	Exploratory	Pragmatic	
1	7	38	45
2	7	10	17
3	4	9	13
4	1	6	7
5	0	2	2
6	1	4	5
8	0	1	1
Total	20	70	90

Table 23: Comparison classification by data type

Comparison classification	Data type					Total
	binary	categorical	continuous	not reported	time to event	
additional comparison	21	2	22	0	5	50
composite outcome	15	0	0	0	5	20
extra time point	3	0	7	0	0	10
single outcome	36	3	41	5	29	114
Total	75	5	70	5	39	194

Table 24: Number of comparisons by data type

Number of comparisons	Data type					Total
	binary	categorical	continuous	not reported	time to event	
1	52	3	41	5	34	135
2	9	1	19	0	3	32
3	6	1	9	0	2	18
4	2	0	1	0	0	3
5	2	0	0	0	0	2
6	2	0	0	0	0	2
7	1	0	0	0	0	1
8	1	0	0	0	0	1
Total	75	5	70	5	39	194

Completeness of reporting of comparisons

The majority of comparisons were either fully (77%) or partially (12%) reported (Table 25) (See Appendix 4: Completeness of reporting for definition). 93% of binary outcomes and 69% of continuous outcomes were fully reported, possibly reflecting the greater level of detail required for a continuous outcome to be considered fully reported.

Table 25: Completeness of reporting of comparisons by data type

	Data type					
Completeness of reporting	binary	categorical	continuous	not reported	time to event	Total
fully reported	70	5	48	0	26	149
not reported	2	0	3	5	3	13
partially reported	0	0	18	0	5	23
qualitative	3	0	1	0	5	9
Total	75	5	70	5	39	194

Trial-based measure of completeness of reporting

Using the method described in Chapter 3, a trial-based measure was created for completeness of reporting. Based on this measure, 76% of trials fully reported all of their primary outcomes, 10% did not fully report any of their primary outcomes, and 14% fully reported some but not all of their primary outcomes (Table 26).

Table 26: Number of trials and proportion of comparisons completely reported

Proportion of comparisons fully reported	Number of trials	Percent
0%	9	10.0
25%	3	3.3
33%	1	1.1
40%	1	1.1
50%	2	2.2
67%	3	3.3
75%	2	2.2
80%	1	1.1
100%	68	75.6
Total	90	100.0

Threshold for a fully reported trial

The threshold for a fully reported trial was set at 100%. That is, trials that fully report all (ie 100%) of their comparisons are considered to be fully reported, and are compared with those that fully report less than 100% of their comparisons. A sensitivity analysis was conducted by changing the threshold for a completely reported trial to 66% (that is, a completely reported trial is one that fully reports more than 66% of its comparisons). The impact of changing the threshold to 0% (trials that do not fully report any of their comparisons are compared with those that fully report at least one comparison) was also investigated. It should be noted that there are only 9 trials that do not fully report any of their comparisons, and the small numbers make this analysis unreliable.

There were 45 trials with a single comparison, of which 39 (87%) were fully reported (Table 27).

If the threshold for a fully reported trial is 100% then 68 trials (76%) would be classified as fully reported. If it is changed to 66% then 74 trials (82%) would be considered to be fully reported. A threshold of 0% indicates that 81 trials (90%) fully report at least one comparison.

Table 27: Number of comparisons by number of trials with 100% fully reported comparisons

Number of comparisons in a trial	Proportion of trials with 100% of outcomes fully reported			Total
	0%	>1 and <100	100%	
1	6	0	39	45
2	0	2	15	17
3	0	3	10	13
4	1	5	1	7
5	0	2	0	2
6	2	1	2	5
8	0	0	1	1
Total	9	13	68	90

e. To explore trial characteristics in the protocol that may be associated with the selective reporting of clinical trials

Variables in the protocol that might be associated with selective reporting were investigated using binary logistic regression. There are two endpoints related to selective reporting:

1. Discrepancy in the identity of the primary outcome
2. Completeness of reporting

The covariates included in the models are listed in Table 28. All univariate analyses are detailed in Appendix 6: Univariate analyses (including analyses adjusted for number of outcomes or number of comparisons) and multivariate analyses in Appendix 7: Multivariate models.

Table 28: Variables in the protocol that might predict selective reporting

Covariate	Variable short name	Classification
Design	Not included in model as only 8/90 trials were crossover designs in the comparison data set (and 12/97 in trials with at least one identifiable primary)	0: crossover 1: parallel
Purpose	purpose	0: exploratory 1: pragmatic
Administration	administration	0: single centre 1: multi-centre national 2: multi-centre international
Commercial funding (protocol)	hrecdrugfund	0: no commercial funding 1: commercial funding available
Proposed sample size	hrecss200code	1: ≤ 200 2: > 200
Completeness of the sample size calculation	sscompcode2	0: nil 1: partial 2: complete
Allocation concealment	hrecalloc	0: not adequate 1: adequate
Sequence generation	hrecseqgen	0: not adequate 1: adequate
Blinding	hrecblind	0: not double blind 1: double blind
Number of comparisons	numbercomp	0: ≥ 2 comparisons 1: 1 comparison

Note: Variable short name is listed here to facilitate interpretation of the regression models in the appendices.

Results of logistic regression for the endpoint "Discrepancy in the identity of the primary outcome"

The analysis for the endpoint "Discrepancy in the identity of the primary outcome" was conducted using the 97 trials for which 1 to 4 primary outcomes were identifiable. Of these 97 trials, 64 (66%) declared all of their outcomes in both the protocol and the publication (Table 29).

Table 29: Proportion of outcomes declared in both protocol and publication by number of outcomes

	Proportion of outcomes declared in both the protocol and the publication			
Number of outcomes	0%	>0% but <100%	100%	Total
1	7	0	48	55
2	6	7	10	23
3 or more	1	12	6	19
Total	14	19	64	97

If it is assumed that in order for a trial to be without selective reporting that 100% of its outcomes must be declared in both the protocol and the publication, then trials with commercial funding, a sample size over 200, complete documentation of the sample size calculation and a single outcome are less likely to selectively report when investigated in a univariate analysis (See Table 30). When each variable is adjusted for the number of outcomes in the trial the same variables remain significant (Table 31). The significant variables were included in a multivariate analysis, the results of which indicate that trials with a partial (OR3.9, 95%CI 0.53-28.1) or completely (OR12.9, 95%CI 1.9-86.1) documented sample size calculation are more likely to report their outcomes in both the protocol and the publication than those without a sample size calculation (Table 32). The multivariate model also indicates that trials with more than one outcome (OR 0.04, 95%CI 0.01 - 0.17) are less likely to report their outcomes in both the protocol and the publication than those with a single outcome (Table 32).

Note that in all results tables a p value marked with an asterisk (*) denotes a global p value for that variable.

Table 30: Univariate analyses for the endpoint discrepancy in the identity of the primary outcome (threshold 100%)

Covariate	Categories	P value	Odds ratio	95% CI	
				Lower	upper
Design	0: crossover 1: parallel	.96	.97	.23	4.14
Purpose	0: exploratory 1: pragmatic	.66	1.26	.46	3.42
Administration	0: single centre 1: multi-centre national 2: multi-centre international	*.26	1.35	.39	4.72
			2.21	.85	5.73
Commercial funding	0: no commercial funding 1: commercial funding available	.03	2.72	1.13	6.50
Target sample size	1: ≤ 200 2: > 200	.09	2.12	.90	4.97
Completeness of sample size calculation	0: nil 1: partial 2: complete	*.007	7.22	1.44	36.22
			9.41	2.31	38.31
Allocation concealment	0: not adequate 1: adequate	.48	1.36	.59	3.19
Sequence generation	0: not adequate 1: adequate	.92	1.05	.42	2.60
Blinding	0: not double blind 1: double blind	.31	1.55	.67	3.62
Number of outcomes	1: 1 outcome 2: more than 1 outcome	<0.001	.08	.03	.22

Table 31: Variables adjusted for number of outcomes for the endpoint discrepancy in the identity of the primary outcome (threshold 100%)

Covariate	Categories	P value	Odds ratio	95% CI	
				Lower	upper
Design	0: crossover 1: parallel	.79	.79	.14	4.39
Purpose	0: exploratory 1: pragmatic	.78	1.18	.36	3.90
Administration	0: single centre 1: multi-centre national 2: multi-centre international	*.46	1.09	.25	4.85
			1.95	.63	6.05
Commercial funding	0: no commercial funding 1: commercial funding available	.02	3.50	1.18	10.39
Target sample size	1: ≤ 200 2: > 200	.04	3.11	1.06	9.12
Completeness of sample size calculation	0: nil 1: partial 2: complete	*.004	5.44	.82	36.22
			20.9	3.34	130.67
Allocation concealment	0: not adequate 1: adequate	.73	1.19	.44	3.28
Sequence generation	0: not adequate 1: adequate	.92	.95	.32	2.79
Blinding	0: not double blind 1: double blind	.15	2.13	.75	6.03

Table 32: Multivariate model for the endpoint discrepancy in the identity of the primary outcome (threshold 100%)

Covariate	Categories	P value	Odds ratio	95% CI	
				Lower	upper
Commercial funding	0: no commercial funding 1: commercial funding available	.091	2.77	.85	9.00
Completeness of sample size calculation	0: nil 1: partial 2: complete	*.027	3.86	.53	28.1
			12.89	1.93	86.08
Number of outcomes	1: 1 outcome 2: more than 1 outcome	<0.001	.04	.01	.17

Results of logistic regression for the endpoint "Completeness of reporting"

The analysis for the endpoint "Completeness of reporting" was conducted using the 90 trials for which 1 to 4 primary outcomes were identifiable that did not involve global outcomes or multiple comparisons.

If it is assumed that in order for a trial to be completely reported it must fully report 100% of its primary outcomes then the only variable significant in univariate analysis is the number of comparisons, with trials with single comparisons being more likely to fully report than those with 2 or more comparisons (Table 33 and Table 35). Repeating the univariate analyses adjusting for the number of comparisons, then the completeness of documentation of the sample size calculation also becomes significant with trials with a partial or completely documented sample size calculation being more likely to fully report than those without (Table 34 and Table 36). The final model therefore included the completeness of the sample size calculation and the number of comparisons (Table 37).

Trials with only one comparison were 4.5 times more likely to fully report all of their comparisons than trials with more than one comparison. Trials with a partially documented sample size calculation in the protocol were 7.6 times as likely as those without such a calculation, and trials with a complete sample size calculation were 4.8 times as likely as those without a calculation to fully report all of their comparisons.

If the threshold for fully reported is changed to 66% then the completeness of documentation of the sample size calculation is the only variable that remains significant (OR 8.5 for partially documented compared with nil, and OR 6.4 for completely documented compared with nil).

Table 33: Number of comparisons by proportion of trials with fully reported comparisons

Number of comparisons	Proportion of trials with fully reported comparisons			Total
	0% fully reported	>0% but <100%	100% fully reported	
1	6	0	39	45
2	0	2	15	17
3	0	3	10	13
4	1	5	1	7
5	0	2	0	2
6	2	1	2	5
8	0	0	1	1
Total	9	13	68	90

Table 34: Completeness of sample size calculation by proportion of trials with fully reported comparisons

Completeness of sample size calculation		Proportion of trials with fully reported comparisons		Total
		<100% fully reported	100% fully reported	
	nil	6	6	12
	Partial	3	16	19
	Complete	13	46	59
Total		22	68	90

Table 35: Univariate analysis for the endpoint completeness of reporting (threshold 100%)

Covariate	Categories	P value	Odds ratio	95% CI	
				Lower	upper
Purpose	0: exploratory 1: pragmatic	.95	1.04	.39	3.28
Administration	0: single centre 1: multi-centre national	.95	1.28	.28	5.98
	2: multi-centre international		1.08	.37	3.18
Commercial funding	0: no commercial funding 1: commercial funding available	.18	1.96	.74	5.19
Target sample size	1: ≤ 200 2: > 200	.91	1.06	.40	2.77
Completeness of sample size calculation	0: nil 1: partial	.09	5.33	1.00	28.44
	2: complete		3.54	.98	12.83
Allocation concealment	0: not adequate 1: adequate	.95	.36	2.49	
Sequence generation	0: not adequate 1: adequate	.46	.68	.25	1.88
Blinding	0: not double blind 1: double blind	.72	1.19	.46	3.13
Number of comparisons	0: ≥ 2 comparisons 1: 1 comparison	.02	3.59	1.25	10.29

Table 36: Variables adjusted for number of outcomes for the endpoint completeness of reporting (threshold 100%)

Covariate	Categories	P value	Odds ratio	95% CI	
				Lower	upper
Purpose	0: exploratory 1: pragmatic	.74	.82	.25	2.71
Administration	0: single centre 1: multi-centre national	.96	1.10	.22	5.47
	2: multi-centre international		1.18	.39	3.64
Commercial funding	0: no commercial funding 1: commercial funding available	.16	2.06	.75	5.69
Target sample size	1: ≤ 200 2: > 200	.73	1.19	.44	3.26
Completeness of sample size calculation	0: nil 1: partial	.05	7.65	1.26	46.47
	2: complete		4.80	1.16	19.81
Allocation concealment	0: not adequate 1: adequate	.64	1.27	.46	3.56
Sequence generation	0: not adequate 1: adequate	.56	.73	.25	2.1
Blinding	0: not double blind 1: double blind	.34	1.67	.59	4.71

Table 37: Multivariate model for the endpoint completeness of reporting (threshold 100%)

Covariate	Categories	P value	Odds ratio	95% CI	
				Lower	upper
Completeness of sample size calculation	0: nil 1: partial	.053	7.650	1.26	46.47
	2: complete		4.802	1.16	19.81
Number of comparisons	0: ≥ 2 comparisons 1: 1 comparison	.010	4.465	1.43	13.98

f. To explore other (non-protocol) factors that may be associated with the selective reporting of clinical trials

Trial-based investigation of statistical significance

In 26 trials, all of the comparisons were statistically significant, of which 24 fully reported all of their comparisons (Table 38). Trials where all of the comparisons are statistically significant are more likely to fully report all of their comparisons (p=0.06). Of the 45 trials with a single comparison, 20 (44%) were statistically significant (Table 39). All comparisons were statistically significant in 6 (13%) of the 45 trials with more than one comparison.

Table 38: Proportion statistically significant and proportion fully reported

Proportion of trials with a comparison p<=0.05	Proportion of trials with a comparison where the primary outcome was fully reported in the publication			Total
	0%	>0% but <100%	100%	
0%	8	6	34	48
>0% but <100%	0	6	10	16
100%	1	1	24	26
Total	9	13	68	90

Exact ordered categorical test: p=0.06

Table 39: Number of comparisons by proportion statistically significant

Number of comparisons	Proportion of trials with a comparison that was statistically significant			Total
	0%	>0% but <100%	100%	
1	25	0	20	45
2	11	2	4	17
3	6	6	1	13
4	2	5	0	7
5	0	1	1	2
6	3	2	0	5
8	1	0	0	1
Total	48	16	26	90

Comparison-based investigation of statistical significance

Of the 194 comparisons a p value was not reported for 32 (Table 40). The odds of fully reporting a comparison are greater if the result is statistically significant when compared to those that are not statistically significant (OR 2.2, 95% CI 0.9-5.3, p=0.08).

Table 40: Statistical significance by completeness of reporting of comparison

P value		Completeness of reporting				Total
		Not reported	Partially reported	Fully reported	qualitative	
	p>0.05	0	18	75	6	99
	P not reported	13	0	19	0	32
	P <= 0.05	0	5	55	3	63
Total		13	23	149	9	194

Multiple time points and global outcomes

There were 14 global outcomes and 19 involving multiple time points (that is, more than 2 time points), including 5 outcomes that met both criteria.

Global outcomes are those where the outcome consists of multiple components and could be reported with a single global measure with or without measures for each of the components. There were 14 trials with global outcomes, 5 of which came from one trial. (Note: this trial had a total of 5 primary global outcomes including 2 quality of life measures, both specified in the protocol as being the primary outcome measures. The publication reported 3 additional global outcomes specified as primary in the publication but not in the protocol.)

Thirteen of the 14 trials with global outcomes could have been considered to be trials with no identifiable primary outcome owing to the very broad "definition" of the outcome that was used in each case. The global outcomes identified included "biochemical response", "anthropometric measures", "compliance", "dissatisfaction", "growth" and "return to normal physiological functions". In all cases, a number of tests or measures were used with no clear single outcome and no overall summary measure anticipated. "Safety" (in 2 trials), when recorded as a primary outcome for which there is no specific definition, also comes under this category.

The primary outcome in 2 trials was quality of life, the most obvious example of a global outcome. Instruments used to measure quality of life usually consist of multiple dimensions, and each dimension consists of multiple questions. A single, global measure of quality of life may be reported, with results for each dimension, and results for each question. To further complicate matters, this outcome is often measured over multiple time points.

There were 19 primary outcomes where there were more than 2 time points involved. In all cases either the protocol or the publication described that the outcome was (to be) routinely measured at multiple, specified intervals during the course of trial, but did not specify a particular time point of interest. The most common way to present multiple time points was to present the results for each time point at which the outcome was measured, or to present the change in the value between one or more time points. One option is to assume that the primary time point is the final measure, although it is not usually evident in either the protocol or the publication that this is the case. The ideal situation would be an analysis that included a single, global test, or individual tests adjusted for multiple outcomes. Insufficient detail was reported in the protocols and most publications to be able to determine the degree to which multiple time points had been taken into account.

Secondary Objective 1: to explore issues relating to the sample size, its calculation and reporting

a. Do trial protocols contain adequate details of the target sample size?

All trials had a target sample size mentioned in the protocol, although in 15 cases there was no evidence that this number was based on a formal power calculation (Table 41).

Table 41: Sample size details in protocol

Characteristic	Classification	Number
Is there a target sample size?	yes - appropriate formula	88
	yes - but no calculation	15
The sample size is	≤200	49
	201-500	23
	>500	31
The completeness of the sample size calculation	Incomplete	16
	Partial	20
	Complete	67
The outcome used to calculate sample size	Yes	86
	No	17
The expected treatment effect	Yes	84
	No	19
Was the α error pre-specified	Yes	73
	No	30
The specified value of α	0.01	2
	0.025	1
	0.05	70
	Not reported	30
Was the power pre-specified	Yes	80
	No	23
The specified power	<0.8	2
	0.8	59
	0.85	4
	0.9	13
	0.95	2
	Not reported	23
Was the alternative hypothesis specified?	Yes	38
	No	65

b. What is the completeness of documentation of the sample size calculation in the protocol?

The completeness of reporting of the sample size calculation in the protocol was classified as complete (reported the outcome used, the estimated size of the effect of the intervention, the level of significance and power), partial (as a minimum reported the outcome used for the power calculation) or nil/none. Sixteen trials were classified as not having a sample size calculation in the protocol, and none of these provided any details of the elements of the sample size calculation except for two which documented some elements but not the outcome used. Of the 19 trials classified as partially reported all but 4 documented the expected effect size, 5 documented the significance level and 11 the power level.

c. What factors in the protocol are associated with the completeness of documentation of the sample size calculation in the protocol?

Variables that were considered to potentially be associated with the completeness of documentation of the sample size calculation in the protocol (nil v complete or partial) were the purpose of the trial, administration and the availability of commercial funding. Univariate analysis suggested that multi-centre international trials were more likely than single trials, and trials with commercial funding available at the time of submission to the REC, were significantly more likely to have a complete or partial sample size calculation in the protocol (See Appendix 6.3: Univariate analyses for completeness of documentation of the sample size calculation). When all of these variables were included in a multivariate model, multi-centre international trials remained significant with such trials being more likely to include a complete or partial sample size calculation in the protocol than single centre trials (OR 15, 95%CI 3-71, p=0.001) (See Appendix 7.3: Multivariate models for completeness of documentation of the sample size calculation).

Table 42: Administration by completeness of the sample size calculation

		Completeness of documentation of the sample size calculation		
Administration		Nil	Complete or partial	Total
	Single centre	8	19	27
	Multi-centre national	2	12	14
	Multi-centre international	2	47	49
Total		12	78	90

d. Is the completeness of documentation of the sample size calculation in the protocol related to the adequacy of reporting of the power calculation in the publication?

There were 67 trials with a complete sample size calculation in the protocol of which 50 had an adequately reported power calculation in the publication (see Table 43). The odds of a power calculation being adequately reported in the publication were almost 5 times greater if the sample size was completely documented in the protocol when compared to those that were not documented (Table 44).

Table 43: Completeness of the sample size calculation (a)

Adequacy of the reporting of the power calculation	Completeness of the documentation of the sample size calculation in the protocol			Total
	Nil	Partial	Complete	
Not adequate	10	7	17	34
adequate	6	13	50	69
Total	16	20	67	103

Table 44: Completeness of the sample size calculation (b)

Completeness of the documentation of the sample size calculation in the protocol	Odds ratio	P value	95.0% C.I.		Global p value
			Lower	Upper	
nil	1	.03			0.025
partial	3.1	.11	.8	12.14	
complete	4.9	.007	1.55	15.51	

e. What other trial-related factors are associated with the adequacy of reporting of the power calculation in the publication?

Variables in the protocol that might be associated with the adequacy of reporting of the power calculation in the publication include the purpose, administration, availability of commercial funding and the completeness of the sample size calculation in the protocol. Univariate analysis suggested that multi-centre international, pragmatic trials with a complete sample size calculation in the protocol were more likely to adequately report the power calculation in the publication (See Appendix 6.4: Univariate analyses for adequacy of reporting of the power calculation). When all of these variables were included in a multivariate model, multi-centre international trials were more likely than single centre trials (OR 15, 95%CI 3-71, p=0.001) to adequately report a power calculation in the publication.

The protocol was more likely to have a complete or partial sample size calculation in trials with sample sizes over 200 (p=0.005) (see Table 45). The publication was also more likely to adequately report a power calculation if the number of patients randomised was more than 200 (see Table 46).

Table 45: Sample size and completeness of reporting of power calculation in protocol

		Completeness of the sample size calculation in the protocol		
		Nil	Complete or partial	Total
Sample size in protocol	<=200	14	35	49
	>200	3	51	54
Total		17	86	103

Fisher's Exact test (2 sided) p=0.003

Table 46: Completeness of sample size calculation in protocol and adequacy of reporting of power calculation in publication

		Adequacy of reporting of power calculation in publication		
		Inadequate	adequate	Total
Completeness of the sample size calculation in the protocol	Nil	11	6	17
	Partial or complete	23	63	86
Total		34	69	103

Fisher's Exact test (2 sided) p=0.004

Table 47: Reported commercial funding and adequacy of reporting of power calculation in publication

Commercial funding reported in publication		Reporting of power calculation in publication		Total
		Inadequate	Adequate	
	No commercial funding available	12	27	39
	Commercial funding available	22	42	64
Total		34	69	103

Fisher's Exact test (2 sided) p=0.83

Table 48: Numbers randomised and adequacy of power calculation in publication

		Reporting of power calculation in publication		Total
		Inadequate	Adequate	
Number of participants randomised	≤ 200	24	21	45
	>200	10	48	58
Total		34	69	103

Fisher's Exact test (2 sided) p<0.001

Secondary Objective 2: to explore other relationships between the protocol and the publication

f. allocation concealment and sequence generation

g. use of blinding and placebo

h. adverse events

i. journal

j. exclusions

a. Allocation concealment and sequence generation

The methods of sequence generation and allocation concealment were inconsistently and poorly documented in both protocols and publications (Table 49). A code of "adequate" was assigned only when the method was both documented and the method described was sufficient to minimise bias. For example, a trial that reports that an envelope system was used but either does not describe how the system was implemented, or describes a system of implementation that could introduce bias, was deemed to have an inadequate method of allocation concealment.

16 trials had an adequate method of allocation concealment described in both the protocol and the publication (Table 49). An adequate method of sequence generation was described in 31 protocols and 38 publications, but only 16 described an adequate method in both documents.

Table 49: Allocation concealment and sequence generation

In protocol	In publication					
	Allocation Concealment			Sequence Generation		
	Adequate	Inadequate	Total	Adequate	Inadequate	Total
Adequate	16	28	44	16	15	31
Inadequate	0	59	59	22	50	72
Total	16	87	103	38	65	103

The method of allocation concealment was described in 36 trial publications (Table 50). Of these, 9 reported that sealed envelopes were used (in only 1 case was it possible to ascertain that the method used was adequate), 15 utilised a central telephone service and 9 reported that randomisation was centralised but the method used was unclear.

The method of sequence generation was described in 39 publications. Of these, 38 reported a method that was considered to be adequate (5 used minimisation, 7 permuted blocks, 6 random number tables, 4 dynamic balancing and 16 stated that a computer was used to generate the sequence).

Table 50: Method of allocation concealment and sequence generation (publication)

Method of sequence generation	Adequacy of method reported			Method of allocation concealment	Adequacy of method reported		
	Adequate	Not adequate	Total		Adequate	Not adequate	Total
computer	16	0	16	centralised method unclear	0	9	9
dynamic balancing	4	0	4	central telephone	15	0	15
minimisation	5	0	5	sealed envelopes	1	0	1
permuted blocks	7	0	7	sealed envelopes method unclear	0	8	8
random number table	6	0	6	other	0	3	3
other	0	1	1	not reported	0	67	67
not reported	0	64	64				
Total	38	65	103	Total	16	87	103

b. Use of blinding and placebo

There were 49 trial protocols that mentioned the use of a placebo in the protocol (48 of the 55 double blind trials and 1 single blind trial used a placebo) of which 31 provided a description of the placebo. In only 17 cases was the description of the placebo such that it was possible to ascertain that the placebo was adequate.

Two of the 55 trials declared as double blind in the protocol were not reported as double blind in the publication, and 2 of the 48 trials not declared as double blind in the protocol were reported as double blind in the publication. Three trials that declared that either the health care practitioner or the patient would be blinded in the protocol did not mention this blinding in the publication.

Table 51: Blinding in protocol and publication

Blinding mentioned in protocol		Blinding mentioned in publication		Total
		Not double blind	Double blind	
	Not double blind	46	2	48
	Double blind	2	53	55
Total		48	55	103

c. Adverse events

In 82 protocols it was stated that there would be some form of monitoring for adverse events, and 77 reported adverse events in some form in the publication (Table 52).

Table 52: Adverse events in protocol and publication

		Adverse events reported (in publication)		Total
		No	Yes	
Adverse events to be monitored (in protocol)	No	15	6	21
	Yes	11	71	82
Total		26	77	103

d. Journal type

Univariate analyses suggest that pragmatic, multi-centre national trials and international trials, sample sizes over 200, with a complete sample size calculation in the protocol and adequate documentation of allocation concealment in the protocol are all more likely to be published in general than specialty journals (see Appendix 6.5: Univariate analyses for journal type).

The multivariate model reveals that a projected sample size of more than 200 (OR 11, 95% CI 2-51, p=0.003) and adequate allocation concealment (OR 3.8, 95% CI 1.2-12, p=0.021) remain significant (see Appendix 7.5: Multivariate model for journal type).

e. Exclusions

There was an explicit statement that intention-to-treat analyses were to be conducted in 45 trials (See Table 53). Of the 40 trials reporting that there were exclusions, 25 (63%) reported the number excluded by treatment arm and 6 reported that the same criteria were applied to each treatment arm (See Table 54).

Table 53: Exclusions from analysis

Characteristics	Classification	Number
Did the publication explicitly state that intention to treat analyses were conducted?	Yes No	45 58
Did the publication report that patients were excluded from the analysis (of the primary outcome)?	Explicitly reported no exclusions Gave impression no exclusions Explicitly reported that there were exclusions Exclusions not mentioned or unclear	19 30 41 7+6
If the publication explicitly reported that there were exclusions, were the number excluded by treatment arm reported?	No Yes Not applicable	16 25 62
If the publication explicitly reported that there were exclusions, was it reported that the same criteria were applied to each group?	No Yes Not applicable	35 6 62
Reasons for exclusions	did not start allocated intervention false inclusions (ineligible) non-compliers other not applicable	12 6 6 19 60

Table 54: Intention to treat and exclusions

Participants excluded from analysis (of primary outcome/s)	Intention-to-treat analyses conducted		Total
	Yes	No	
Explicitly reported no exclusions	17	2	19
Gave impression no exclusions	11	19	30
Explicitly reported that there were exclusions	14	27	41
Exclusions not mentioned or unclear	2	5	7
Explicitly reported no exclusions	1	5	6
Total	45	58	103

The protocol-specific variables purpose, administration, commercial funding, allocation concealment, sequence generation, sample size and completeness of documentation of the sample size calculation were investigated both in univariate analyses and in the multivariate model (see Appendix 6.6: Univariate analyses for exclusions). There was no evidence of a relationship between any of the variables and exclusions from the trial analysis.

Secondary Objective 3: to explore the impact of the availability of commercial funding on the trial protocol and publication

Of the 65 trials indicating that commercial funding was available at the time of submission to the REC, 56 acknowledged that funding in the publication (Table 55). For 30 trials there was no indication that commercial funding was available in either the REC submission or the publication. Eight of the 38 trials indicating that commercial funding was not available at the time of submission to the REC, acknowledged a commercial funding source in the publication. Of the commercially funded trials 63% were multi-centre and international (Table 59), and 72% had a complete sample size calculation in the protocol (Table 60).

Table 55: Declaration of commercial funding

	Commercial funding declared in publication		
Commercial funding declared in protocol	No	Yes	Total
No	30	8	38
Yes	9	56	65
Total	39	64	103

In univariate analysis, trials with commercial funding available at the time of submission to the REC were significantly more likely to:

- Have at least one clearly identifiable primary outcome in the protocol (OR 4.7, 95%CI 1.1-19.3, p=0.033) (Table 56)
- State that adverse events would be monitored in the protocol (OR 19, 95%CI 5-70, p<0.001) (Table 57)
- Mention adverse events in the publication (OR 6, 95%CI 2-13, p=0.001) (Table 58)

The relationship between the trial characteristics described in Table 28 and commercial funding available at the time of submission to the REC were explored univariately. This suggested that double-blind, multicentre trials with a complete or partial sample size calculation and adequate description of sequence generation were more likely to be commercially funded. The variables that were significant in univariate analysis were included in a multivariate model (see Appendix 7.6: Multivariate model for commercial funding in protocol). Commercially funded trials

were significantly more likely to be double blind (OR 2.6, 95%CI 1.1-6.4, p=.04) and multi-centre national (OR 2.6, 95%CI .7-9) or international trials (OR 4, 95%CI 1.5-11, global p=.02).

Table 56: Commercial funding by primary outcome in protocol

		Primary outcome identifiable in protocol		Total
		No	Yes	
Commercial funding available	No	7	31	38
	Yes	3	62	65
Total		10	93	103

Table 57: Commercial funding by adverse events monitored

		Adverse events to be monitored (in protocol)		Total
		No	Yes	
Commercial funding available	No	18	20	38
	Yes	3	62	65
Total		21	82	103

Table 58: Commercial funding by adverse events reported

		Adverse events reported (in publication)		Total
		0	1	
Commercial funding available	No	17	21	38
	Yes	9	56	65
Total		26	77	103

Table 59: Commercial funding by administration

		Administration			Total
		Single centre	Multi-centre national	Multi-centre international	
Commercial funding available	No	20	6	12	38
	Yes	13	11	41	65
Total		33	17	53	103

Table 60: Commercial funding by sample size calculation

		Completeness of sample size calculation			Total
		Incomplete	Partial	Complete	
Commercial funding available	No	11	7	20	38
	Yes	5	13	47	65
Total		16	20	67	103

Chapter 5: Discussion and conclusions of follow-up study

In Chapters 2, 3 and 4 the background, methods and results of the follow-up study were described. This chapter will attempt to put these results into context through a discussion of:

- Summary of key findings
- Discussion
    - Issues pertaining to the selective reporting of primary outcomes
    - Issues pertaining to sample size
    - Other issues
- Limitations of the project
- Implications
- Conclusions

Summary of key findings

This study set out to examine issues in the selective reporting of the primary outcomes of randomised controlled trials through the conduct of a follow-up study of trials submitted to a research ethics committee. It was demonstrated that selective reporting exists in some form in a significant proportion of trial publications. In relation to the specific primary objectives of the study it was found that:

- Most trials had at least 1 identifiable primary outcome (97 of 103).
- Many trials declared more than one primary outcome in the protocol (59 of 103).
- There was evidence of the selective reporting of the identity of the primary outcome with 15% of outcomes declared in the publication as being the primary outcome not being declared as such in the protocol, and almost half of these were not mentioned as outcomes in the protocol at all. Similarly, 17% of outcomes declared as the primary outcome in the protocol were not declared as being the primary outcome in the publication.
- Half of the trials included in the analysis of completeness of reporting (45 of 90) reported more than one comparison based on a primary outcome.

- Most trials (68 of 90) completely reported all of their comparisons. Trials were more likely to completely report all of their comparisons if there was a single primary comparison and a completely documented sample size calculation in the protocol.
- Although all of the outcomes that were reported as the primary outcomes in the publication were the same as all of the primary outcomes declared in the protocol in most trials (64 of 97), there were often discrepancies between the trial protocol and the trial publication in relation to the identity of the primary outcome (or outcomes). The identity of the primary outcomes was more likely to be the same in the protocol and the publication in trials with a complete sample size calculation in the protocol (and possibly those with commercial funding), and was less likely to be the same if the trial had more than one outcome.
- The documented definitions of primary outcomes were insufficient to allow readers to determine if there were discrepancies between the trial protocol and the trial publication in relation to those definitions.
- Trials where all comparisons were statistically significant were more likely to fully report all of their comparisons.
- Trials with a completely or partially documented sample size calculation in the protocol consistently appeared in the multivariate regression models as being significantly less likely to selectively report the primary outcome.
- Multi-centre, international trials were more likely to have a complete or partial sample size calculation in the protocol, and were also more likely to adequately report a power calculation in the publication.
- Randomisation details (allocation concealment and sequence generation) were poorly documented in most trial protocols.
- Trials with commercial funding available at the time the protocol was submitted to the REC were more likely to be double-blind, multi-centre national or international trials.

Discussion

Issues pertaining to the selective reporting of primary outcomes

As described in Chapter 1, a good solution to multiple testing in most studies is to focus on a single primary outcome and, to ensure that the choice of outcome is not data derived, that outcome should be clearly specified before the trial commences. It is not inappropriate, however, for a trial to have more than one primary outcome (or comparison) but trials intending to address more than one outcome need to take the multiplicity into consideration through the use of appropriate statistical methods. The two main methods available are to adjust the nominal significance level or use a multivariate technique. It was not possible to investigate whether appropriate adjustment was made for multiple comparisons in the publications of trials included in this study as the feasibility study indicated that this information was poorly reported in trial publications.

In around 25% of protocols it was necessary to infer that the outcome was a primary outcome based on the sample size calculation or objectives as there was no clear statement of the primary outcome in the protocol. Although it may be reasonable to make such inferences, it is not uncommon for a trial to be powered to address an intermediate outcome rather than the outcome of primary clinical interest. It is therefore possible that at least some outcomes were misclassified as being primary outcomes. Pragmatic trials were more likely to clearly state the primary outcome in both the protocol and the publication, and it was more common to have to infer the primary outcomes in explanatory trials. This is not unexpected given the more investigative nature of the latter. Although the purpose of the trial did not appear as a significant factor in any of the multivariate models, it would be wise to treat primary outcome information inferred from the publications of explanatory trials with a degree of caution.

While it is not possible to ascertain the motive for elevating an outcome to, or demoting an outcome from being the primary, in many cases it is likely to be a post hoc, data-derived decision. There was evidence that selective reporting of the identity of the primary was less likely if the sample size was larger, there was a complete sample size calculation in the protocol and the trial had a single comparison. If these 3 factors can be

thought to be surrogate measure for the quality of the trial protocol, then it would suggest that trials with better quality protocols are less likely to be guilty of this form of selective reporting when it comes to the publication of the trial's results.

A concept worthy of further thought is a method by which the degree of severity of selective reporting of outcomes could be classified, some forms of selective reporting potentially having greater importance and impact than others. An outcome declared as the primary in the protocol, which is not reported in the publication, implies a greater degree of selection, for example, than a change in the process for adjudicating the outcome. A possible framework for a severity scale is described in Table 61. This represents my interpretation of a possible framework and more work would be required to obtain a range of opinions regarding what is more and less severe selective reporting if this was to be developed into an assessment tool with wider applicability.

Table 61: Classification of severity of selective reporting

Severity of selective reporting of outcome	Level	Details
Severe	1	• Primary outcome in protocol not reported in publication • Primary outcome in publication not mentioned in protocol
	2	• Primary outcome in protocol reported in publication but not as primary outcome • Primary outcome in publication mentioned in protocol but not as primary outcome • Primary outcome is a composite outcome that has been changed between the protocol and the publication through addition or removal of a component
Moderate	3	• A change in the measurement criteria. This could include: o A change in the instrument used o A change in the definition of the outcome. eg a new (more accurate) test is introduced; eg new criteria for defining an event are introduced
	4	• A change in the primary time point (given that an outcome is measured over multiple time points)
Mild	5	• Change in statistical analysis (eg analysis unadjusted versus adjusted for other variables; a change in the statistical test used)
	6	• Change in processes for adjudication (eg committee, data review, pathology review, etc)

The manner in which the outcome is to be assessed should be clearly defined to ensure consistency across patients. The choice of outcome measure, particularly if it involves a specific test or other investigation, should be documented in sufficient detail for people

reading the protocol to be able to apply the same measure in the same way across individuals and sites, and would ideally be supported by evidence of the measure's reliability, validity, repeatability, sensitivity and specificity. All of these details should be recorded in the trial protocol. More than 30% of the primary outcomes included in this study were poorly defined in either the protocol or the publication to the extent that it was not possible to determine if the outcomes were in fact the same.

In some cases, trial investigators may assume that it is "understood" that a standard definition applies to a particular term, and that explicit definition in the protocol is not required. An example is time to event outcomes where the point at which the time period of interest is judged to start and end should be clearly defined. It is often assumed that time starts at the date of randomisation, but this is not always a valid assumption, with some trials starting the clock at the time the participant commences treatment, rather than when randomised. In addition, if this time point ends at the date of a specific event then what defines an event, and on what basis do we measure whether and when an event has occurred?

Alternatively, the instrument or test to be used to measure the outcome may be documented in detail, with a number of investigations to occur over a specified time period (eg specific test performed at baseline and once a month for 12 months). Problems arise when the way in which the data will be interpreted is not declared. In the example of monthly measurements for 12 months, the trial investigators could intend to report the results for all time points, for selected time points, for a single time point (eg the last time point). In addition, the change in a value between any 2 time points may be reported, or a pooled measure such as an area-under-the-curve may be used. This opens the way for trial investigators to select the time point or analysis method that places the results of their trial in the best light.

There was a relationship between the number of comparisons and full reporting, with the proportion of trials fully reporting all of their comparisons decreasing as the number of comparisons increased (87% of those with a single comparison compared to 27% of those with 4 or more comparisons). This finding is not surprising given that trials with a single comparison have fewer options to choose from, and that the investigators of trials

with more than one comparison may place higher priority on one comparison over another. They may or may not have documented this priority in the trial protocol.

One reason for completely reporting one comparison but not another is statistical significance. Although the current study focused on the reporting of primary outcomes rather than all outcomes, it was still able to demonstrate that trials in which all of the comparisons were statistically significant were more likely to fully report all of those comparisons, confirming the results of the landmark studies completed by Chan *et al* (Chan AW & Altman 2003;Chan, Hrobjartsson A, Haahr, Gotzsche, & Altman 2004a). who demonstrated an association between completeness of reporting and statistical significance. In their study, the odds of an efficacy outcome being fully reported was more than doubled if that outcome was statistically significant (OR 2.4, 95%CI 1.4 – 4.0). The odds of a harm (or safety) outcome being reported if it was statistically significant was even greater (OR 4.7, 95%CI 1.8 – 12.0). This is consistent with the findings of my cohort which also found that the odds of fully reporting a comparison were greater if the result was statistically significant (OR 2.2, 95% CI 0.9-5.3, p=0.08).

While the current study and that performed by Chan *et al* are similar, there are key differences. The former focuses on the primary outcome and, for the analysis of completeness of reporting, breaks each trial down to the level of the comparison rather than the outcome. This was necessary in order to take into account the substantial number of trials with more than 2 arms, and more than 1 outcome. The previous study did not find it necessary to break the data down into comparisons, and analysed it by creating 2x2 tables for each trial (completeness of reporting x statistical significance), calculating an odds ratio for each trial, then pooling the data in a meta-analysis using a random effects model. Trials with empty rows or columns were not included, resulting in the exclusion of 49 trials from the analysis of efficacy outcomes, and 54 from the analysis of harm outcomes. The current study analysed the data based on the frequency of significant comparisons in each trial allowing me to explore the impact of changing the threshold for what could be considered a "positive" trial, and allowed all trials with eligible outcomes to be included in the analysis.

Issues pertaining to sample size

A fixed sample size calculated before the trial begins enables the investigator to assess the feasibility of the trial and to make adjustments where appropriate. The number of participants required to detect a useful treatment difference may, for example, necessitate a multi-centre rather than single centre trial, or amendment of the eligibility criteria for the trial. In theory, the sample size and the components used in its calculation are therefore important markers of a trial's quality and its chances of successful accrual. In this study, trials with a completely or partially documented sample size calculation in the protocol consistently appeared in the multivariate regression models as being significantly less likely to selectively report the primary outcome when compared to trials with no sample size calculation documented. This variable would therefore appear to be a proxy measure for the quality of the trial based on the protocol.

Other issues

It is evident from the trials included in this study that important aspects of the design and conduct of clinical trials (such as the method of randomization and details of how blinding of the intervention will be achieved) are not always documented in clinical trial protocols, and it would be unwise to rely solely on these documents (as submitted to an ethics committee) to determine the quality and validity of a clinical trial. The best quality trial protocols appeared to be multi-centre, international trials with a commercial funding source – perhaps not surprising given the significant resources required to adequately support the conduct of high quality clinical trials research. Given the anticipated financial outlay, commercial sponsors are likely to want to invest in the conduct of a sound study and hence adequately resource the design and development phase through the employment of appropriately qualified statisticians and other methodologists.

Limitations

At the time the trials included in this follow-up study were submitted to the ethics committee the standards expected of trial protocols were somewhat different to those expected now, a decade later. While researchers were becoming more aware of the problem of publication bias, the movement behind the prospective registration of trials was in its infancy. Neither the CONSORT statement (CONSORT Group 1996) nor the ICH guidelines (International Conference on Harmonisation 1996) would emerge until 1996. The main reasons for compiling a protocol at the time that the

protocols included in the follow-up study were prepared were likely to have been to impress potential funding agencies, obtain ethics approval, to attract site investigators to participate in multi-centre studies, and to assist in the conduct of the study at each site by providing some basic rules. In the case of some of the trials included in this study the "protocol" was simply the grant application subsequently submitted to a potential funding agency such as the NHMRC. The notion that the protocol could be used as a mechanism for addressing publication bias and other issues in the conduct of clinical trials would not have been widely recognised or discussed at the time. While the actual protocol document is still not required today in order to register a trial in a prospective trials register, there are key items required that can only be obtained from documents. These items include the health condition being studied, the details of the intervention, the key inclusion and exclusion criteria, the target sample size, the primary outcome and secondary outcomes (World Health Organization 2006).

Given the era in which the trials included in this cohort were designed and conducted, it could be said that the issues identified in this protocol are of historical interest and not relevant in the current context, where the guidelines provided by documents such as the guidelines for good clinical research practice (GCRP) produced by the International Conference on Harmonization, and the CONSORT statement, both mentioned previously. While the CONSORT statement may have impacted on the quality of reporting in the publication arising from a clinical trial, it is unlikely to have had an impact on the quality of the trial protocol, given that the issues identified by this cohort study were disparities between the protocol and the publication.

In relation to GCRP it is important to note that they are guidelines and hence primarily voluntary unless an organization such as a regulatory agency mandates compliance. This has happened recently in the case of the European Directive on clinical trials (European Commission. 2001). The degree to which clinical trial protocols comply with the various guidelines produced by ICH (including ICH E9: Statistical principles for clinical trials (International Conference on Harmonisation 1998)) is unknown, as is whether or not compliance with such guidelines leads to improvements in the quality of clinical trials research. In addition, compliance has generally been a requirement of the private sector, and adoption by the public sector has not been widespread, or even possible, given the significant resources required for

their implementation. Indeed, when the European Directive was introduced and made applicable to all clinical trials research, concern was expressed by prominent international researchers that it would "effectively end all clinical research except for those trials which are commercially-inspired, and drug-company sponsored" (www.saveeuropeanresearch.org; accessed 23rd March 2006). Hemminki and Kellokumpu-Lentinen reported in 2006 that the Directive appears to have had a negative impact on the amount of investigator-initiated research conducted (Hemminki & Kellokumpu-Lehtinen 2006).

Any retrospective investigation of clinical trial protocols should take the changes in expectations over time into account. It should also be kept in mind that the best we can hope to achieve by looking back is an assessment of the quality of documentation, not the quality of the clinical trials the documentation is associated with. Absence of information on allocation concealment and sequence generation in the trial protocol, for example, does not signify that the method of randomisation actually implemented can be judged to be inadequate. As demonstrated by Soares *et al* for example, in a study that compared 56 protocols for trials conducted by a US-based collaborative oncology group with their resulting publications, the reporting of methods in the publications does not necessarily reflect the methodological quality of the associated protocols (Soares et al. 2004).

Owing to the complexity of outcomes with multiple time points or multiple dimensions (global outcomes) these were excluded from the analysis of completeness of reporting. It is possible that these complex outcomes may be more prone to incomplete reporting for the very reason that they are complex, and that the size of the problem of complete reporting has been underestimated in this study.

An important issue not addressed in this study is the impact of the lack of an analysis plan in clinical trial protocols (including lack of *a priori* declaration of sub-groups of interest). An original intention was to examine sub-groups, however none of the trials examined in the feasibility study had an analysis plan in the protocol and hence this project was subsequently abandoned. While it is expected that an analysis plan will be developed and documented before the analysis is conducted, the need to include an analysis plan in the trial protocol submitted to an ethics committee is unclear. While it

could be argued that including an analysis plan in the protocol would improve the quality of clinical trials and the quality of reporting, it is not clear whether detailed plans are available at the beginning of most trials. The lack of an analysis plan could make it easier for a statistical test to be selected based on the results obtained from multiple tests. Readers currently rely on the authors of manuscripts to declare the nature of their analysis plan (including whether subgroups investigated were declared *a priori*). Regulatory agencies such as the US Food and Drug Administration (FDA) may suggest that the analysis plan be included as part of the documentation of the quality of evidence supporting an effectiveness claim. (U.S Food and Drug Administration 1998) Even then, the lodgement of the analysis plan would appear to be a suggestion rather than a requirement.

Implications

There is a need for all those associated with the conduct of clinical trials, and the evaluation of clinical trial protocols, to pay more attention to the documentation of key issues in the protocol. This includes the documentation of the sample size calculation and the rationale for the choices made for each of the components, and the methods used to achieve randomization and blinding of the interventions (if used). In particular, there needs to be an improvement in the way in which trial investigators declare which outcomes are the primary outcomes, as well as how those outcomes are to be measured.

Implications for research ethics committees

Ethics committees have a unique opportunity to improve the quality of the trials they approve, particularly single-centre trials, before the trial starts recruiting participants. A relatively simple measure for them to take may be to encourage trial investigators to improve the way in which they calculate and document their sample size, as the adequacy of documentation of the sample size in this follow-up study was associated with an increase in the odds of selective reporting. It is possible that adequate documentation of the sample size calculation may simply reflect the resources that were available to the trial investigators at the time the trial was designed, including appropriate statistical support. Whatever the mechanism is behind this apparent effect it would seem logical that trials would benefit from the early active involvement of a statistician or other methodologist.

An important initiative implemented by the REC whose trials were included in this study in the late 1990's, was a "clinic" held once a week to provide statistical and methodological support to individuals in the institution embarking on a clinical research project. This service was abandoned in 2005 due to lack of funds which is unfortunate for many reasons, particularly the lost potential to strengthen the methodological quality of trials that do not have a commercial sponsor. Appropriate statistical and methodological support could make these smaller, single-centre studies less vulnerable to selective analysis and reporting. Chapter 6 and Chapter 7 will discuss the role and responsibility of ethics committees to ensure the conduct of high quality research, and possible mechanisms to ensure they are able to do so.

A role for prospective trial registers

There may be a role for the slowly increasing number of prospective clinical trial registers to act as repositories for the registration of protocols, protocol amendments, analysis plans and other trial-related information (such as minutes of safety and data monitoring committees) about individual clinical trials not required or available at the beginning of the trial by funding agencies or ethics committees. Despite the best of intentions, however, even prospective lodgement of an analysis plan would essentially be an honour system, as the only people who will really know if a trial was conducted according to protocol, or if subgroups and other aspects of the analysis were truly declared *a priori*, are the people in control of the trial's data.

Implications for trial investigators and systematic reviewers

The relationship between statistical significance and completeness of reporting is a logical one given the desire of trial investigators to place their trial in a positive light so that journals will publish the results, journal subscribers will read the results, and health care practitioners and policy makers will incorporate the results into practice. However, the fact that we might be getting the full picture for the "positive" news, but only an incomplete picture for the "negative" news, is a matter of considerable concern.

The users of the published results of randomised trials (including systematic reviewers) should be acutely aware that they may only be seeing part of the story in the trial publication, and hence of the bias that may ensue as a result. Before relying on the results of one or more trials, they should be cognisant of the outcomes they

consider to be of primary clinical relevance, and should question the absence of those outcomes in the publications of trials addressing the question of interest.

The trials included in this follow-up study were a subset of all the randomised trials submitted to the REC over a 5 year period. The subset was selected based on the ability of the trial investigators to publish the results in a peer-reviewed journal rather than the full cohort of commenced studies. The process of peer-review may itself lead to improvements in the quality of reporting in a manuscript and by the time the manuscript has been published major discrepancies may have been queried and dealt with. It is likely that the same cannot be said for conference abstracts and other interim or unpublished reports from randomised trials which may therefore be even more prone to selective reporting.

Future research

There are many questions unanswered by this follow-up study that would be worthy of further investigation. The first would be to examine the impact initiatives such as Good Clinical Practice (introduced in 1996) or the European Directive (introduced in 2001) have had on the quality of randomised clinical trial protocols. (International Conference on Harmonisation 1998) (European Commission. 2001) Although GCP primarily targets commercial industry trials, it would be interesting to see if or how it has impacted on the design and conduct of non-commercially funded studies. As GCP recommends that protocols contain statistical analysis plans it would also be worth investigating whether it has had an influence on the reporting of this information in protocols considered by ethics committees since 1996.

The aim of those involved in the conduct of systematic reviews and meta-analysis is to identify and include information from all of the available, high quality evidence (ideally randomised controlled trials). This includes information that may only be available in the form of conference abstracts and other grey literature. In a study of 494 abstracts published in the proceedings of the American Society of Clinical Oncology, for example, it was found that key methodological information was often missing from trial abstracts, including information on allocation concealment, blinding, eligibility criteria and the description of the interventions. (Hopewell &

Clarke 2005) Further, a systematic review of studies that had evaluated the effect of the inclusion and exclusion of grey literature on the results of meta-analyses of randomised trials concluded that the inclusion of data from conference abstracts can be problematic as the relevant data may often be incomplete. (Hopewell et al. 2002) At the same time, others have suggested that excluding grey literature can "lead to exaggerated estimates of intervention effectiveness". (McAuley et al. 2001)

A study comparing the protocol with the conference abstracts would be a worthwhile undertaking. This would enable the size of the problem of selective reporting in conference abstracts to be quantified, particularly in relation to the selection of outcomes for reporting in the abstract. Such a study has not been undertaken to date, possibly due to problems in obtaining access to trial protocols and the challenge of identifying corresponding abstracts. (Chan AW et al. 2006)

Statistical methods to aid in the assessment of and adjust for selective reporting of outcomes in meta-analyses need to be developed evaluated. Methods suggested as being useful in assessing publication bias include the graphical (such as funnel plots) and others such as the imputation of missing data, the trim-an-fill method, regression methods, etc. (Rothstein, Sutton, & Borenstein 2005;Williamson & Gamble 2005)

It seems logical that providing methodological support to trial investigators at the time the trial is being designed would lead to improvements in that design. It is likely, though, that those who are most in need often do not have access to the resources necessary to pay for that expertise – specifically non-commercial, single-centre trials. If resources could be found there is potential to both improve the methodological quality of research, and to save the time and efforts of ethics committees, funding agencies, participants who consent to take part and others by ensuring that the standard of research is high. Such support comes at a financial cost and health authorities and others responsible for providing support to researchers may be unwilling to make the necessary investment. An investigation of the value of such support would therefore be worth pursuing.

Conclusion

Selective reporting of the results of clinical trials leads to the dissemination of incomplete and hence misleading information to those who rely on the published results of clinical trials. Trial investigators have a scientific and a moral responsibility to ensure that a true and accurate picture of their trial is presented. Initiatives such as the Public Library of Science's online journal "*Clinical Trials*", which promises that "publication decisions will not be affected by the direction of results, size or perceived importance of the trial", will hopefully start to overcome the difficulties some investigators have publishing negative or inconclusive research (http://clinicaltrials.plosjournals.org/index.html). Internet publishing also has the potential to overcome the problems caused by restrictive word limits on published manuscripts, with the ability to supplement publications with additional data online.

The recent acceptance by many health sectors of the need to prospectively register clinical trials has been encouraging. (Berlin & Ghersi 2005) Registration early in the history of a clinical trial could provide those of us interested in trial quality with a unique opportunity to address some of the issues arising from poor documentation (including inadequately developed trial protocols) and thus possibly prevent (or at least minimise) problems caused by the selective reporting of trial outcomes and other trial details.

Chapter 6: Impact of shared scientific or ethical review of multi-centre clinical research on the quality and process of clinical research (A systematic review)

This chapter reports all aspects of the design, conduct and reporting of a systematic review aiming to evaluate the impact of central scientific and/or ethical review of multi-centre clinical trial protocols on the clinical research process. The details will be presented as follows:

- Background to the systematic review
- Methods, including:
    - Eligibility criteria for studies, participants, interventions and outcome measures
    - Objectives
    - Search strategy to identify studies
- Results
- Discussion

Background to systematic review

Modern researchers accept that in order to comply with the requirements of the Declaration of Helsinki, which in turn informs the various guidelines for the conduct of ethical research in place around the world, research involving human subjects needs to be evaluated and approved by an appropriately constituted research ethics committee (REC: See Glossary on page xiv). (World Medical Association 2000) The most recent ratification of the Declaration of Helsinki specifically states that:

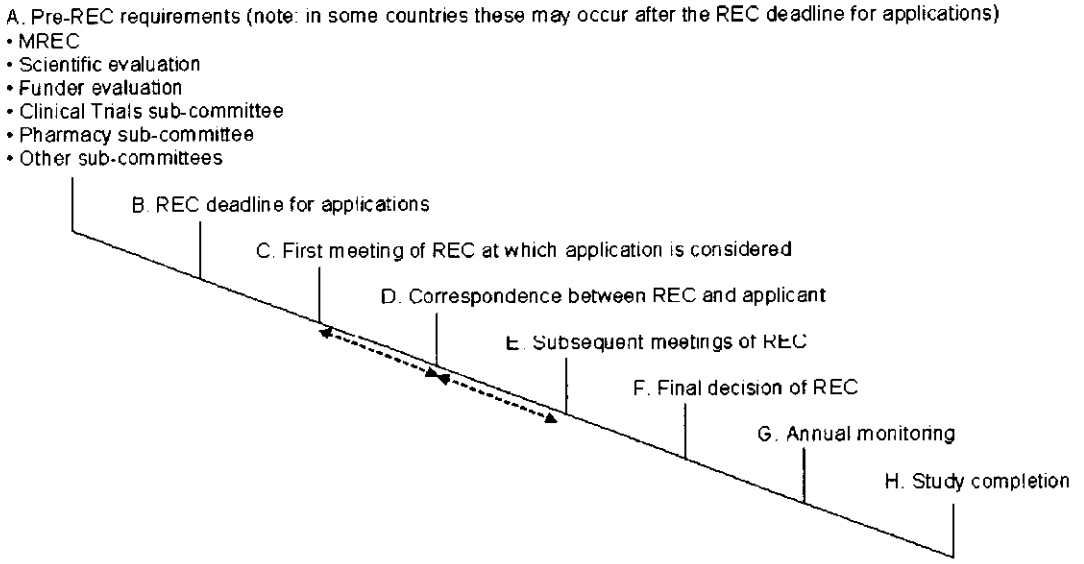
> *"The design and performance of each experimental procedure involving human subjects should be clearly formulated in an experimental protocol. This protocol should be submitted for consideration, comment, guidance, and where appropriate, approval to a specially appointed ethical review committee, which must be independent of the investigator, the sponsor or any other kind of undue influence. This independent committee should be in conformity with the laws and regulations of the country in which the research experiment is performed."(World Medical Association 2000)*

As described in Chapter 1, the volume of clinical trials research being conducted has increased considerably over the last decade. As most advances in clinical medicine are relatively modest, evaluating the effectiveness of new treatments usually requires large-scale clinical trials, ideally involving thousands of patients, often across multiple centres in multiple countries. To improve their efficiency, the design of these trials can be complicated and involve a baffling array of methodological and other scientific challenges.

As the volume and complexity of clinical research increases it is inevitable that the workload of research ethics committees (RECs) will also increase. In a Review of Institutional Review Boards (IRBs) in the USA in 1998 it was reported that the volume of work of most IRBs had increased by 42% since 1974-5 when the average number of proposals reviewed by an IRB was 43 per annum. In 1998 some IRBs reviewed as many as 2000 proposals. (US Dept Health and Human Services 1998) A study of a single hospital-based REC in the UK reported a steady increase in workload over ten years from 66 protocols per year to 302 ten years later. (Cookson 1992)

The work of RECs is not restricted to reviewing proposals as they also monitor studies they have previously approved, review amendments to research protocols, review adverse events and perform other tasks (see Figure 6: Ethics approval process). Some commentators believe that RECs are in crisis and that their effectiveness is in jeopardy. (US Dept Health and Human Services 1998) The USA IRB Review concluded that the major challenges facing RECs are the changing research environment (the increase in the number of research proposals, multi-site research, commercialised research and patient consumerism) and the fact that they are "reviewing too much, too quickly, with too little expertise". (US Dept Health and Human Services 1998)

Figure 6: Ethics approval process



When Savulescu and colleagues posed the question "are research ethics committees behaving ethically?" it provoked an appropriately varied range of responses: from researchers who could relate to the frustrations of the authors to REC members overwhelmed by the expectations placed on a group who are essentially "conscientious, sincere and disinterested" amateurs. (Pierce 1997;Savulescu, Chalmers, & Blunt 1996;Stone & Blogg 1997) Membership of ethics committees is expected to include members of the lay community and other non-scientists, yet these individuals are required to have significant scientific knowledge in order to meet their ethical obligations, scientific inadequacies having recognisable ethical implications. (NHMRC 2001)It could be argued, for example, that it is unethical to recruit patients to a study that is poorly designed and which cannot provide an unbiased answer to the question posed. In many cases, RECs themselves recognise that they do not have the expertise necessary to assess scientific and safety issues, a problem that is exacerbated by the expectation that RECs act independently, which has been interpreted in some countries as including acting independently of each other. (Commonwealth Department of Health and Ageing 1999) The workload burden of RECs is further complicated by the pressure they are under to make decisions within relatively short time frames, and by an increasingly litigious society, with legal and indemnity issues taking up growing amounts of time on REC agenda. (Hendrick 2001)

From the perspective of clinical researchers, a common complaint is the amount of time and effort it takes for multi-centre clinical research to be submitted to and processed by RECs. The research protocol must be approved by the ethics committee at every institution participating in the study. If the study is being conducted in 20 hospitals there are usually 20 associated ethics committees, often with 20 different application forms, all with different requirements of the researchers. If the study involves more than one country then the cross-cultural issues can further complicate matters. As a result, there is a perception amongst clinical researchers that there is unnecessary duplication of effort across multiple ethics committees for the same multi-centre clinical trial. (Burman, Reves, Cohn, & Schooley 2001;Crooks, Colman, & Campbell 1996;Levine 2001;Wolf, Croughan, & Lo 2002)

A number of countries have implemented centralised systems aimed at improving the evaluation process, primarily in terms of the time taken for trials to obtain ethics approval and hence commence recruitment. (Christian, Goldberg, Killen, Abrams, McCabe, Mauer, & Wittes 2002;UK National Health Service 1997) The systems vary in detail and involve centralisation of the entire ethical review process (in the UK, for example), or elements of the process, particularly scientific review (for example, a pilot project being conducted in Australia: http://www.health.nsw.gov.au/public-health/rad/Ethics/Sharedassess/share_index.html). Critics of some centralised review systems suggest that they increase the burden on researchers and ethics committees by adding another level of bureaucracy. (Alberti 2000) The reason for this is possibly difficulties encountered achieving a balance between aspect of research that are reviewed centrally, and those that are reviewed locally. In addition, models for reviewing the scientific aspects of a trial protocol vary across countries, with some requiring review and approval by a funding agency prior to ethical review (the assumption being that the funder has conducted a review of the science and considered it to be adequate), and others requiring ethical approval before the funding agency gives approval. An assumption in the whole process is that ethics committees and funding agencies have the same expectations regarding scientific quality.

The intention of the peer review process is to confirm the scientific validity of the research that has been (or will be) conducted. It has also been suggested that there is

an ethical basis for peer review, in that "they are sustained in the course of careful examination, vigorous exchange of views, sound argument, and connections to the values of the society in which they occur". (Fletcher & Fletcher 1999) The effectiveness of peer review may be influenced by quality of the referees selected, the number selected, the instructions they are given, and whether or not they are masked to the identity of the manuscript authors.

As demonstrated by 2 relevant systematic reviews, the impact of peer review on the quality of research or research publications is uncertain. (Demicheli & Di Pietrantonj 2003;Jefferson et al. 2006) Both reviews attempted to identify comparative evidence addressing the value of peer review. Both categorised the processes of peer review as: i) different ways of assessing, assigning or masking submissions, ii)) different ways of eliciting internal or external opinions, iii) different decision-making procedures (group or single person) or iv) different types of feedback to authors and subsequent revision of submissions.

The review by Demicheli and Di Pietrantonj set out to identify comparative studies investigating the effectiveness of editorial peer review (defined as "procedures aimed at assessing and ensuring the scientific quality of output") in improving the quality of grant applications (Demicheli & Di Pietrantonj 2003). No studies were found that compared the effects of peer review with doing nothing and hence it is not possible to make a judgement about the impact of peer review of grant applications on the quality of research.

Jefferson *et al* concluded that there was insufficient evidence to be able to make a judgement about the value of editorial peer review in ensuring the quality of biomedical research publications. (Jefferson, Rudin, Brodney Folse, & Davidoff 2006) Of the 28 studies included in this review, only 4 specifically assessed the effects of peer review on study validity, only one of which was a randomised trial. In this study, 82 manuscripts submitted for publication were randomly allocated to either joint statistical-clinical review or clinical review only (Arnau et al. 2003). The authors of the report of this trial concluded that adding statistical review improved the quality of the final manuscript.

There is a need for more efficient and effective systems to facilitate the ethical review of multi-centre clinical research. It is apparent that a number of countries have developed and implemented their own unique systems and it is important that their impact be evaluated to determine whether they help or hinder the clinical research process. At the same time it should be possible to ascertain whether these systems ultimately influence the quality of approved clinical research.

Note: a protocol for this review has been published on the Cochrane Library (Ghersi & Dickersin 2004) The format of this chapter follows that required of a Cochrane methodology review.

Objectives of systematic review

To evaluate the impact of central (or "shared") scientific and/or ethical review of multi-centre clinical trial protocols on the clinical research process. That is, does centralising all or part of the ethical review process improve the quality of approved research, minimise unnecessary delay and result in improved decision-making.

Criteria for considering studies for this review

Types of studies

Prospective or retrospective comparative studies with two or more comparison groups were considered potentially eligible for inclusion in this review. That is, studies that compared a centralised ethical review process with a non-centralised process. These groups may be generated by random or other methods and could include historical comparisons. All studies must have reported original data. All potentially eligible studies were considered regardless of publication status.

It was decided *a priori* that if no comparative studies could be identified then single-armed studies would be described, recognising the limitations of such studies. As all studies ultimately identified by the search strategy can be described as either case studies or case series, the following definitions of these study types was applied:

Case study: describes an experience with a single study submitted to 1 or more committees.

Case series: describes an experience with 2 or more studies submitted to (usually) 1 or more committees.

The method used to select cases for inclusion in case series, when reported, has been documented in the Table of Included Studies using the following classification.

- Selected case series: studies where it is clear that information on a selected or incomplete set of trials were reported
- Consecutive case series: studies where it is clear that the intention was for information on all trials within a specified time period to be included. eg first 100 applications in a particular year, all applications in a particular year.
- Case series unknown: case series where it is not possible to tell how cases were selected

Types of participants

Clinical trials submitted to one or more human research ethics committees for approval. It was expected that studies and participants would be identified in the following combinations:

- a single trial protocol considered by multiple committees
- multiple trial protocols considered by multiple committees
- all trial protocols submitted to a single committee
- all trial protocols submitted to multiple committees

Types of interventions

- Centralized or shared review
    - where the entire review process (scientific and ethical review) is shared or centralised
    - institutional ethics committees may then be responsible for reviewing the study from a local perspective but should not repeat the work performed centrally
    - this may include systems where ethical review is performed centrally and institutional ethics committees monitor the study locally after it has been approved centrally

- Mixed review
  - where part of the review process is shared (eg scientific review) and other parts of the review process are decentralised (eg ethical review)
- Decentralized review
  - where the entire review process (scientific and ethical review) is performed by multiple institutional ethics committees. There is no sharing of processes.

Types of outcome measures

- Impact on clinical research
  - impact on study quality
  - type of problem identified (type of change requested)
  - time to first patient recruited
  - time to last patient recruited
  - rate of accrual / recruitment period
- ethics committee decision
  - time to ethics committee decision (defined as the period between the date that the application was received by REC and the date of their decision)
  - time to actionable approval document
  - approved (un/conditional yes or no)
- Impact on the ethics committee
  - time taken to review each protocol
  - resources consumed
  - average number of re-submissions per protocol
- Acceptability of process (by investigators and researchers, by sponsors and funders, by ethics committees)

There are a number of potential confounding factors that may be associated with the time it takes to obtain ethics approval. These include:

- Administrative tasks unrelated to ethical review but expected of RECs. These vary across countries but include:
  - indemnity issues
  - regulatory issues

- Resource issues (eg number of staff and workload at the "secretariat" of each REC)
- REC processes (even RECs within a single country function differently. For example, if they have an "executive" or other mechanism to expedite review, in the way they distribute work to REC members, how many REC members review each protocol, if a primary reviewer is identified, etc)

An attempt was made to collect data on these confounding factors to aid in the interpretation of the results of the review.

Search strategy for identification of studies

The Cochrane Methodology Register was searched as well as the databases MEDLINE, PREMEDLINE, EMBASE and CINAHL. The first search strategy defined in Table 62 was applied to MEDLINE (via Ovid) on 24th March 2003 and was updated on 4th March 2005. Before applying the updated search the keywords and MeSH headings of the previously identified eligible studies were examined and it was discovered that almost all eligible studies had been allocated the term "multicenter studies", and that attempts to narrow the search based on study design was of limited use. When updating the search the second strategy in Table 62 was therefore applied. The reference lists and bibliographies of eligible studies were also searched.

Table 62: Search strategy

Search 1

1 exp ethical review/ or exp ethics committees/ or exp ethics, institutional/ or exp ethics, research/ (12179)
2 exp Advisory Committees/ (4983)
3 exp Professional Staff Committees/ (12791)
4 exp Peer Review/ (7148)
5 or/1-4 (27025)
6 (ethic$ adj review).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading] (193)
7 (ethic$ adj committee).mp. [mp=title, abstract, cas registry/ec number word, mesh subject heading] (1405)
8 or/6-7 (1575)
9 5 or 8 (27740)
10 limit 9 to (controlled clinical trial or randomized controlled trial) (249)
11 exp epidemiologic studies/ or exp clinical trials/ or exp feasibility studies/ or exp intervention studies/ or exp pilot projects/ or exp sampling studies/ or exp Epidemiologic Research Design/ (1049656)
12 evaluation.mp. (356471)
13 audit.mp. (8925)
14 11 or 12 or 13 (1328863)
15 9 and 14 (3693)
16 10 or 15 (3791)

Search 2

1 exp ethical review/ or exp ethics committees/ or exp ethics, institutional/ or exp ethics, research/ or Ethics Committees, Research/ (9373)
2 exp Advisory Committees/ (3326)
3 exp Professional Staff Committees/ (9952)
4 exp Peer Review/ (7625)
5 (ethic$ adj review).mp. (1523)
6 (ethic$ adj committee).mp. (1341)
7 IRB$.mp. (1908)
8 or/1-7 (26150)
9 exp Multicenter Studies/ (8866)
10 8 and 9 (224)

Methods of the review

All citations were imported into a Reference Manage database. Two individuals (DG and a research assistant) separately examined all citations (including the abstract) identified by the search strategy and screened out clearly ineligible studies. Full papers were extracted for citations where both individuals agreed that the reference might relate to a potentially eligible study, as well as for those where there was disagreement. These full reports were read to determine the design of the reported study.

The intention was to evaluate the quality of any non-randomised controlled studies we found using the criteria outlined in the MERGE instrument. As we were unable to identify any controlled studies it was not possible to use this instrument. The only study types identified were case studies and case series (as described above) and the type of each study has been documented in the Methods column of Appendix 8.1: Characteristics of included studies.

Two reviewers (DG, KD) extracted outcome data for each eligible study. Data extracted included:

- study design, method of allocation concealment (if applicable), inclusion and exclusion criteria, interventions (including the characteristics of each committee, if available) and recorded outcomes for each study
- descriptive data about each study (study population, intervention, outcomes, etc)
- information on potential confounders

Statistical pooling of the results of studies was not appropriate or feasible given the nature of the studies identified. A table has been created for each outcome, summarising the information available for each outcome, by study (see Appendix 8.3: Data tables: systematic review).

Description of studies

The above search strategy was applied on the 24th April 2003 resulting in the identification of 2,531 references. These were imported into bibliographic software (Reference Manager). A duplicate check performed using Reference Manager reduced the number of references to 1,922. The updated MEDLINE search performed on 4th March 2005 resulted in the identification of an additional 74 citations. After preliminary screening of the abstracts to identify potentially eligible studies, 1720 references were excluded. The full paper was retrieved for the remaining 276 citations and each was assessed for inclusion in the review by DG.

As already noted, no controlled studies were identified. Of the 34 studies included, 23 were case studies and 11 were case series. An additional 23 potentially eligible studies were excluded for reasons outlined in Appendix 8.1: Characteristics of included studies. The remaining 219 short-listed studies were deemed to be ineligible mainly because the full article did not report original data. Studies that had the potential to contribute information to the review but were not strictly eligible were excluded. The details for these studies are described in Appendix 8.2: Characteristics of excluded studies (page 235).

Studies identified in the intervention category "Centralized or shared review" were all case studies of a single trial submitted to a central committee and then to multiple local committees. The intervention for these studies is therefore referred to in this review as "Central then local REC/s". No studies were identified in the intervention category "mixed review". Decentralized review included studies describing the experiences of a central review committee only (referred to as "Central REC") or a local ethical review committee only (referred to as "Local REC/s"). The intervention category for each study is described in the Intervention column of the table "Characteristics of included studies" (See Appendix 8.2: Characteristics of excluded studies, page 229).

Of the 11 case series, 7 were classified as Local REC/s and 4 as Central REC (Table 63). Of the 23 case studies, 14 were classified as Local REC/s, and 9 as Central then Local REC/s. Note: One article ((Dunn, Arscott, & Mann 2000a)) reported 2 case

studies: one classified as Local REC/s, the second classified as Central then Local REC/s.

Table 63: Type of study by intervention

Study type	Type of intervention			Total
	Local REC	Central then Local REC/s	Central REC	
Case study	14	9	0	23
Case series	7	0	4	11
Total	21	9	4	34

The oldest study identified was a case series of 79 projects submitted to a local REC between 1969 and 1970. (Gray 1975) However, the majority (17) of studies in this review reported experiences during the 1990's. Many studies were performed in the UK, with others performed in other European countries including Spain and Finland. There were also studies involving ethics committees in the USA, Canada, Australia and New Zealand.

The quality of available studies was poor. Most case studies appear to be opportunistic reports of experiences with a single study with no stated *a priori* objectives. Where identifiable, the objective of the reported study has been documented in the "Notes" column of Appendix 8.1: Characteristics of included studies. The quality of case series was variable. One study was conducted, for example, by an individual member of a REC reviewing his files and another examined only projects in selected specialties (Boyce 2002b;Gray 1975). Others examined all studies submitted over a defined period (eg (Faccini, Bennett, & Reid 1982;Keinonen et al. 2001)

Results

Impact on clinical research

Although 12 studies reported the changes requested by ethics committees in order for the trial to obtain approval, it is not possible to determine if the requested modifications were reasonable or resulted in an improvement in the quality of the studies submitted (Appendix 8.3: Data tables: systematic review; Table 72, Table 73 and Table 74). The nature of the changes requested include those that could be described as "classic" ethical issues, such as the wording of patient information sheets, consent forms, safety, etc. Methodological issues included the choice of comparator, statistical issues (including the sample size calculation), compliance issues, selection criteria, etc. This would suggest that RECs consider the quality of trials important in their decision to approve or reject an application.

Two studies suggested that the ethical review process resulted in a delay in the time taken to recruit participants to their studies but evidence was not provided.

Ethics committee decision

Most studies (19) reported on the time taken to receive ethics approval, with a diverse range of reported approval times (see Appendix 8.3: Data tables: systematic review; Table 77, Table 78 and Table 79). Some studies reported approval times as long as 298, 346 and 408 days although if these studies are generalisable, these would appear to be extreme cases. It is not possible to tease out where the time was spent or why - the time from submission to approval usually involving activity on the part of both the applicant and the ethics committee. One case series involving a central REC reported a median time from first meeting to approval of 64 days (Boyce 2002a). The two case series involving more than 2 studies reported median times from submission to approval of 64 days (Dal Re, Espada, & Ortega 1999) and 45 days (Ortega & Dal Re 1995b), with the former also suggesting that trials without queries take less time to obtain approval.

Most studies would appear to be approved without change or be given conditional approval at the first meeting of the ethics committee (See Appendix 8.3: Data tables: systematic review; Table 80, Table 81 and Table 82). Most studies would also appear

to be successful in obtaining ethics approval (See Appendix 8.3: Data tables: systematic review; Table 83, Table 84 and Table 85).

Impact on the ethics committee

None of the included studies reported on the impact that sharing part or all of the review process between a central REC and local RECs has on the work or decisions made by ethics committees.

Acceptability of process

None of the included studies reported on the acceptability of the process by investigators, researchers, sponsors, funders, ethics committees or others with a vested interest in the outcome of ethics review.

If resources consumed by applicants in order to make a submission are considered a measure of acceptability, then 14 studies provided estimates of resource usage (Table 86). The costs incurred include preparing the documentation (photocopying, postage, telephone), the time used to prepare and handle requests for changes (research staff, secretarial support, etc), fees charged by ethics committees to consider an application, and the impact of having to prepare multiple copies of multiple applications for multiple ethics committees.

Discussion

Given the importance of the ethical review process in the conduct of high quality clinical research, and the criticisms of this process, the poor quality of studies investigating their effectiveness is disappointing. Most of the information supporting debate is based on selected case studies of bad experiences that are unlikely to provide unbiased assessments of any problems that may or may not exist.

Performing any task adequately will take time, and one could question the appropriateness of "benchmarking" the time it should take for an ethics committee to perform the tasks with which it is charged, particularly considering the nature of those tasks. On the other hand, researchers are generally working within a restricted time frame (whether this is due to commercial pressures, the requirements of non-commercial funding agencies, or other reasons) and often need to demonstrate adequate progress in order to justify the resources expended. Obtaining ethics approval might be a deliverable in a funding contract, for example, and it can be exasperating for researchers when delays are for reasons that are out of their control. Commercial pressures do not, however, carry much weight with ethics committees, with one commentator suggesting that commercial practice "is not a matter of the greatest importance for patients taking part in research projects". (Ross 1994)

One apparent cause of delay would appear to be the time taken to respond to requests for changes. Again, it can be frustrating when requests are made to alter a trial that has already been reviewed by numerous committees, particularly in the case of multi-centre research when requests for changes to details such as design features (such as the intervention) may not be possible or practicable. Requests for clarification or change may, in some instances, reflect the lack of expertise available to the committee at the time the study was being considered. In other instances, the requests for change may be reasonable, and researchers should take their share of responsibility for delays that occur as the result of incomplete or carelessly compiled submissions. As a result of his review of 353 applications to an ethics committee between 1997 and 2000, Boyce concluded that "More care and effort by researchers in preparing applications, particularly information leaflets, would shorten approval time". (Boyce 2002b)

One reason for lack of consistency of decision making across ethics committees may be the remit of local ethics committees to consider "local issues" when they consider studies. Although these local issues may not always be obvious to the researcher, the resulting variability across RECs "is not inherently inappropriate". (Silverman, Hull, & Sugarman 2001) It is also not particularly surprising given that each committee is composed of different individuals, with their own values and beliefs. Local RECs are autonomous bodies and are given reasonable leeway by their governing agencies to interpret and apply the rules and regulations that apply to them.

One of the local issues assessed by RECs is the ability of the researcher making the submission to conduct the research appropriately. In some cases, rejection of the study may in fact be rejection of the researcher. Watling and Dewhurst cite a case where the reason for the rejection of a study was the lack of suitability of the local investigator. (Watling & Dewhurst 1995) The "committee felt unable to state this objection and opted instead for a less credible alternative", the reasons being to avoid the risk of being sued for defamation and "to avoid local controversy and confrontation". (Watling & Dewhurst 1995) Whether rejection of a submitted study is an appropriate way to handle situations such as these is debatable: researchers have the right to know if they have been deemed to be inappropriate as investigators by an ethics committee.

While some complaints made by researchers may be justified, it is possible that the ethical review process may be the only opportunity researchers have to express their discontent with what is becoming an increasingly regulated and monitored research environment. In 1999, for example, a member of an ethics committee that had been subjected to criticism suggested that the complaint submitted by the researchers was "misdirected, mistimed and possibly harmful". (Alexander 1999;Larcombe & Mott 1999) The proposal submitted by the researchers was perceived to have a number of problems, including issues around confidentiality, and was submitted at a time of major change in the local research ethics committee system in the UK, including the implementation of international standards for good clinical research practice. (Alexander 1999;International Conference on Harmonisation 1996) The author goes on to remind readers that "not all proposals are acceptable and members of ethics committees, especially lay members, spend much effort protecting the interests of the

patients for no discernable reward. Repetitious criticism will erode the willingness necessary to perform this necessary function". (Alexander 1999)

It is of interest to note that some researchers may approach the ethical review process expecting to have problems. Druml *et al,* for example, surveyed a group of physicians to assess the reputation and acceptance of a University-based research ethics committee in their region. They found that most of the respondents who had experience in the submission of an application gave the committee satisfactory ratings, while respondents without the experience of submitting and application tended to judge the committee negatively. (Druml et al. 1999)

There are many ways in which the relationship between researchers and RECs can be improved. It is important that the effect that the changing research environment has had on the process of ethics approval be recognised and openly discussed by researchers, ethics committees and others involved in the clinical research process. Appropriate investment needs to be made in the infrastructure necessary for RECs to function effectively and efficiently, including adequate funding of REC secretariats. There is also a need for researchers to be willing to contribute their scientific expertise to the process of ethics review. For researchers to volunteer to be part of the REC process they will, however, need to perceive REC membership as an activity that holds both appeal and reward. An improvement in the quality of submissions would also be useful, perhaps starting with standardisation of the various parts of the application process—including the forms used—and through improved education of both RECs and researchers.

RECs provide a valuable yet undervalued service to both researchers and participants. Researchers are, after all, ultimately accountable to the people who will consent to participate in their research. The difficulty is how to achieve a balance between the core function of an REC to review ethics, and the associated paperwork and bureaucracy. There is clearly room for improvement and the only way to be sure that any mechanisms introduced with the intention of making such improvements are evaluated appropriately.

Chapter 7: Evaluation of the NSW Health Shared Scientific Assessment Scheme

This chapter reports all aspects of the design, conduct and reporting of a prospective, single-arm, follow-up study of multi-centre clinical drug trials submitted to the NSW Health Shared Scientific Assessment Scheme (SSAS) between 1st January 2003 and 31st December 2003. The aim was to assess the influence of SSAS on Human Research Ethics Committees as well as on multi-centre clinical trials. The details will be presented as follows:

- Background to shared assessment
- Methods
- Results
- Discussion

Background

In 1996 a report was published by the National Health and Medical Research Council (NHMRC), which presented the results of a review of the role and functioning of Institutional Ethics Committees. (NHMRC 2002) This comprehensive report raised a number of important issues, including the problems caused by the lack of clarity regarding the ability of one REC to accept the decisions made by another. As a result, it is common practice in Australia for multi-centre research applications to be reviewed separately by multiple institutions. This may in turn result in (whether real or perceived) delays in obtaining approval, inconsistencies across RECs (submission processes, decisions made, changes requested) and unnecessary duplication of effort.

Traditionally, ethical review committees [also referred to as Health Research Ethics Committees (RECs)] have functioned in isolation from other RECs. The membership of an REC is expected to include representatives from the lay community and other non-scientists, yet these individuals are expected to acquire quite a detailed understanding of the methodology of clinical research in order to meet their ethical obligations. The 1996 review reported that:

> "In assessing a research proposal, IECs are concerned primarily with ensuring that the rights of the research subject take precedence over expected benefits

to knowledge. An understanding of the scientific and safety or privacy aspects of a protocol is an essential component of this review. From the submissions, many IECs do not believe they have the necessary expertise to assess science and safety issues. (NHMRC 2002)

Further:

"One submission noted that "...a central review of scientific merit could improve efficiency, decrease approval time, and take some the burden from ethics committees" ... This approach could streamline the review process, reduce duplication of effort by IECs and allow more efficient use of the available expert opinion." (NHMRC 2002)

Shared Assessment Schemes Internationally

Recent years have seen a number of countries, including the United Kingdom (UK), Czechoslovakia, Germany and New Zealand, implement centralised systems aimed at improving the evaluation process, primarily in terms of the time taken for trials to obtain ethics approval and hence commence recruitment. Critics of centralised review systems suggest that the opposite is the case, with such systems being accused of increasing the burden on researchers and ethics committees by adding another level of bureaucracy.

The only system to have been reported in any detail is in the UK where Multi-centre Research Ethics Committees (MRECs) were implemented in 1997. The goals of MRECs were to simplify and speed up the process of ethical review for multi-centre research, and to improve consistency between ethics committees. (Tully et al. 2000) At that time, the UK was divided into 11 health care regions and an MREC was established for each region. The principal researcher submitted the research proposal to their designated MREC who then considered the proposal. It was expected that scientific peer review would be completed prior to submission to the MREC. An approval letter was issued and disseminated by the principal researchers to local researchers who then submitted it with their application to their local research ethics committee (LREC). The LREC was permitted to change the consent form or patient information sheet but could not change the protocol.

In 2000, Alberti discussed some of the early anecdotal feedback on the MREC system. The concern expressed by researchers and ethics committees in the UK was that the new system, far from achieving its goals, may have added "yet another layer of bureaucracy ... making the process even more labyrinthine". (Alberti 2000) A major problem appeared to be the LREC's perceived loss of independence, and it was felt that this may have primarily been due to poor communication between MRECs and LRECs. Mechanisms were put in place to improve this.

Alberti's anecdotal findings were supported by a study reported at the same time by Tully *et al* who conducted a prospective study aimed at evaluating the MREC system. (Alberti 2000) This case study involved a single, multi-centre research proposal, submitted to an MREC in September 1998. After approval by the MREC, the proposal was submitted to 125 LRECs (50 (40%) of which had executive sub-committees that could consider expedited review). The authors were interested in a number of outcomes, including the time to approval by each LREC, and the number of non-local changes that were requested.

Approximately 88% of proposals were approved within 8 weeks of submission (50% approved within 4 weeks) if the LREC had an Executive subcommittee. Those without an Executive subcommittee took almost twice as long (15 weeks, 25% approved within 4 weeks). Non-local changes requested included changes to information sheets (18% of committees), changes to consent procedure (10%), changes to questionnaire (7%) and changes to methods of recruiting subjects (7%). The authors concluded that improvements in the system had occurred. Although some problems remained they were primarily structural and logistic and not due to substandard work of LRECs.

Shared Assessment Schemes in Australia

In 1999 the NHMRC released its National Statement on Ethical Conduct in Research Involving Humans incorporating many of the recommendations made in the review.(NHMRC 2001) Section 3 of the National Statement is dedicated to multi-centre research and allows RECs to "accept a scientific/technical assessment of the research by another institution or organization". This clause has resulted in various

models for sharing scientific or ethical review. At least one state is looking at implementing a single, central ethical review where RECs would not be responsible for initial approval but would monitor trials. In Victoria, a group of hospitals involved in the treatment of patients with cancer are investigating shared ethical review, agreeing to abide by the decision made by any one of the ethics committees in the group. NSW will be piloting a process for sharing review of scientific and safety aspects, leaving RECs with responsibility for reviewing the trial from a local ethical perspective.

About the NSW Health Shared Scientific Assessment Scheme

In February 2003 NSW Health established the Shared Scientific Assessment Scheme, the intention being to provide a central committee that would review the scientific components of a clinical trial before submission to any of the 22 RECs affiliated with NSW Health. LRECs remained responsible for reviewing the ethics of the trial at a local level. It was hoped that the activities of the committee would not prolong the time taken to obtain ethics approval and might alleviate some of the workload of LRECs (particularly those without access to scientific expertise at a local level).

For a clinical trial to be eligible for consideration by SSAS as part of the pilot it needed to be randomised and potentially involve 3 or more RECs affiliated with the NSW Department of Health. The trial could either be submitted by the Sponsor (see Glossary on page xiv for definition) before they submitted to any institutional RECs (RECs) or, if the Sponsor did not submit the trial to SSAS, a REC could choose to submit the trial to SSAS themselves.

The Shared Scientific Assessment Committee (SSAC) met once a month to discuss submitted trials. Submission deadlines were published on the SSAS web site and circulated to all REC Executive Officers.

Following each SSAC meeting a response was sent to the relevant Applicant informing them of the result of the meeting. Correspondence continued until a Final Report was agreed on, the intention being that this report would be submitted with the application to each REC. The Final Report could be sent by SSAS to the Sponsor for dissemination to RECs, or directly to RECs who the Sponsor had indicated would be participating in the trial on the SSAS Application Form. It was the Sponsor's decision as to who would ultimately provide the REC with the SSAS Final Report.

The RECs were advised to consider the trial as per usual practice for that REC, replacing their normal scientific review process with the SSAS Final Report.

Complete details of the SSAS process are described in the SSAS Manual. This is available from the NSW Health website. (http://www.health.nsw.gov.au/public-health/rad/Ethics/Sharedassess/share_index.html).

Aim

The aim of this study was to evaluate the NSW Health Shared Scientific Assessment Scheme (SSAS) by assessing its influence on Human Research Ethics Committees as well as on multi-centre clinical trials. It was hypothesised that Human Research Ethics Committees affiliated with NSW Health would find SSAS to be useful in the ethical review of multi-centre clinical drug trials, and that the SSAS process would not significantly prolong the time taken for multi-centre clinical drug trials to obtain ethics approval.

Methods

The study was a prospective, single-arm, follow-up study of multi-centre clinical drug trials submitted to the NSW Health Shared Scientific Assessment Scheme (SSAS) between 1st January 2003 and 31st December 2003.

Trials were followed for at least 4 months or until a final decision was made by each REC, whichever occurred first. Based on the experience of Multi-centre Research Ethics Committees (MRECs) in the UK it was anticipated that a final decision would be made by each REC within 4-5 months of submission in 90% of cases (Tully, Ninis, Booy, & Viner 2000).

If 25 trials were submitted to SSAS during the pilot, and the percentage of trials where SSAS made a difference to an outcome was 50%, this would provide 95% confidence intervals of ±19.6%.

Ethical issues

Ethical approval to conduct this study (known as "the Evaluation") was obtained from the University of Sydney Human Research Ethics Committee. Written informed consent was obtained from all Applicants submitting eligible trials to use their trial in the evaluation of the SSAS as part of the application process to the SSAS. Written

informed consent was also obtained from RECs who were asked to consider participating in the Evaluation at one of their meetings. The Chair then signed the consent form on behalf of the REC and sent it to the SSAS Secretariat, along with a completed Baseline Information form.

Data Collection

When the Final Report was returned to the Applicant it was accompanied by one SSAS Evaluation Form (EF) for each REC listed by the Applicant in the SSAS Application Form. The Applicant was asked to include an EF with each application made to each REC, and to keep a log of forms to enable them to be tracked.

RECs were asked to treat trials that had been through SSAS as normal, replacing their usual scientific review process with SSAS and the Final Report. After making a decision regarding the trial at their meeting, RECs were asked to allocate a maximum of 10 minutes in total to discuss the following 3 questions:

- In the general opinion of the REC, did the SSAS Final Report reduce the overall time taken to consider the trial at the meeting?
- In the general opinion of the REC, did the SSAS Final Report improve the committee's confidence in their decision?
- Was the information provided by the SSAS Final Report useful?

Following this discussion, the EF was then completed by the Executive Officer or Chair on behalf of the REC. They were asked to ensure that the opinions expressed were those of the committee as a whole and not of the individual/s completing the form. Completed forms were sent to the SSAS Secretariat who then masked the identity of the individual trials, Applicants and RECs and forwarded the forms to the Evaluator (DG).

Please note that all contact with ethics committees and trial sponsors was performed by staff at NSW Health. The author was blinded to the identity of both the ethics committee and the trial sponsors and was unable to contact individuals directly.

Results part 1: About the RECs

There are 22 RECs affiliated with NSW Health eligible to participate in the pilot. All but one indicated their agreement to participate in the SSAS Evaluation in writing, and provided baseline information on committee membership, the relevant standard practices of their committee and their workload experience during the year 2002.

Each REC reviewed an average of 132 research proposals in 2002 (range 5 – 446; see Table 64) Fifteen RECs were able to provide information on the approximate proportion of research projects that were multi-centre drug trials. On average 35% of research projects reviewed by RECs are clinical drug trials, and about half of these (45%) are multi-centre. (See Figure 7).

The average number of members on a REC is 15. Most committees (13) committees meet once each month, 5 meet every second month, 2 every 6 weeks and 1 REC meets once every 3 months.

Table 64: About RECs in NSW

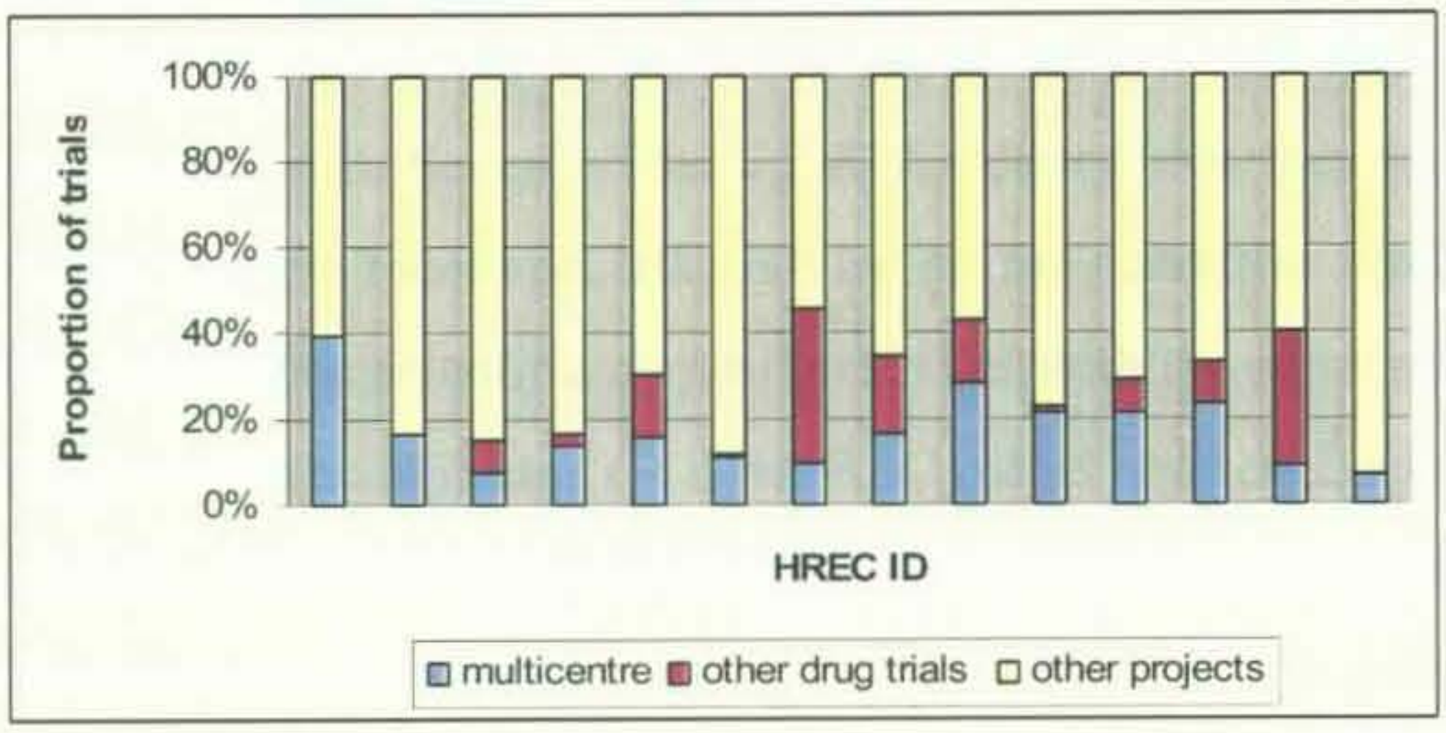
Item	Total across RECs	Average	Range
Number of members	261	15 members	9 – 27 members
Average length of meetings	n/a	3	1.5 – 5.5 hours
Number of research projects reviewed in 2002	2766	132	15 – 446 projects
- Clinical drug trials [Note 1]	695	35	0 – 140
- Multi-centre clinical drug trials [Note 2]	313	16.5	0 – 55
Average time REC Secretariats spend preparing each multi-centre drug trial for an REC meeting [Note 3]	n/a	3.35	1 – 8 hours
Average time REC members spend preparing for meetings [Note 3]	n/a	5.4	2 – 10 hours

Note 1: 3 RECs did not review any clinical drug trials. The average for this and the following question is therefore calculated based on 18 RECs.

Note 2: One REC with a larger workload (>300 projects) answered this question as "lots" so unable to include this data. The estimated proportion of trials that are multi-centre is therefore likely to be an underestimate

Note 3: 2 RECs did not answer these questions

Figure 7: Proportion of trials that are multi-centre and/or drug trials



Note: Two committees are closely affiliated and filled out one baseline form between them.

Access to specialist expertise

Fourteen RECs indicated that they either have a specialist committee, use external experts, or both. Seven of the 21 RECs stated that they had access to a committee formed to give them specialist advice on the scientific / technical aspects of clinical drug trials. The scientific committees meet with the same regularity as the REC with which they are associated (ie if the REC meets once each month, so does the scientific committee). Ten RECs use external experts (some RECs with scientific committees also use external experts).

Thirteen RECs indicated that they assign members to look at applications in more detail. Individuals are usually assigned according to their content expertise and/or to ensure that a minimum number of members have looked at a submission in detail.

Table 65: About scientific advice obtained by RECs in NSW

Item	Total across RECs	Average	Range
How many people are on the scientific committee (7 responses)	80	11	4 – 19
Average length of scientific committee meetings (7 responses)	n/a	2.2 hours	1 – 3 hours
How many hours of external expertise would you access each year (8 responses) Note 1	449	50	1 – 100 hours

Note 1: 2 of the 10 RECs using outside experts did not indicate the number of hours accessed each year

Estimated resources

The average REC expended a total of 953 person-hours per year preparing for meetings, plus 485 hours per year attending REC meetings. This is the full-time-equivalent of just over 4 days a week per committee which, when extrapolated to all 22 RECs, is comparable to a full-time state-wide committee composed of at least 16 people. Added to this is an estimated 2,315 person-hours spent on specialist review in NSW, equivalent to at least 1.4 people dedicated full time to this task state-wide.

Table 66: Person-hours spent on REC or scientific review

Item	Total across RECs	Average	Range
Estimated person-hours per year spent attending REC meetings (21 responses) (number of REC members in NSW x length of meetings x number of meetings each year)	10,181 FTE 6.06	485	54 – 1,254
Estimated person-hours per year preparing for meetings (19 responses) (number of REC members in NSW x average time REC members spend preparing x number of meetings each year)	18,098 FTE 10.77	953	243 – 3,240
Estimated person-hours per year for scientific review meetings (number of members of scientific committees in NSW x length of meetings x number of meetings each year)	1,866 FTE 1.11	133	0 – 540
Estimated person-hours per year for external scientific review (as estimated by REC)	449 FTE 0.27	50	1 – 100
Estimated person-hours per year spent by REC secretariats preparing multi-centre drug trials for REC meetings (number of multi-centre drug trials x time REC secretariat spends preparing each)	1,055 FTE 0.63	59	0 – 184

Assumption: FTE are calculated assuming a 35 hour week, 48 weeks per year (ie 4 weeks annual leave)

Results Part 2: About the Submitted Trials

The first meeting of SSAC was held on 10th February 2003. As of 26th February 2004, 27 projects had been submitted to SSAS, of which 22 have been approved and hence have a final report available. One project has been rejected, one has been withdrawn and 3 have yet to be approved.

SSAS takes an average of 9 days to respond to an Applicant following a SSAC meeting, and Applicants take an average of 24 days to respond for the first time to specific queries. (See Figure 8 and Table 67). The application is the responsibility of the applicant for an average of 33 days (range 3 to 125 days; median 31 days), and the responsibility of SSAS for an average of 44 days (range 24-81 days; median 41 days). The average time from submission to the first response to the applicant by SSAS is 25 days (range 17 to 53 days, median 25 days).

As outlined on page 5, Sponsors were asked to keep a log of RECs and document the date they received an actionable approval document. Unfortunately only one of these log forms was returned to the SSAS Secretariat and it is not possible to report on the time taken to obtain an actionable approval document.

Figure 8: Time from submission to SSAS to Final Report

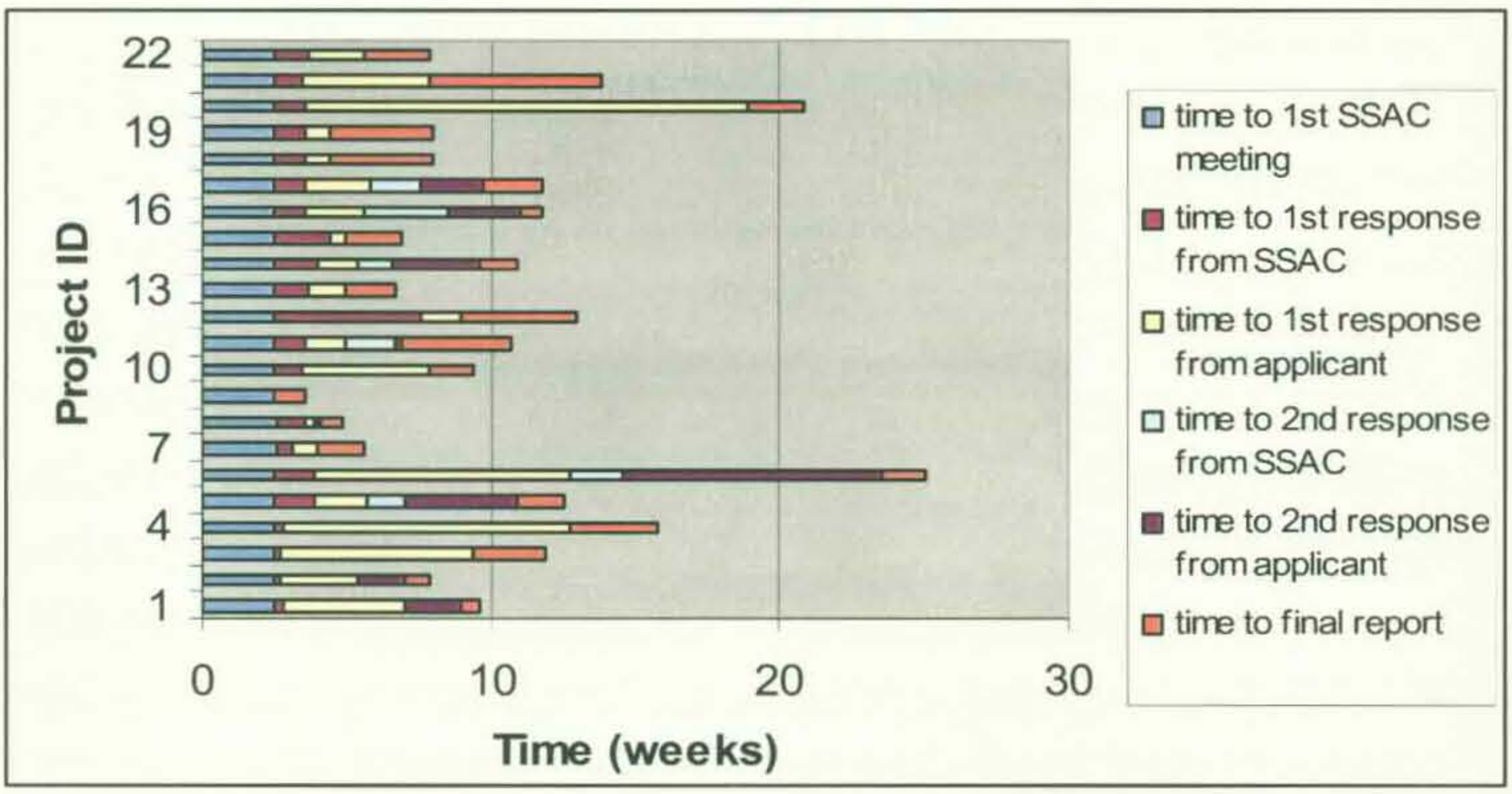


Table 67: Time taken for each step of the SSAS process

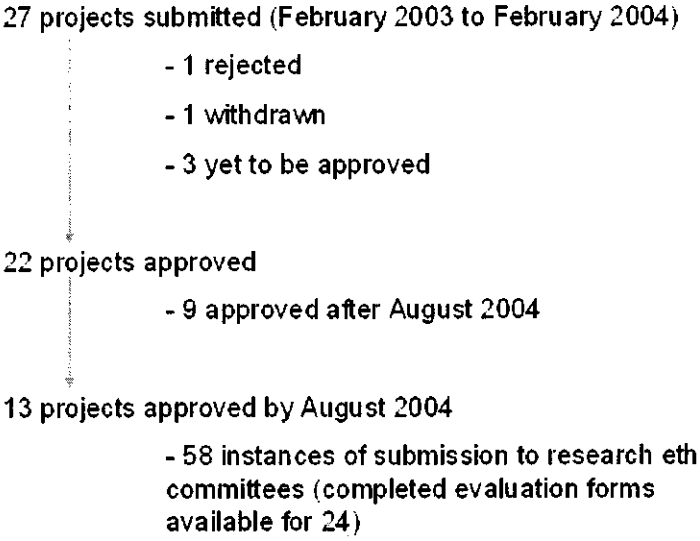
Time interval	Average number of days (median)	Range
Time to 1st SSAC meeting	17 (17)	17 – 18
Time to 1st response from SSAC	9 (8)	2 – 36
Time to 1st response from applicant	24 (13)	2 – 107
Time to 2nd response from SSAC	11 (12)	1 – 21
Time to 2nd response from applicant	19 (15)	1 – 63
Time to final report	15 (13)	4 – 32
Overall time from submission to SSAS to Final report	75 (71)	25 – 175

Results Part 3: the evaluation

Trial by REC Data Available for Evaluation

Completed *Evaluation Forms* (See Appendix 9: SSAS invitations and data collection forms and Figure 9) are available for the 13 trials submitted to SSAS between February and August 2004. The Applicants of all 13 trials agreed to the inclusion of their trial in the SSAS Evaluation.

Figure 9: flow chart of studies submitted to SSAS



In the application to SSAS, Sponsors are asked to list those RECs to which they intend to submit the trial. Based on information included in SSAS application forms there were potentially 70 instances when the 13 trials should have been submitted to eligible RECs. It is known that in 12 of these instances the trial was not submitted to the nominated REC, and that completed evaluation forms are available for 24 of the 58 instances when the trial was submitted to the nominated REC; a response rate of 47%. It is not known if the reason for this is that Sponsors did not make the Final Report available to RECs, or if RECs did not complete the EF for trials considered by their committee. Staff at NSW Health continued to encourage Sponsors and Ethics Committees throughout the duration of the pilot. Meetings were held with members of ethics committees and representatives of trial sponsors to further educate individuals on the reason for the pilot and the need for evaluation. NSW Health staff also

followed up individuals with at least 1 telephone call in an attempt to obtain completed forms.

Six RECs reviewed 3 or more of the SSAS evaluated trials, and 4 trials were submitted to 3 or more RECs. Unfortunately, 9 completed EFs were for trials where the REC indicated that the Final Report had been submitted to SSAS after the trial had been considered by the REC. These will be referred to as "late" EFs and the remainder will be referred to as "on-time". In the case of three of the late EFs the RECs considered the SSAS report after the meeting and suggested additional changes, if necessary, in their second response to the applicant.

On 8 occasions the date of the Final Report was after the date of the REC meeting. Six of these were "late" EFs as stated above, and the remaining 2 appeared to defer their decision until the Final Report became available. There were 16 on-time evaluation forms for 11 trials submitted to 6 RECs.

SSAS Evaluation Form for RECs

The main results are based on data provided by the 16 on-time EFs plus the 2 EFs relating to trials where the decision of the REC was deferred until the Final Report became available. Results are reported for the 9 late EFs if relevant and available. Each question on the form will be addressed in order, with a summary of responses from each REC. Some of the comments made on the evaluation forms by RECs are included for additional information. These were selected based on their relevance to the question being addressed.

In the general opinion of the REC, did the SSAS Final Report reduce the overall time taken to consider the trial at the meeting?

Seventeen EFs indicated that the Final Report reduced the overall time taken to consider the trial at the REC meeting. One REC was uncertain of the Final Report's influence on the time taken to consider the particular trial at the meeting, but the points raised by SSAS were discussed by that REC in some detail, specifically in the context of the study being carried out at that REC's institution.

"The SSAS Final report assessed the scientific merit of the study and hence did not require review by our Clinical Trials Sub-Committee. The project went directly to our Ethics Committee for ethical review"

"It reduced the time undertaken by the REC considering scientific issues and raised important areas for the REC to discuss"

"A study such as this would normally be reviewed by 2 people and discussed at the scientific sub-committee meeting. Approximate time saved = 8 hours."

In the general opinion of the REC, did the SSAS Final Report improve the committee's confidence in their decision?

All 18 EFs indicated that the SSAS Final Report improved the RECs confidence in their decision. In addition, in 4 of the 9 occasions when the Final Report was late, the RECs involved reported that it improved the committee's confidence in their decision.

"... the report confirmed that the RECs decision was the correct one..."

Was the information provided by the SSAS Final Report useful?

17 EFs indicated that RECs found the report to be either "very useful" (13) or "reasonably useful" (4) (this question was not answered on the remaining form). Four of the late evaluations also found the Final Report to be either very or reasonably useful.

"Information very useful in clarifying some issues raised by REC"

"It would have been very useful if it had been received prior to the review"

Decision made at the REC meeting

Three applications were approved unchanged, 14 were approved with changes and one was 'decision pending'. Eight of the nine late EFs related to trials approved with changes and one decision pending.

The changes requested by RECs, other than those relating to ethics, could be categorised as issues relating to:

- Changes to Patient Information Sheet and/or Consent Form
- Insurance and Indemnity
- Restrictions placed by Sponsors on publication / presentation / data ownership
- Clarification regarding trial and local context (eg drug storage, availability of specific tests within health service, etc)
- Funding and financial issues (eg who pays for the intervention)
- Trial management (access to data, data collection forms, clarification of the existence of a Data and Safety Monitoring Board

and

- Science

When the SSAS Final Report was available (and the Sponsor had not made changes requested by SSAS) it was not uncommon for RECs to explicitly ask for SSAS recommendations to be incorporated.

Other issues that may have caused a delay in the review of the trial by the REC

The issues causing delay were similar to the changes requested following the first REC meeting. The most common issues reported that may have caused delay in review were legal issues and incomplete applications (eg missing clinical trial agreements, copies of data collection forms, etc).

SSAS Evaluation Form for Applicants

All applicants were asked to complete an evaluation form for each trial submitted to SSAS. Unfortunately only 3 applicants completed this form. It is not clear why Applicants did not complete and return this form.

Applicants were asked to rate, on a scale from 1 (very satisfied) to 10 (very dissatisfied), their experience with SSAS regarding their trial. The three applicants were very satisfied with their experience with an average score of 3 (range 2-4).

Aspects of SSAS the applicants are happy with were:

> *"Compared to other states the SSAS Report appears to speed ethical approval of the trial up, as the 3 NSW sites were amongst the first sites to receive approval - so seems to assist! All interactions with SSAS were pleasant and professional. Glad to only have to answer the scientific questions once for 3 sites rather than 3 times over"*

> *"Easy, straightforward application form. Good idea to have a central committee that will have the expertise not necessarily available at all RECs. Sponsor being able to apply and have direct contact is beneficial."*

> *"The scientific review was sound. The time scale and communication was good."*

Aspects of SSAS the applicants are not happy with include:

> *"Required follow up by CRA to obtain questions posed by the committee and the Final Report quickly"*

> *"Increased length of time for submission, all RECs needing SSAC approval are initially held up."*

> *"It's a whole other form to complete and copies to make. Still have to do the full ethics application."*

Other issues that may have caused a delay in the review of the trial by RECs include:

> *"Because REC meetings are not aligned to the SSAC meetings this process can take a lot longer. Would be better if the SSAC meeting output could feed into a round of REC meetings at all the NSW RECs"*

> *"Additional process required by University ethics committee costing 6-8 weeks"*

Results Part 4: End of study

In January 2004 *End of Study* forms were sent to the participating RECs and members of the SSAS Committee with a request to return them to the Secretariat by 27th February 2004 (see End of Study Forms on pages 255, 257 and 259). All 21 of the participating RECs, and 7 of the 8 SSAS Committee members completed *End of Study* forms. Please note: one institution has 2 RECs and each REC is considered as a separate entity although baseline details were reported on one form.

Medicines Australia is the national association representing the prescription medicines industry in Australia. They were asked to by the SSAS Secretariat to distribute the End of Study forms to all their members. RECs were also asked to distribute the End of Study forms to their researchers. This resulted in the completion of 21 *End of Study* forms, and 17 of these were completed by Sponsors who had submitted trials to the SSAS Committee as part of the Pilot.

Participation in Pilot

Of the 21 RECs that completed end of study forms, 16 indicated that they had reviewed a trial that had been submitted to SSAS. The 5 that had not reviewed any SSAS reviewed trials all indicated that no trials that were eligible for SSAS had been submitted to their committee. Three RECs insisted that Sponsors use SSAS, 6 suggested it and 1 did not answer the question. The remaining 6 RECs indicated that they had either been provided with the Final Report by the Sponsor as part of the application (2), or dealt with each trial on a case-by-case basis. One REC mentioned the difficulty they had determining the eligibility of a trial for SSAS, particularly regarding the criterion that trials needed to involve 3 or more sites in NSW. Some RECs also mentioned the difficulty they had determining if a trial that had been submitted to them had also been submitted to SSAS.

17 of the 21 Sponsors that completed end-of-study forms indicated that they had submitted a trial for consideration by SSAS as part of the pilot. One Sponsor did not answer the question and the remaining 3 did not submit a trial for the following reasons:

"Company had no relevant studies at the time of the evaluation scheme, however, I would be concerned about adding to the evaluation time for EC review- even though I understand the aim of SSAS is not to do this. Discussion with larger hospitals with "established" ECs indicate that there is likely to be a strong preference for "in-house" review due to potential medico-legal issues with running studies at the sites."

"No perceived advantage to direct application to RECs"

"The only information I received was at the NHMRC Conference in 2002 and so was not familiar with the submission process. Also, there was some concern that this was an "extra step" in the EC approval process, adding additional weeks to the approval process"

Eight Sponsors submitted trials to SSAS because at least one REC insisted they use it (plus one Sponsor who indicated that "*a couple of REC preferred that we used the SSAS prior to submission to the REC*"), and 8 chose to use SSAS without prompting by an REC (one of which indicated that they thought it was compulsory).

Resources and Support

Satisfaction with advice and support by NSW Health

RECs, Sponsors and SSAC members were all asked to rate their satisfaction with the advice and support provided by NSW Health regarding the SSAS on a Likert scale marked from 1 (very dissatisfied) to 10 (very satisfied). The level of satisfaction was generally high with an average rating of 7 for Sponsors (range 3 to 10), 8 for RECs (range 3 to 10) and 9 for REC members (range 7 to 10).

Two RECs reported low levels of satisfaction (<4). One was concerned with the speed of the SSAS response and the other received the Final Report after the project had been considered by SSAS, pointing out the problem RECs have determining if a project had been submitted to SSAS.

Two Sponsors also reported low levels of satisfaction (<4). One had been informed of SSAS by an REC and felt that "*information regarding this pilot scheme could have been distributed in better/more timely manner*". The other Sponsor was concerned about the expertise of the individuals on SSAS:

"For studies in highly specialised areas, where the SSAC clearly has limited expertise, external experts should be chosen to ensure that the opinion of the body which the expert represents is reflected. This may require opinions from more than one external expert; The selection of external expert(s) should be more transparent and should have input from the sponsor of the study; There should be a clear appeals process. This should be communicated to the sponsor when the decision is forwarded; The sponsor should have access to the minutes of the SSAC meeting relevant to the application; Applications should be considered in a more expedient manner."

Suggested improvements to support structure

RECs appeared to be frustrated when the SSAS Final Report was made available to them after the trial had been considered by the REC. The other issue identified was

the lack of information available to RECs on the progress of trials through the SSAS, particularly reasons for delay.

> *"need more communication between SSAS coordinator and peripheral REC sites and investigators. ie when approvals are delayed"*

> *"companies planning multi-centre studies should submit the protocol to SSAS before it is sent to any RECs"*

> *"SSAS report received after project was considered by REC; REC not always aware if project submitted to SSAS"*

There was one response from a non-participating REC:

> *"most ... multi-centre trials we review are across the states - would be useful to have shared assessment arrangement with the other states (I know - a big ask)"*

The suggested improvements to the support structure made by Sponsors were more varied. Most (but not all) indicated that they had submitted at least one trial to the scheme. Two respondents did not think the scheme had improved the process. Two indicated that they were only made aware of the scheme by an REC and a third suggested that RECs should "*advise at the time of submission of a protocol of the existence of the SSAS*". One Sponsor felt that they questions they had been asked were ambiguous (it is unclear whether these questions related to the Application Form or to correspondence relating to the Final Report), one had difficulty finding the web site, and two Sponsors would have liked to have received more prompt assistance from the Secretariat.

The SSAC members did not suggest any changes to the support structure.

Information provision

Most RECs (16) indicated that they felt adequately informed of the SSAS and its processes (4 did not answer the question). One did not feel adequately informed, feeling unclear about the time frame of the process. One of the adequately informed respondents indicated that they were "*not aware of individuals on the committee or their expertise*" (although this information is on the SSAS web site).

13 RECs found the manual useful and 9 found the web site useful. The remainder either did not find these resources useful or did not answer the question. 14 Sponsors felt adequately informed about the SSAS and its processes (2 did not answer the question). The 5 who did not made the following comments:

> *"inconsistent advice"*
>
> *"It was through one particular REC that I found out about SSAS and I therefore did my own search on internet to obtain additional information"*
>
> *"... the only information I received was at the NHMRC conference. It was also difficult to see the benefits of the scheme except for reducing the burden on smaller ECs"*
>
> *"Appeals process was inadequately explained. It appeared to be non-existent."*
>
> *"It was only through an REC that I was notified of this scheme. I feel that information regarding this pilot scheme could have been distributed in better/more timely manner"*

Time spent by SSAC members preparing for meetings

SSAC members were asked to indicate how much time, an average, they spent preparing for each SSAC meeting. The average time spent preparing was 5 hours (range 2-12 hours; median 4 hours). All 7 members providing feedback indicated that they felt the meetings were held at an acceptable time.

The Final Report

RECs, Sponsors and SSAC members were all asked to consider their experience with SSAS during the Pilot and indicate whether there was anything about the Final Report they would like to see changed. The majority did not think any changes were necessary and only 7 suggested changes (see Table 68). One REC suggesting no change be made to the Final Report made the following comment:

> *"The committee found that the report helped them to quickly expedite the application and prevented them from becoming entrenched in having to seek technical advice about highly complex scientific issues, the committee were therefore more able to quickly assess the application on the ethical issues, in the light of the technical advice from the SSAS."*

Table 68: Suggestions for the final report

	Missing	No	Yes	If yes, changes suggested
REC	5	12	4	Conclusion/summary placed at front of report, recommendations at the beginning A list of members could be useful with their area of expertise. Names of reviewers and their comments could be useful.
Sponsor	7	11	3	Come earlier A more detailed explanation for the decision taken by the SSAC needs to be provided. This should discuss all data taken into consideration in reaching the decision. In our experience, clinical advice that is not yet accepted clinical practice anywhere should not be binding
SSAC	1	6	0	No changes suggested
Total	**13**	**29**	**7**	

Continuation of the SSAS Scheme

Applicants, RECs and SSAS were all asked the same questions regarding continuation of the scheme.

Do you think the scheme should continue?
and
If the scheme continues, are there any changes to the process that should be made?

The majority of responding RECs (16 of 21) and SSAC (6 of 7) members thought the scheme should continue, the remainder being undecided (see Table 69). The majority of responding Sponsors (9) also thought the scheme should continue or were undecided – only 4 thought the scheme should cease. The reasons given for this included a preference for centralisation of the entire ethical review process and a concern that centralised scientific review leads to longer approval times:

> *"Only worthwhile if all EC buy in to the output of SSAC. Otherwise it's just adding another step and slowing whole process down."*

> *"Very easy to use, and if all NSW RECs would accept it, this would give a consistent review across the state as well as potentially speeding the process up, if RECs will accept the scientific evaluation and only focus on other issues such as patient informed consent etc"*

If the scheme continues, what form should it take?
and
If you feel the scheme should continue in an expanded form, in what way do you think it should expand?

Respondents were divided as to the form SSAS should take if it continues in an expanded form with 16 respondents suggesting that it should continue in its current form, 17 in an expanded form and 11 suggesting another form (see Table 70).

There responses were not consistent in identifying any single, preferred way in which the scheme should expand (see Table 71). Some respondents suggested that the committee's role should expand beyond scientific review to provide expert opinion on an ad-hoc basis to RECs, or to full ethical review for multi-centre trials.

Sponsors made the following suggestions:

> *"Would need to continue in a form that was accepted and supported by sites should make sure that RECs are aware of the scheme and will use it"*

> *"After learning more about the scheme and now that more RECs are aware if the scheme also, it will help in improving timelines"*

> *"Replace site by site ethics approvals. Be a true multi-centre ethics review committee"*

> *"Submission to SSAS should be optional"*

Suggestions made by SSAC members included:

> *"I would consider requiring all drug trials with 3 or more centres to go to the SSAC. The exact threshold (3, 4 or 5 centres) should be set to ensure that the SSAC considers no more than 4 new trials per meeting"*

> *"To have an impact this (*all multicentre research*) should be the ultimate goal. However, I'm not convinced that our experience so far tells us that this is possible"*

> *"Probably need 2+ sites for review of protocol; interstate cooperation"*

> *"It would be beneficial if the committee could look at some of the smaller studies being conducted"*

"Provide expert opinion to those RECs who require a second opinion for a single site trial. Some RECs do not have the expertise available to them."

Suggestions made by RECs included:

"Consulting role for "difficult" single-centre studies

"Matters of risk management, hospital drug budgets and embedded research costs are a quasi function and no alternative system exists to monitor such issues"

"Accept referrals when institution feels it des not have expertise to assess a particular multicentre trial"

"Possible use as a quality control function for scientific subcommittee; as a reference panel for RECs with difficult cases to refer for assistance"

Table 69: Do you think the scheme should continue?

	Yes	No	Undecided	Missing
REC	16	0	4	1
Sponsor	9	4	7	1
SSAC	6	0	1	0
Total	**31**	**4**	**12**	**2**

Table 70: If the scheme continues, what form should it take?

	Current form	Expanded form	Other Note 1
REC	6	7	5
Sponsor	8	6	5
SSAC	2	4	1
Total	**16**	**17**	**11**

Note 1: Some of the respondents ticking "other" ticked all boxes.

Table 71: How should the scheme expand?

	All multi-centre drug trials	All multi-centre clinical research	All multi-centre research (including epidemiological)	Expand to all options	Other
REC	1	1	2	6	3 Note 1
Sponsor	3	0	1	0	5 Note 2
SSAC	0	0	1	1	4 Note 3
Total	**4**	**1**	**4**	**7**	**12**

Note 1: 2 suggested all multi-centre drug and device trials; 1 suggested all multi-centre drug trials plus referrals when REC does not have expertise required to assess.

Note 2: 3 Sponsors suggested centralised ethical review; 1 suggested the form taken should be acceptable to sites; and 1 suggested that SSAS should be optional

Note 3: 1 SSAC member suggested multi-centre drug trials with 3 or more centres; 1 suggested looking at smaller studies; 1 suggested all multi-centre drug trials and research as well as providing expert opinion for single-site trials when required by RECs without access to necessary expertise.

Other issues raised

Other issues raised by Sponsors included:

"There is no point in having the system if it just makes doing research in NSW harder / less competitive however I do think a centralised review is a good idea but not at a cost of delaying study start up"

"The Secretariat was fantastic to work with, very approachable and contactable, very helpful with completing the forms and discussing the processes, etc. I realise that it is difficult to ask RECs to alter their practises, however it would be so useful to have a tick-box on all REC application forms to state whether the study has already gone to the SSAC and if so, please attach final report. We sent the report to all sites but found that not all RECs received it (where applicable) so it seemed wasted. Perhaps this is a sponsor/investigator education issue though."

"Patient information sheet - either SSAS rules or should not be involved"

"The SSAS should not necessarily consider all multi-centre trials, as only 1 NSW site may be involved and the usefulness of the SSAC reviewing the trial may not be appropriate as it will tie up resources with little contribution"

"I found the criticisms raised in certain issues to be very doubtful but the concept is very good"

"On attempting to resolve a number of issues raised by the SSAC, the Secretary and Chairman were inadequately informed about the appeals process; Request to provide details of the external expert (qualification, background, etc) was declined. Need more transparency an option for face-to-face meetings between the sponsor and the Chairman."

"Trial submitted by Sponsor- greatly improved approval process"

The issues raised by SSAC include:

"Process is still duplicated by the individual ethics committees"

"We run the risk of another bureaucracy unless we can get "buy in" from Ethics Committees"

"The REC need better access to our comments etc"

"If the role expands, the membership should expand"

Additional questions asked of Applicants

Is the application process easy to understand?

Seventeen applicants indicated that the application process was easy to understand, although one indicated that the appeals process was not. One applicant felt that it is not clear when SSAS should, or should not, be used.

Does the application process involve an acceptable amount of work?

Fifteen applicants indicated that the process involved an acceptable amount of work, and 3 felt it was extra work or duplication. There were 2 missing responses to this question.

"Very easy and user friendly application form"
"Easier than a regular REC"

Was the SSAS consistent in the feedback it provided

Eight applicants felt that the feedback provided by SSAS was consistent. The remaining applicants either did not answer this question or only saw 1 SSAS-reviewed trial and felt unable to comment on consistency. One applicant was concerned about the merit of some of the criticisms.

Additional questions asked of RECs

Would you be prepared to replace your current system of scientific assessment with the SSAS evaluation?

Eleven RECs indicated that they would be prepared to replace their current method of scientific assessment with the SSAS evaluation. Five were uncertain and 4 would not replace their current method. There was one missing response to this question.

The reasons given by most of the committees for their willingness to replace their current scientific assessment with SSAS were:

- The SSAS process was more rigorous and thorough than most RECs were able to perform themselves
- Lack of access to the necessary expertise locally
- Confidence in the SSAC members, and
- Increased efficiency and reduced duplication across RECs

For example:

> *"Multi-centre clinical drug trials are increasingly complex and advice from SSAS on the scientific aspects are invaluable, especially to a committee which due to its rural location would not necessarily have easy access to such invaluable advice. We would for the reasons stated welcome continuance of such a valuable source of technical information and advice."*

Although 9 RECs were uncertain or unwilling to replace their current scientific assessment with SSAS, 3 RECs indicated that this was due to the lack of opportunity to test the scheme (lack of trials that meet the eligibility criteria), and another was unsure what replacing their current assessment would mean. Two RECs were satisfied with their current in-house assessment, and one indicated that the "*process could well be adopted if processing time was faster.*" One committee made the following comment:

> *"REC review forms are one aspect of risk management of research - science review is a necessary aspect of this process. (REC) Scientific review is a key mode of learning for clinical researchers. To devolve, even if in part, this responsibility would equal a loss of research skill. The system of research governance is already deregulated."*

Additional questions asked of SSAC

If SSAS continues, are there any changes to the process that should be made?

SSAC members made the following suggestions for changes to the SSAS process:

> *"It would be great to have 2 expert reviewers. Also we need a neurologist on the committee."*

> *"Most of the issues raised by the SSAC not adequately addressed by companies / sponsors but accepted by SSAC"*

> *"The process continues to be too slow to provide timely feedback to the RECs"*

> *"It is essential that Ethics Committees have the information from the SSAS before they hand out their schedules for their meetings"*

SSAS Evaluation Discussion

Before commencing this evaluation of SSAS it was hypothesised that RECs would find it useful in the ethical review of multi-centre clinical drug trials. Based on the experience of most of the 21 participating RECs this would indeed appear to be the case. Most reported that they found the SSAS Final Report to be useful and that it improved their confidence in their decision making regarding the trial. They also felt that the Final Report reduced the time taken to consider the trial at their meeting. Both RECs and Sponsors thought the scheme should continue in some form, although there were some issues identified that needed to be resolved if their full support was to be obtained.

The first of these issues was the perceived time taken to obtain ethics approval at multiple sites and whether SSAS has a negative impact on this. Unfortunately it was not possible to quantify the impact SSAS has on the time it takes to obtain ethics approval in the Pilot, because Sponsors did not return the log form that would have provided data on the time to an actionable approval document. In addition, the evaluation did not have a concurrent control arm of trials not submitted to SSAS to which the trials submitted to SSAS could be compared.

Caution needs to be taken when interpreting time delays in the scientific and ethical review process as there are many factors that may contribute to a delay in approval. For example, most RECs publicise submission deadlines which potential applicants may miss. As is evident from the trials in this evaluation, at least some delay can be attributed to the Sponsor – in this case delays resulting from the time taken to respond to queries made by SSAS. Insurance, legal and indemnity issues, as well as incomplete applications, are also factors that result in review delays. It is important to note that, unlike RECs in many other countries, Australian RECs carry a significant proportion of the regulatory burden. For example, for clinical trials submitted as part of the Clinical Trial Notification (CTN) scheme, RECs are expected to assess toxicological and safety data (Commonwealth Department of Health and Ageing 1999). Obtaining approval to proceed with the trial from the TGA, once REC approval has been obtained and the CTN form has been signed by all the relevant parties, is a

simple process of forwarding the completed CTN form with a fee to the TGA (Commonwealth Department of Health and Ageing 2001a). By way of contrast, Institutional Review Boards (IRBs) in the United States are expected to consider if the trial protocol is scientifically sound when reviewing drug research but this is "not the IRB's primary concern ... [and] an IRB may rely on the FDA, institutions, scientific review committees, funding agencies (*e.g.*, NIH), or others for this determination" (2001;U.S Food and Drug Administration 2002).

Preconceived notions concerning the value of the SSAS process is a problem that can only be addressed through demonstrating the reliability, usefulness and timeliness of the scheme (or not) over time. It is possible that the impact of SSAS on factors such as the time taken to obtain ethical approval may only be seen once acceptance of the scheme becomes more widespread, and RECs develop trust and confidence in the SSAS process and its product. We should also consider the quality of ethical review and the possibility that pressure to hasten the process, and the workload burden experienced by some committees, could impact on the quality of ethical review. There will be occasions when delay in ethical approval resulting from requests for clarification or change will be appropriate.

It is important to note that, while SSAS may result in an initial delay this may wash out in the longer term as multiple sites may approve trial sooner – initial delay may be misleading.

Moves have already been made to address some of the resource and support issues encountered during the conduct of the pilot. A mechanism for communicating information to RECs about the status of applications made to SSAS has been put into place. There is a need to continue to disseminate information about SSAS to Sponsors although the best way to go about achieving this is open to debate.

There are some limitations in the design of this evaluation. One is the method used to obtain the data used to estimate resource usage. While some data are relatively accurate (eg number of projects reviewed) others will be estimates only, based on the experience of the Executive Officer. The estimated time spent by REC members preparing for meetings, is one example. Although a guesstimate, the data can still

provide us with a best estimate of resource expenditure in the absence of a more detailed resource utilisation study, the latter being beyond the scope of this evaluation.

Another limitation is the poor response rate (47%) regarding the return of Evaluation Forms on individual trials by RECs. By cross-referencing the End of Study Forms with the Evaluation Forms we know that at least two of the missing EFs relate to trials where the SSAS Final Report was received after the REC meeting, and that the RECs of about one third of the missing EFs indicated that they would be happy to replace their current system of scientific review with SSAS.

A methodological challenge was identifying the best way to obtain valid and useful information about aspects of decision-making when the entity making that decision is a group of individuals, and considering the restricted access by the investigator to those entities. While RECs were asked to discuss the questions as a group, the forms were completed by individuals and it is therefore inevitable that some responses may not truly represent the opinions of the entire group. Another limitation, particularly in the estimation of workload, is that RECs were asked to make some "best guesses" rather than conduct detailed time-and-motion studies.

Even though the majority of RECs felt adequately informed, and most were satisfied with the support provided by the Department of Health, it is evident that there were some problems experienced during the pilot regarding the communication of information relating to the trials being considered by SSAS, and the expected timeframes. Surprisingly few seemed to have access the manual or the SSAS web site and reasons for this need to be explored further. The SSAS web site could be a valuable tool in the dissemination of information to RECs, but only if RECs have the ability and the inclination to visit the site.

Based on the findings of the End of Study survey, there was clear support for SSAS to continue beyond the initial 12 month pilot. It was recommended that further evaluation be carried out in order to:

- Increase the number of studies submitted to SSAS and hence available for evaluation.

- Investigate the potential impact of expanding SSAS to include all multi-centre trials or all multicentre research. This may require the formation of more than one committee.
- Obtain data on the time taken to obtain REC approval, perhaps obtaining this information from RECs rather than Sponsors.
- Explore further the other factors that may contribute to delay in ethical approval, particularly regulatory, legal and indemnity issues.
- Explore the best way in which to communicate information to RECs about SSAS and trials submitted to SSAS, and identify possible barriers to communication.

It was also recommended that mechanisms for disseminating information about SSAS to Sponsors be investigated, as it was evident that information was not filtering through the relevant people within each organisation.

An issue facing RECs and SSAS is the need to share information in a way that does not divulge aspects of a trial that are considered to be confidential, but at the same time enables RECs to be aware of trials that have been submitted to SSAS. This would save resources by avoiding unnecessary duplication across RECs. It would also be useful if RECs could check the status of a particular application, thereby helping them to anticipate submissions and effectively manage their workload.

In conclusion, there would appear to be a potentially valuable role for centralised scientific review of multi-centre clinical research. It could improve the consistency of decision-making across ethics committees by ensuring an underlying minimum scientific standard, as well as ensuring that individual ethics committee's meet their ethical obligation to consider the science of a trial in its deliberations. At the same time, central scientific review allows individual RECs to maintain their autonomy and to focus on their primary task: that is, to ensure the conduct of ethical clinical research in the context of the community they represent and serve. Evaluation is required to determine whether centralised scientific review has an impact on the quality of research, as well as whether it unreasonably extends the time required to obtain ethics approval.

Chapter 8: Thesis overview, discussion and conclusions

As I argued at the start of this thesis, there is an ethical obligation on the part of all those associated with clinical trials research to ensure the methodological rigour and scientific integrity of this research. There is also an obligation to ensure that the results are made publicly available in a manner that is not misleading and does not misrepresent the data. This will ensure that the best quality research is available as the basis for health care decisions.

Quality and randomised controlled trials

Interest in the quality of randomised clinical trial research is not restricted to those directly involved in their design and conduct. A government body or charitable organisation funding a clinical trial will want to ensure that their money is being invested wisely in a trial that is of direct relevance to health care practice, has a reasonable chance of reaching a valid conclusion, with its findings being made available to people making decisions about health care preactice. A pharmaceutical sponsor will also want to comply with the appropriate regulatory requirements, particularly ICH GCP. Institutional ethics committees in their role as patient representatives, need to ensure that the patients on whose behalf they act are participating in research that is ethically sound.

The concept of "quality" for a randomised clinical trial is therefore multi-dimensional, and the emphasis given to any one dimension will depend on the viewpoint of the individual or group. Dimensions of quality include:

- Methodology
- Relevance
- Conduct
- Reporting

Methodology and Reporting

This thesis concentrated on the quality of methodology by considering the design of trials, and the quality of reporting by considering the content of published articles. It

is composed of 3 substantial studies, the first of which identified and quantified issues relevant to trial quality by comparing the protocol and related document lodged with an ethics committee with resulting publications (Chapters 2 to 5). This helped to highlight areas in both methodology and reporting where there is room for improvement. The second study investigated the published experiences with shared scientific or ethical review of multi-centre research in an attempt to evaluate the impact this has on the quality of clinical trials and of decision making. The study found that evidence of the effectiveness or not of these systems was lacking. The third study addressed some of these deficiencies, evaluating the impact of a shared scientific assessment scheme implemented in NSW had on the functioning and decision-making of human research ethics committees.

Key factors which have been shown to have a direct impact on the quality of randomised clinical trials were all poorly documented in the protocols of trials in my study, and the content of a protocol was not always consistent with the information subsequently reported in the publication. These factors include the methods used to generate the sequence in which interventions were allocated, allocation concealment, the methods used to blind interventions including inadequate descriptions of placebos, and the sample size calculation. All of these factors should be considered a fundamental component of any decision to approve a trial made by an ethics committee, but it is possible that at least some investigators consider them to be more "administrative" issues that relate to the conduct but not the science of the trial.

My follow-up study demonstrated that selective reporting existed in some form in a significant proportion of the included trials. This selective reporting included:

- selection of which outcomes were reported (discrepancy in identity), and
- selection of the amount of information reported for an outcome (completeness of reporting)
- selection of how an outcome is defined (discrepancy in definition)

It is to be expected that there will be some differences between the protocol and the publication given that trials usually occur over prolonged periods (sometimes several years). For example, the results of other research may become available while a trial

is ongoing, which might require justifiable alterations in one or more aspects its design or conduct. A new, more sensitive technique for measuring the outcome of interest may become available, again requiring an alteration to the trial protocol. There could, however, be more insidious reasons for changes to the methodology and selective reporting. These include deliberate manipulation of one or more aspects of a trial to allow it to be portrayed in a manner that authors may perceive to be more appealing to journal editors or readers, possibly making an intervention look more or less effective than it really is. The main cause for concern is when any differences and the reasons for them are not mentioned in the publication, because the end user of the trial publication is then unable to make an informed judgement about the validity of the changes in design of study itself, or its relevance to their decision-making needs.

Relevance

The perfectly designed trial that is seeking a conclusive answer to a question that will not be of interest in practice could be considered to be of poor "quality". It is also possible that such trials would also have difficulties recruiting participants. The concept of relevance is important in any one (or all) of the components of a question: that is, the patient population, the intervention and comparator, and the outcome. Examples of problems would be eligibility criteria that exclude those participants who would benefit most from the identification of an effective new intervention; the choice of a comparator that is not accepted standard practice; or an outcome measure using a technique that is not widely available.

There are also societal consequences of the conduct of irrelevant trials. The resources consumed by such a trial could have been used more efficiently and to greater benefit elsewhere. Some may also question the ethics of recruiting patients onto a randomised trial if the results of that trial are unlikely to be incorporated into practice. The same ethical question could be asked of trials that are unable to recruit a sufficient number of patients to answer the question posed, trials with a sample size calculation that considerably over (or under) estimated the potential size of the treatment effect, or trials based on an outcome that is not clinically relevant.

Conduct

A range of rules and regulations exist that can (and in some cases must) be applied when clinical trials research is being conducted. The most widely known are the International Conference on Harmonisation (ICH) Guidelines for Good Clinical Practice (GCP). These guidelines recommend minimum standards of practice for the various aspects of the conduct of a clinical trial, and they include recommendations for the trial protocol. The recommendations made by GCP are not mandatory, but regulatory agencies and some ethics committees may insist on compliance with the guidelines if a trial is to be approved.

Guidelines such as GCP have an important role in establishing a benchmark for trials. While such guidelines can be important and relevant, their impact on the any aspect of the quality of clinical trials research is not known. The value of undertaking some tasks has certainly been questioned, particularly given the cost entailed in their achievement. Concerns have been expressed, in particular, by academic researchers in relation to the potentially negative impact of the European Commission Directive 2001/20/EC on clinical trials introduced in May 2004. Many were concerned at the time, and this concern remains today that the Directive would have a negative impact on publicly funded clinical trials research in Europe. The intention of the directive would appear to be to "simplify and harmonise the administrative provisions governing such trials by establishing a clear, transparent procedure and creating conditions conducive to effective coordination of such clinical trials in the Community". (European Commission. 2001) The fear was that the "labour intensive, bureaucratic and expensive endeavour of running a clinical trial would become worse", and that only the commercial sector would have the resources to conduct trials. (Hemminki & Kellokumpu-Lehtinen 2006) In a recent BMJ editorial it was reported that, since the introduction of the Directive, the number of new trials conducted by a large independent European cancer trials organisation had fallen, fewer patients had been enrolled, costs had increased (including insurance costs), and trial initiation had slowed, "mostly the result of the increased workload of ethics committees". (Hemminki & Kellokumpu-Lehtinen 2006)

While guidelines such as GCP are intended to ensure the quality of the conduct of clinical trials, it is arguable that improving conduct results in an improvement in the

other (arguably more important) aspects of trial quality. In at least some cases it would appear that overly onerous regulation of conduct has a negative impact on the quality of clinical trials, specifically on relevance, by making it more difficult to initiate and successfully complete publicly funded (investigator initiated) studies.

Who is responsible for trial quality?

Responsibility for all aspects of a clinical trial should remain with the trial investigators, who are ultimately accountable to the trial sponsors (financial and non-financial) and to the people who consent to participate. The effectiveness of self-regulation is debatable though, and there needs to be a third party willing to act on behalf of the community to ensure the integrity of research is maintained. The obvious agencies to take on this responsibility would be research ethics committees.

Although the scientific validity of a clinical trial is clearly pertinent to its ethical integrity, at least some ethics committees admit they do not have the skills necessary to evaluate adequately the science of the trials they are asked to assess. Modern ethics committees also have daunting workloads and are under significant pressure to deliver expedient decisions within tight timeframes.

Many modern trials are multi-centre (68% of trials in the follow-up study in this thesis) and will therefore be reviewed by more than one ethics committee. It would therefore seem logical that one way to reduce the workload of ethics committees would be to reduce unnecessary repetition, particularly the repetition of those tasks they consider themselves to be ill-qualified to conduct effectively. A mechanism that would allow non-local tasks to be performed centrally, leaving local committees to consider local issues, is therefore appealing – although experience with such mechanisms has been variable, as I have shown in my systematic review in Chapter 6. As indicated by this review, the studies that were identified provide insufficient evidence to determine whether centralised scientific or ethical review improves the quality of trials or the quality of decisions made by RECs. However, the pilot of a shared scientific review in NSW, described in Chapter 7, does suggest that there is certainly potential for such systems to lead to improvements, but there is clearly more that needs to be learned.

There is potentially a greater role for health care journals in ensuring the quality of the clinical trials research they publish. The CONSORT statement and its implementation by many key journals has already gone some way to improve the quality of reporting, although this in itself will not address the issue of selective reporting (CONSORT Group 1996). Some journals (such as BioMed Central) encourage trial investigators to publish their full protocols (in the case of BioMed Central this is online publication). One journal editor has suggested that publication of the protocol "allows reviewers and readers to suggest improvements to be made to the study before it begins"(Godlee F 2001).

By demonstrating a commitment to the prospective registration of trials, journals affiliated with the International Committee of Medical Journal Editors (ICMJE) have increased the acceptance and utilisation of the existing registers, forcing trial investigators to register minimum information about their trials (at inception) as a precondition for publication. The Australian Clinical Trials Registry, for example, saw the number of trials submitted for registration increase from 346 to 651 during the 2 weeks period immediately prior to the deadline for registration set by the ICMJE of 13th September 2006 if the trial was to eligible for future publication in these journals (Australian Clinical Trials Registry 2005).

Some journals have taken this commitment a step further. The British Medical Journal (BMJ), for example, requires authors of manuscripts reporting the results of randomised trials to not only comply with the CONSORT statement, but also to submit their protocol with the manuscript, so the editors and reviews can refer to the former if necessary as part of the peer review process. (Jones & Abbasi 2004) The belief is "that identifying deviation from the protocol is another important step in ensuring that the findings of a study are reported with honesty and transparency". Unfortunately, the BMJ has found that authors are reluctant to provide their protocols, but the reasons for this reluctance do not appear to have been reported.

What can be done to improve quality?

There are many opportunities that could be taken advantage of to improve the quality of clinical trials, at all stages of the research process, involving all of those with a vested interest in clinical trials research. The first is to improve the expertise of the

principle investigators for trials, as well as those directly involved in the day-to-day conduct of research, perhaps restricting the ability to undertake clinical trials to those who are "certified" to do so.

To maximise the relevance of trials, barriers to investigator-initiated clinical trials research need to minimised, while at the same time ensuring the methodological quality of that research. Investigator access to methodological and biostatistical expertise could be improved, starting with raising the profile (and recognition) of the importance of this type of expertise in the development of a trial protocol. This would require the allocation of funding to allow investigators to procure this expertise.

As discussed in Chapters 6 and 7, quality could also be addressed by improving REC access to the necessary expertise (such as shared scientific review) as well as by improving the quality of data reported in published manuscripts (through consideration of protocol at time of manuscript submission, for example).

Further research

If some aspects of ethical review (including scientific review) for multi-centre clinical research could be shared or concentrated in a single committee, this should lead to improvements in trial quality and facilitate the whole process of clinical research. The effectiveness of centralised review has not been adequately evaluated, as was shown in Chapter 6. The best way to do this would be through a randomised trial in which individual ethics committees (as a 'cluster') would be randomised to either continue as normal (that is, the entire ethical review process is conducted locally for all trials) or to a process of local review following central review. Randomising individual trials for consideration, or not, by the central committee would be cumbersome and not practicable. There is a possibility of "contamination" across RECs who may want to share information but measures could be taken to minimise the problems this might cause. Trial investigators, for example, would need to be kept unaware of the identity of the "intervention" to which each ethics committee had been allocated. The intervention in the trial outlined here would need to be well defined, as would the distinction between the role of the central committee and that of the local committee. Outcomes of interest would include the quality of the trial protocol after it has been

approved at a local level, the time taken to obtain approval and the costs involved in the approval process.

The impact of making the original protocol available to journal editors and referees could also be evaluated in a randomised fashion. If a condition of submission is that the protocol be provided along with the manuscript, trials could then be randomly allocated so that some referees would receive the protocol along with the manuscript and others would only receive the manuscript. The instructions given to the referee regarding the use of the protocol would need to be defined, and could be either prescriptive or pragmatic (possibly no instruction at all). If differences between the protocol and the manuscript were detected, authors could then be asked to address these differences in their manuscripts, thereby improving the quality of the report. The outcomes would include the quality of the published article and resource implications such as the time taken for the peer-review process.

Research in selective reporting and forms of publication bias has largely focused on randomised trials. The problem is probably worse for observational studies but there is a dearth of research in this area (Godlee F 2001). There is a need for further research in order to describe and quantify the selective reporting of observational studies, particularly given their increased use in systematic reviews in areas where randomised trials are not possible or feasible.

Conclusion

Trial investigators have a scientific and a moral responsibility to ensure that a true and accurate representation of their trial is made available to the ethics committees who are being asked to approve the research, to the people who are invited to participate in that research, and to the readers of any report produced as a result. This will ensure that health care decisions are made based on research of the highest quality - in all its dimensions.

Appendices

Appendix 1: Expectations of a Human Research Ethics Committee regarding the science of a clinical trial

The following has been extracted from the Human Research Ethics Handbook ((NHMRC 2002))

RECs need to be satisfied that the research design can produce valid results and can protect the welfare and rights of research participants. To be satisfied, an REC may seek or receive advice from an individual, a scientific committee in its institution, an external expert, or it may include an *additional* person who has specific expertise in the particular type of research. It is not possible to provide a comprehensive list of relevant considerations for every research approval. However, the following matters will usually require consideration:

The project

- Is there a clear hypothesis?
- Is the research question useful? Is the research worthwhile?
- Is the research likely to yield new information, enhance understanding or clarify existing uncertainty?
- Has this, or similar, research been carried out before in the same, or similar, contexts?
- Can the research proposal be supported by a systematic review of the literature that would demonstrate the importance of the research question and that it builds upon the results of previous research?
- If indicated, have perspectives of potential participant groups, the wider community, or other disciplines been incorporated into the research proposal?
- Are the aims of the proposal clear?
- Does the value of the project appear to be adequate to justify its conduct with humans?

Research methodology

- Are all aspects of the research methodology clearly described?
- Is the REC satisfied that the methodology is appropriate to the achievement of the aims of the project?

NS 12.2
An REC must consider all aspects of the design of a clinical trial and be satisfied that:
(a) the trial is directed to answering a specific question or questions;
(b) there is a scientifically valid hypothesis being tested which offers a realistic possibility that the interventions being studied will be at least as effective as standard treatment;
(c) where the research is therapeutic, and is therefore intended and likely to be of direct benefit to participants, there is an acceptable balance between the risks and benefits of the trial;
(d) the methodology provides:
    (i) a rationale for the selection of appropriate participants;
    (ii) an appropriate method of recruitment;

(iii) adequate, understandable information for the purpose of obtaining participant consent;
(iv) a clear description of the intervention and observation to be conducted;
(v) a sample size adequate to demonstrate clinically and statistically significant effects;

(e) it has access to adequate expertise or advice to consider the safety of the drugs, medical devices or other intervention under investigation; and
(f) it is familiar with the requirements of the Therapeutic Goods Administration (TGA) in relation to unregistered drugs and devices, particularly the Clinical Trial Notification (CTN) and Clinical Trial Exemption (CTX) schemes, where relevant.

12.2(a)
The complexity of design in some clinical trials does not relieve an REC of the need to be satisfied that a research question is identified and set out clearly in information sheets for participants.

12.2(b)
The study design should be appropriate to the clinical question being asked.

12.2(c)
Risks may not be confined to adverse effects on physical health but could extend to emotional, economic and other types of disadvantage. In seeking to establish whether benefits and risks associated with a trial are acceptably balanced, an REC should consider whether such a balance is struck not only at the level of the entire participant population but also for individual participants. For instance, planning of a trial should seek to avoid situations in which the likelihood of benefits is predictably higher than average for one identifiable group of participants whilst risks are predictably higher for another.

12.2(d) (i) and (ii)
The proposal should clearly identify how the classes of participants have been selected so as to permit the best extrapolation of trial results to the patient groups to whom the new treatment, if successful, is to be administered.

12.2(d) (iii)
The REC should scrutinise the participant information statement carefully. Medical terminology and abbreviations should be avoided or explained in plain language. The document should clearly explain the purpose of the trial and give a detailed account of the nature of interventions and procedures to be employed, as well as any risks involved. The latter include possible effects of drugs, medical devices or any changes to existing therapies, as well as the interventions to be used in assessing these effects.

Where applicable, it should be stated that interventions will be randomly assigned and that participants may receive inactive or unproven interventions. Special care should be taken with vulnerable participants, such as those with incurable diseases, who may be particularly disposed to try new therapies, as well as those with whom communication is difficult. Potential participants should be informed about available alternative treatments and advised that they may discontinue participation in the study at any time without prejudice to their ongoing medical treatment.

The REC should also consider the ongoing availability of a drug that is proposed for incorporation in a clinical trial. It would be reasonable for the REC to seek assurances from the sponsor that, in the absence of observation of detrimental side effects or of inefficacy, the drug will remain available until the trial is completed. Details should also be provided about:

- compensation and treatment available in the event of trial-related injury to the participant;
- issues of confidentiality;
- contact details in case of emergency; and
- the name of an independent person with whom concerns about the study could be discussed.

12.2(d) (iv)

It should be clear to an REC precisely what the amounts and frequency of dosages of the drug will be, and the kinds and frequency of tests or monitoring involving hospital or clinic attendance that will be required.

Dosages of unmarketed drugs should be based upon pre-clinical and early phase clinical trial data. Care should be taken to ensure that these are consistent with earlier studies. Often, the duration of therapy is much longer in later phase studies than in earlier ones. If the proposed treatment period is significantly longer than those for which data exist interim safety reports will be necessary and should be built into the study design.

12.2(d)(v)

A justification of the proposed sample size, based on the primary endpoint of the study, should be provided. Details should be given about expected clinically important differences between the test and control therapies and the expected variability of the outcome variables. Calculations of the required sample size based on such information are referred to as 'power calculations'. The sample size required for the conduct of any comparative study is directly proportional to the 'power' of the study and the natural variation in the outcome of interest in the population, and inversely proportional to the size of the difference the researcher wishes to detect.

An ethics committee should be satisfied, usually on the basis of expert opinion, that a clinical trial design indicates that the trial can reliably show a reasonable comparative benefit in relation to the new drug or device simply because sufficient participants are to be studied.

12.2(e)

RECs need to receive competent advice about the scientific details of a proposed project. This includes all of the items listed in NS 12.2(d), that is, the protocol, study design, inclusion and exclusion criteria, endpoints and outcome measures, sample size, dosages and duration of therapy, and methods for analysing results. When seeking advice from non-members, an REC should ensure that confidentiality about all aspects of the proposal is preserved and that the intellectual property of the sponsor is not jeopardised (see also NS 2.19–2.20 on the avoidance of conflicts of interest).

Appendix 2: Australian RECs and monitoring responsibilities

From National Statement on Ethical Conduct in Research Involving Humans (NHMRC 2001)

12.8 An institution or organisation and its REC must require the researcher:
(a) to conduct the trial in compliance with the approved protocol;
(b) to provide reports of the progress of the trial to the REC at a frequency directed by the REC that is related to the degree of risk to participants, but at least annually;
(c) to inform the REC of, and seek its approval of, amendments to the protocol including any:
(i) proposed or undertaken in order to eliminate immediate hazards to participants;
(ii) that may increase the risks to participants; or
(iii) that significantly affect the conduct of the trial;

(d) to inform the REC and the TGA of all serious or unexpected adverse events that occur during the trial and may affect the conduct of the trial or the safety of the participants or their willingness to continue participation in the trial;

(e) to inform the REC as soon as possible of any new information from other published or unpublished studies which may have an impact on the continued ethical acceptability of the trial or which may indicate the need for amendments to the trial protocol;

(f) to inform the REC, giving reasons, if the trial is discontinued before the expected date of completion; and

(g) in relation to trials with implantable medical devices, to confirm the existence of or establish a system for tracking the participant, with consent, for the lifetime of the device, and to report any device incidents to the TGA.

12.9 The institution or organisation and its REC must determine the type and frequency of review appropriate to the drug or device being investigated and to the degree of risk to participants provided that the review occurs at least once a year.

12.10 It may be unethical for a researcher to continue a trial if:

(a) there are or have been substantial deviations from the trial protocol;

(b) side effects of unexpected type, severity, or frequency are encountered; or

(c) as the trial progresses, one of several treatments or procedures being compared proves to be so much better, or worse, than other(s) that continuation of the trial would disadvantage some of the participants.

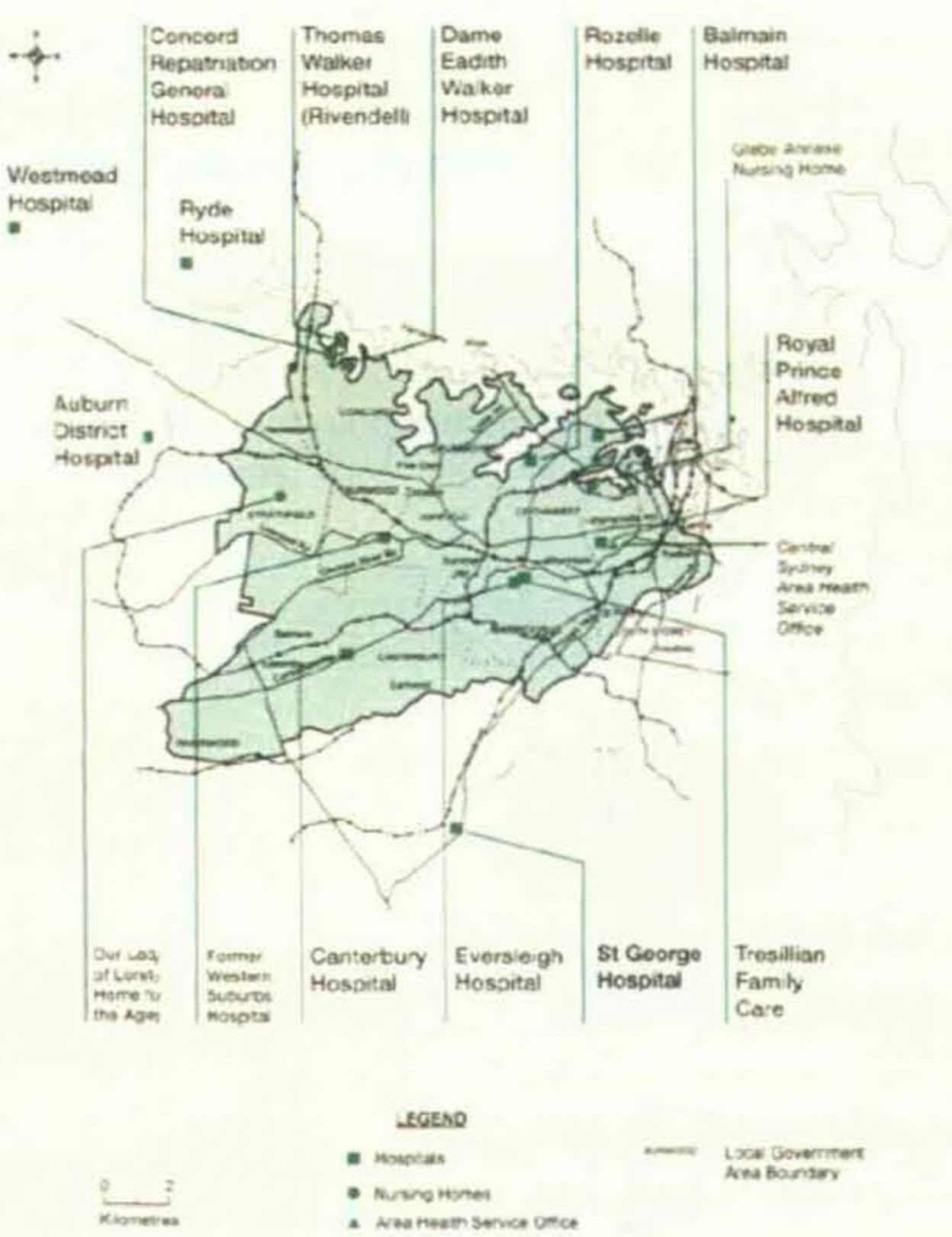
Appendix 3: About CSAHS REC

Extracted from the web site:
http://www.cs.nsw.gov.au/rpa/Research/ethics/default.htm

The CSAHS Ethics Committee (RPAH Zone) reviews research conducted at:

- Royal Prince Alfred Hospital
- Canterbury Hospital
- Balmain Hospital
- Rozelle Hospital
- Institute of Forensic Medicine
- Sydney IVF Pty Ltd
- CSAHS Division of Population Health
- Institute of Respiratory Medicine
- Heart Research Institute
- Centenary Institute of Cancer Medicine & Cell Biology

CSAHS Area Map



The ERC meets each month and, over an average year, considers more than 300 new protocols as well as reviewing progress and compliance of all currently approved studies.

In undertaking an ethical assessment of a proposed project, the ERC considers a number of issues: Is the purpose of the study such that it will usefully advance medical/scientific knowledge? Has it been designed so that a valid conclusion will be reached? What procedures will subjects undergo? Are they unnecessarily painful, arduous, risky or time-consuming? Are questionnaires phrased in such a way that they do not cause anxiety or alarm to subjects? Is any undue inducement being offered to encourage prospective subjects to participate? Have the Subject Information Statement and Consent Form been prepared in clear, concise, plain language giving full details of the procedures, risks and benefits which the subject will face if he/she agrees to participate? Have suitable arrangements been made to ensure that subjects of non-English speaking backgrounds also have the opportunity to take part in the study and to give informed consent to their participation?

Research studies which involve the clinical trial of drugs or devices are assessed by the **Clinical Trials Sub-committee** of the ERC, which also meets monthly. This sub-committee comprises senior clinicians and clinical academics, and has the responsibility for reviewing all the scientific data (such as toxicology, pharmacology and previous clinical experience) on new products to ensure that the expected benefits to subjects outweigh the possible side effects. The sub-committee then forwards the protocol, together with its advice and recommendations, to the ERC for its further consideration.

Clinical Trials Sub-committee

Members of the Clinical Trials Sub-committee have expertise in a wide range of disciplines relevant to clinical trials. All proposed studies involving drugs and devices are reviewed by this Sub-committee before consideration by the Ethics Review Committee. Issues reviewed include safety, technical aspects and scientific validity. If necessary, the Sub-committee can request evaluation by independent (internal or external) assessors.

Appendix 4: Completeness of reporting of primary outcomes

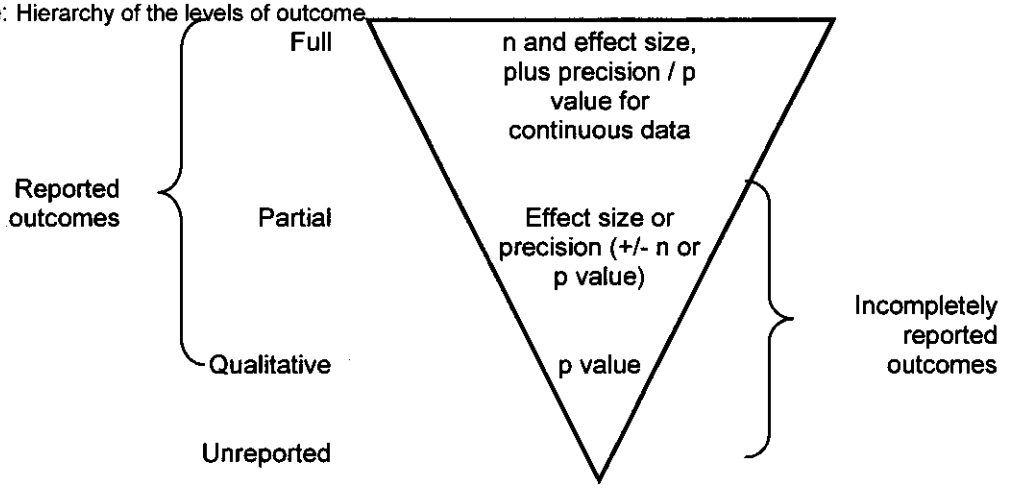
Chan *et al* defined a fully reported outcome as one with enough data for meta-analysis ((Chan AW & Altman 2003)). That is, there was sufficient data to:

- derive the sample sizes per group,
- the effect size, and
- a measure of precision/variability (for continuous outcomes)
    - The standard error could also be derived if sample sizes, effect size, and precise p-value are available.

Incompletely reported outcomes had insufficient data for meta-analysis and were classified:

- partially reported (some data provided),
- qualitatively reported (only the p-value or statement about statistical significance was provided), and
- unreported (no data given).

Figure: Hierarchy of the levels of outcome



Note: reporting (n = number of participants per group)

Table: Amount of data required for meta-analysis of fully reported outcomes (adapted from Chan *et al*)

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

Appendix 5: Data collection forms (follow-up study)

Data collection from the REC file

HREC ID:

From the HREC file

1. Was any funding available to support the conduct of the trial at the time of submission to the HREC?
   - ❑ No funding mentioned
   - ❑ Government agency
   - ❑ Public charity or organisation
   - ❑ Commercial sector
   - ❑ Other (*sp*)

2. Date of actionable approval document: ____ / ____ / _____
3. Date of the last annual report ____ / ____ / _____
4. What was the status of the trial at the time of the last annual report?
   - ❑ Completed (completion date ____ / ____ / _____
   - ❑ In progress
   - ❑ Not yet commenced
   - ❑ Abandoned
   - ❑ Other

5. Is there at least one clearly distinguishable primary outcome in the protocol?
   - ❑ Yes
   - ❑ No

Randomisation

6. Is the trial randomised? (see definition)
   - ❑ Truly randomized
   - ❑ Quasi-randomised
   - ❑ Not randomized
   - ❑ Not assessable
7. Is allocation concealment adequate?
   - ❑ Yes
   - ❑ No
   - ❑ Unclear
8. Is sequence generation adequate?
   - ❑ Yes
   - ❑ No
   - ❑ Unclear
9. Does randomization occur at a time as close to the commencement of treatment as possible?
   - ❑ Yes
   - ❑ No
   - ❑ Unclear

Sample size

10. Is there a target sample size?
    - ❑ Yes, calculated using an appropriate mathematical formula
    - ❑ Yes, but no evidence of a sample size calculation
    - ❑ No

11. Are adequate details of the sample size calculation recorded?	Yes	No
a. the outcome used to calculate sample size	❑	❑ _____
b. the expected treatment effect	❑	❑ _____
c. α error pre-specified	❑	❑
d. β error pre-specified	❑	❑
e. The alternative treatment hypothesis	❑	❑

Use of blinding and placebo

12. The trial is:
    - ❑ Open label
    - ❑ Patient is blinded
      - i. Is it possible for the patient to identify their treatment?
    - ❑ Person administering treatment is blinded
      - ii. Is it possible for the practitioner to identify the treatment?

13. Is a placebo being used?
    - ❑ Yes
    - ❑ No
    - ❑ Unclear

a. If yes, is there a description of the placebo?
   - ❑ Yes
   - ❑ No
   - ❑ Unclear

b. Is the description of the placebo adequate?
   - ❑ Yes
   - ❑ No
   - ❑ Unclear

Adverse events and withdrawals

14. Does the protocol state that adverse events would be monitored (routine laboratory tests)?
    - ❑ Yes
    - ❑ No

15. Does the protocol include a section on the handling of withdrawals (not as an outcome)?
    - ❑ Yes
    - ❑ No

Note: for Question 1, up to 2 responses were collected. If more than 2, the 2 major types of funder were collected.
See glossary for clarification of terms.

Data collection from the publication

HREC ID:

From the publication

Publication characteristics

1. Journal type
   - ❑ General
   - ❑ Specialty
2. Type of report
   - ❑ full publication
   - ❑ short report
   - ❑ letter
   - ❑ other
3. Declared funding (tick *all that apply*)
   - ❑ No funding mentioned
   - ❑ Government agency (eg NHMRC, Department of Health)
   - ❑ Public charity or organisation (eg National Heart Foundation, Cancer Council)
   - ❑ Commercial sector (including pharmaceutical or device industry)
   - ❑ Other (please specify):

Trial characteristics

4. The number of subjects randomised: _____ *(record number)*
5. The reporting of the power calculation is:
   - ❑ adequate
   - ❑ inadequate
6. The number of trial arms is: _____ *(record number)*

7. Patient population	8. Arm 1 *(control / standard / comparator)*	9. Arm 2	10. Arm 3	11. Arm 4

12. The trial was reported is
    - ❑ Single centre
    - ❑ Multicentre
13. and
    - ❑ National
    - ❑ International
14. Reported study design
    - ❑ Parallel
    - ❑ Factorial
    - ❑ Crossover
    - ❑ Equivalence
    - ❑ Cluster
    - ❑ Other (please specify)
15. The purpose of the trial is:
    - ❑ Exploratory
    - ❑ Pragmatic
16. Is a placebo used?
    - ❑ Yes
    - ❑ No
17. If a placebo is used, the description is:
    - ❑ Adequate
    - ❑ Inadequate
    - ❑ Unclear
18. If no placebo, is masking used?
    - ❑ Yes
    - ❑ No
19. If masked, the description of masking is:
    - ❑ Adequate
    - ❑ Inadequate
    - ❑ Unclear
20. Handling of attrition
    - ❑ Adequate
    - ❑ Inadequate
    - ❑ Unclear

Reporting of randomisation

21. The description of sequence generation:
    - ❑ Adequate
    - ❑ Inadequate
    - ❑ Unclear
22. The description of allocation concealment:
    - ❑ Adequate
    - ❑ Inadequate
    - ❑ Unclear
23. Is there at least one clearly distinguishable primary outcome in the publication
    - ❑ Yes
    - ❑ No

Data collection on primary outcome

HREC ID:

Primary Outcome (number _____ of _____)

1. Name:

2. How can you tell it is the primary outcome?

	In protocol	In publication
Clearly stated	☐	☐
Reasonably inferred	☐	☐
Used to calculate sample size	☐	☐
Stated in aims or objectives	☐	☐
Other	☐	☐
Not stated	☐	☐

3. Outcome declaration

Outcome in protocol	Outcome in publication
☐ Y, 1^0	☐ Y, 1^0
☐ Y, 2^0	☐ Y, 2^0
☐ Y, unsp	☐ Y, unsp
☐ N	☐ N

4. Outcome type
   - ☐ Safety
   - ☐ Efficacy
   - ☐ Other / Unclear

5. Defined in protocol as

6. Defined in publication as

7. Are the definitions the same?
   - ☐ Yes
   - ☐ Unable to judge
   - ☐ No – minor change
   - ☐ No – major change

8. If No, how they are different

9. Was the change in definition reported in the publication?
   - ☐ Yes
   - ☐ No

10. Was the outcome measured across multiple time points?
    - ☐ Yes
    - ☐ No

11. Data type reported as:
    - ☐ Continuous
    - ☐ Binary
    - ☐ Categorical
    - ☐ Time to event

12. Elements reported:
    - ☐ n/N
    - ☐ effect size
    - ☐ measure of precision/variability
    - ☐ precise p value
    - ☐ Other (*sp*)

13. What is the p value:

14. Completeness of reporting: meta-analyst
    - ☐ Fully reported
    - ☐ Partially reported
    - ☐ Qualitative
    - ☐ Named but no data reported
    - ☐ Not reported

15. Is this primary outcome mentioned in the abstract?
    - ☐ Yes
    - ☐ No

16. Is this primary outcome used as a basis for any conclusions?
    - ☐ Yes
    - ☐ No

Data collection: Intention to treat and exclusions

Intention to treat

- ☐ states ITT
- ☐ does not state ITT

Exclusions

- ☐ A. Explicitly reported no exclusions
  - Eg stated "intention to treat" and evidence provided
  - Eg explicitly stated that there were no deviations from random allocation
- ☐ B. Gave impression that no exclusions had taken place
  - Eg stated "intention to treat" but no evidence provided
  - Appeared to analyse as randomised
    - Explicitly reported analysed according to random allocation
    - Did not explicitly report analysed according to random allocation
- ☐ C. Explicitly reported that there were exclusions
  - Report number excluded by treatment arm
  - Report that entry criteria applied identically to each group (eg committee revising all eligible cases)
  - Did not analyse as randomised
- ☐ D. Exclusions not mentioned
- ☐ E. Other (details documented)

Reasons for exclusions:
(If there were exclusions / did not analyse as randomised, did they exclude☺

- ☐ Patients who did not start the allocated intervention
- ☐ Non-compliers
- ☐ False inclusions (ineligible)
- ☐ Other
- ☐ Not applicable

Appendix 6: Univariate analyses

Appendix 6.1: Univariate analyses for discrepancy in the identity of the primary outcome

Note: These analyses are based on the 97 trials with at least one identifiable primary outcome.

Appendix 6.1.1: Threshold for discrepancy in the identity of the primary outcome = 100%

Proportion of trials with a discrepancy in the identity of a primary outcome		0: <100% 1: 100%

Design

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	designcode	-.034	.742	.002	1	.964	.967	.226	4.139
	Constant	.693	.707	.961	1	.327	2.000		

Purpose

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	purposecode	.227	.511	.198	1	.657	1.255	.461	3.420
	Constant	.486	.449	1.167	1	.280	1.625		

Administration

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	administration			2.718	2	.257			
	administration(1)	.303	.637	.226	1	.634	1.354	.388	4.722
	administration(2)	.791	.487	2.637	1	.104	2.205	.849	5.728
	Constant	.208	.373	.309	1	.578	1.231		

Commercial funding (protocol)

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECdrugfundcode	.999	.446	5.024	1	.025	2.715	1.134	6.504
	Constant	.057	.338	.029	1	.866	1.059		

Proposed sample size

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECss200code	.750	.436	2.966	1	.085	2.117	.902	4.972
	Constant	.274	.304	.813	1	.367	1.316		

Completeness of the sample size calculation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	sscompcode2			9.821	2	.007			
	sscompcode2(1)	1.977	.823	5.775	1	.016	7.222	1.440	36.224
	sscompcode2(2)	2.242	.716	9.798	1	.002	9.412	2.312	38.312
	Constant	-1.204	.658	3.345	1	.067	.300		

Allocation concealment

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECalloccode	.306	.436	.492	1	.483	1.357	.578	3.187
	Constant	.531	.282	3.546	1	.060	1.700		

Sequence generation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECseqgencode	.044	.465	.009	1	.924	1.045	.420	2.601
	Constant	.649	.257	6.356	1	.012	1.913		

Blinding

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECblindcode	.440	.431	1.041	1	.308	1.553	.667	3.617
	Constant	.425	.312	1.856	1	.173	1.529		

Number of outcomes

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numoutgrp2	-2.506	.511	24.019	1	.000	.082	.030	.222
	Constant	4.359	.834	27.322	1	.000	78.155		

Appendix 6.1.2: Threshold for discrepancy in the identity of the primary outcome = 100%: adjusted for number of outcomes

Proportion of trials with a discrepancy in the identity of a primary outcome		0: <100% 1: 100%

Design

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numoutgrp2	-2.513	.513	24.026	1	.000	.081	.030	.221
	designcode	-.233	.874	.071	1	.789	.792	.143	4.393
	Constant	4.581	1.188	14.874	1	.000	97.603		

Purpose

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numoutgrp2	-2.503	.512	23.939	1	.000	.082	.030	.223
	purposecode	.168	.609	.077	1	.782	1.183	.359	3.903
	Constant	4.223	.960	19.365	1	.000	68.248		

Administration

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numoutgrp2	-2.489	.518	23.125	1	.000	.083	.030	.229
	administration			1.562	2	.458			
	administration(1)	.086	.761	.013	1	.910	1.090	.245	4.846
	administration(2)	.670	.577	1.347	1	.246	1.953	.631	6.051
	Constant	3.978	.918	18.761	1	.000	53.407		

Commercial funding (protocol)

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numoutgrp2	-2.635	.547	23.174	1	.000	.072	.025	.210
	RECdrugfundcode	1.253	.555	5.107	1	.024	3.502	1.181	10.388
	Constant	3.791	.862	19.348	1	.000	44.306		

Proposed sample size

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numoutgrp2	-2.687	.551	23.776	1	.000	.068	.023	.200
	RECss200code	1.134	.549	4.271	1	.039	3.109	1.060	9.118
	Constant	4.040	.851	22.527	1	.000	56.846		

Completeness of the sample size calculation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numoutgrp2	-3.082	.679	20.614	1	.000	.046	.012	.174
	sscompcode2			10.815	2	.004			
	sscompcode2(1)	1.693	.968	3.062	1	.080	5.436	.816	36.217
	sscompcode2(2)	3.040	.935	10.562	1	.001	20.896	3.341	130.668
	Constant	2.908	1.085	7.180	1	.007	18.325		

Allocation concealment

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numoutgrp2	-2.496	.512	23.766	1	.000	.082	.030	.225
	RECalloccode	.176	.515	.117	1	.732	1.193	.435	3.275
	Constant	4.268	.870	24.073	1	.000	71.356		

Sequence generation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numoutgrp2	-2.509	.512	24.001	1	.000	.081	.030	.222
	RECseqgencode	-.057	.552	.011	1	.918	.945	.320	2.789
	Constant	4.380	.859	25.965	1	.000	79.804		

Blinding

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numoutgrp2	-2.609	.532	24.038	1	.000	.074	.026	.209
	RECblindcode	.757	.531	2.032	1	.154	2.131	.753	6.033
	Constant	4.103	.847	23.489	1	.000	60.511		

Appendix 6.1.3: Threshold for discrepancy in the identity of the primary outcome = 66%

Proportion of trials with a discrepancy in the identity of a primary outcome	0: <=66% 1: >66%

Purpose

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B) Lower	Upper
Step 1(a)	purposecode	.405	.533	.578	1	.447	1.500	.527	4.267
	Constant	.693	.463	2.242	1	.134	2.000		

Administration

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B) Lower	Upper
Step 1(a)	administration			5.439	2	.066			
	administration(1)	.440	.658	.447	1	.504	1.553	.428	5.640
	administration(2)	1.216	.526	5.344	1	.021	3.373	1.203	9.453
	Constant	.348	.377	.853	1	.356	1.417		

Commercial funding (protocol)

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B) Lower	Upper
Step 1(a)	RECdrugfundcode	.582	.468	1.544	1	.214	1.789	.715	4.477
	Constant	.651	.356	3.338	1	.068	1.917		

Proposed sample size

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B) Lower	Upper
Step 1(a)	RECss200code	1.124	.479	5.503	1	.019	3.078	1.203	7.875
	Constant	.463	.310	2.233	1	.135	1.588		

Completeness of the sample size calculation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B) Lower	Upper
Step 1(a)	sscompcode2			9.136	2	.010			
	sscompcode2(1)	1.243	.754	2.718	1	.099	3.467	.791	15.197
	sscompcode2(2)	1.955	.654	8.950	1	.003	7.067	1.963	25.443
	Constant	-.470	.570	.680	1	.410	.625		

Allocation concealment

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B) Lower	Upper
Step 1(a)	RECalloccode	.551	.476	1.343	1	.246	1.736	.683	4.410
	Constant	.778	.293	7.045	1	.008	2.176		

Sequence generation

	B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for

								EXP(B)	
								Lower	Upper
Step 1(a)	RECseqgencode	.263	.510	.266	1	.606	1.301	.479	3.532
	Constant	.927	.271	11.691	1	.001	2.526		

Blinding

								95.0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1(a)	RECblindcode	.958	.471	4.132	1	.042	2.607	1.035	6.569
	Constant	.523	.315	2.751	1	.097	1.688		

Appendix 6.1.4: Threshold for discrepancy in the identity of the primary outcome = 66%: adjusted for number of outcomes

Proportion of trials with a discrepancy in the identity of a primary outcome	0: <=66% 1: >66%

Purpose

								95.0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1(a)	numoutgrp2	-1.743	.501	12.083	1	.001	.175	.066	.468
	purposecode	.380	.578	.432	1	.511	1.462	.471	4.538
	Constant	3.301	.932	12.558	1	.000	27.148		

Administration

								95.0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1(a)	numoutgrp2	-1.716	.514	11.150	1	.001	.180	.066	.492
	administration			4.267	2	.118			
	administration(1)	.309	.714	.188	1	.665	1.363	.336	5.527
	administration(2)	1.139	.565	4.066	1	.044	3.124	1.032	9.456
	Constant	2.956	.900	10.778	1	.001	19.212		

Commercial funding (protocol)

								95.0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1(a)	numoutgrp2	-1.741	.504	11.938	1	.001	.175	.065	.471
	RECdrugfundcode	.562	.506	1.236	1	.266	1.755	.651	4.728
	Constant	3.248	.874	13.816	1	.000	25.734		

Proposed sample size

								95.0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1(a)	numoutgrp2	-1.920	.536	12.841	1	.000	.147	.051	.419
	RECss200code	1.361	.536	6.452	1	.011	3.898	1.364	11.137
	Constant	3.201	.853	14.096	1	.000	24.554		

Completeness of the sample size calculation

								95.0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1(a)	numoutgrp2	-2.048	.584	12.279	1	.000	.129	.041	.406
	sscompcode2			9.429	2	.009			
	sscompcode2(1)	.763	.845	.816	1	.366	2.145	.409	11.239
	sscompcode2(2)	2.142	.749	8.170	1	.004	8.515	1.960	36.987
	Constant	2.564	1.050	5.964	1	.015	12.984		

Allocation concealment

								95.0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper

Step 1(a)	numoutgrp2	-1.728	.503	11.816	1	.001	.178	.066	.476
	RECalloccode	.484	.510	.899	1	.343	1.622	.597	4.407
	Constant	3.372	.854	15.599	1	.000	29.124		

Sequence generation

								95.0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1(a)	numoutgrp2	-1.743	.501	12.126	1	.000	.175	.066	.467
	RECseqgencode	.231	.547	.179	1	.673	1.260	.431	3.684
	Constant	3.525	.842	17.517	1	.000	33.968		

Blinding

								95.0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1(a)	numoutgrp2	-1.922	.533	13.015	1	.000	.146	.051	.416
	RECblindcode	1.223	.529	5.335	1	.021	3.396	1.203	9.584
	Constant	3.246	.848	14.646	1	.000	25.694		

Appendix 6.1.5: Threshold for discrepancy in the identity of the primary outcome = 0%: unadjusted for number of outcomes

Proportion of trials with a discrepancy in the identity of a primary outcome	0: <=0% 1: >0%

Purpose

								95.0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1(a)	purposecode	.440	.651	.457	1	.499	1.553	.433	5.564
	Constant	1.447	.556	6.779	1	.009	4.250		

Administration

								95.0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1(a)	administration			8.129	2	.017			
	administration(1)	1.147	.856	1.797	1	.180	3.150	.589	16.859
	administration(2)	1.995	.718	7.728	1	.005	7.350	1.801	29.995
	Constant	.799	.401	3.958	1	.047	2.222		

Commercial funding (protocol)

								95.0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1(a)	RECdrugfundcode	1.017	.589	2.986	1	.084	2.765	.872	8.767
	Constant	1.216	.403	9.131	1	.003	3.375		

Proposed sample size

								95.0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1(a)	RECss200code	1.715	.689	6.197	1	.013	5.556	1.440	21.434
	Constant	1.099	.348	9.957	1	.002	3.000		

Completeness of the sample size calculation

								95.0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1(a)	sscompcode2			18.519	2	.000			
	sscompcode2(1)	2.144	.849	6.377	1	.012	8.533	1.616	45.061
	sscompcode2(2)	3.499	.821	18.147	1	.000	33.067	6.612	165.366
	Constant	-.470	.570	.680	1	.410	.625		

Allocation concealment

								95.0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1(a)	RECalloccode	1.768	.795	4.948	1	.026	5.857	1.234	27.805
	Constant	1.253	.327	14.648	1	.000	3.500		

Sequence generation

	B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for

								EXP(B)	
								Lower	Upper
Step 1(a)	RECseqgencode	1.117	.798	1.957	1	.162	3.055	.639	14.603
	Constant	1.522	.319	22.832	1	.000	4.583		

Blinding

								95.0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1(a)	RECblindcode	1.765	.689	6.560	1	.010	5.844	1.513	22.563
	Constant	1.068	.350	9.334	1	.002	2.909		

Number of outcomes

								95.0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1(a)	numoutgrp2	-.517	.581	.793	1	.373	.596	.191	1.861
	Constant	2.523	.907	7.728	1	.005	12.461		

Appendix 6.1.6: Threshold for discrepancy in the identity of the primary outcome = 0%: adjusted for number of outcomes

Proportion of trials with a discrepancy in the identity of a primary outcome	0: <=0% 1: >0%

Purpose

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numoutgrp2	-.504	.582	.750	1	.386	.604	.193	1.891
	purposecode	.420	.654	.411	1	.521	1.521	.422	5.486
	Constant	2.187	1.039	4.427	1	.035	8.906		

Administration

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numoutgrp2	-.329	.614	.288	1	.592	.720	.216	2.396
	administration			7.714	2	.021			
	administration(1)	1.116	.859	1.690	1	.194	3.054	.567	16.440
	administration(2)	1.956	.721	7.355	1	.007	7.069	1.720	29.053
	Constant	1.292	1.013	1.627	1	.202	3.639		

Commercial funding (protocol)

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numoutgrp2	-.474	.591	.645	1	.422	.622	.196	1.980
	RECdrugfundcode	.997	.591	2.842	1	.092	2.709	.850	8.628
	Constant	1.909	.969	3.879	1	.049	6.746		

Proposed sample size

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numoutgrp2	-.579	.607	.912	1	.340	.560	.171	1.840
	RECss200code	1.737	.693	6.290	1	.012	5.681	1.462	22.084
	Constant	1.922	.949	4.096	1	.043	6.832		

Completeness of the sample size calculation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numoutgrp2	-.423	.715	.350	1	.554	.655	.161	2.659
	sscompcode2			17.949	2	.000			
	sscompcode2(1)	2.021	.871	5.378	1	.020	7.544	1.367	41.619
	sscompcode2(2)	3.472	.824	17.759	1	.000	32.185	6.404	161.757
	Constant	.176	1.225	.021	1	.886	1.192		

Allocation concealment

	B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
							Lower	Upper

Step 1(a)	numoutgrp2	-.432	.600	.518	1	.472	.649	.201	2.103
	RECalloccode	1.742	.796	4.783	1	.029	5.706	1.198	27.175
	Constant	1.881	.949	3.930	1	.047	6.559		

Sequence generation

								95.0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1(a)	numoutgrp2	-.499	.587	.721	1	.396	.607	.192	1.920
	RECseqgencode	1.105	.800	1.907	1	.167	3.019	.629	14.487
	Constant	2.241	.926	5.854	1	.016	9.407		

Blinding

								95.0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1(a)	numoutgrp2	-.633	.611	1.074	1	.300	.531	.160	1.758
	RECblindcode	1.809	.695	6.771	1	.009	6.105	1.563	23.846
	Constant	1.959	.949	4.258	1	.039	7.091		

Appendix 6.2: Univariate analyses for completeness of reporting

Appendix 6.2.1: Threshold for completeness of reporting = 100%

Proportion of trials with fully reported comparisons		0: <100% 1: 100%

Purpose

								95.0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1(a)	purpose	.038	.587	.004	1	.948	1.039	.329	3.283
	Constant	1.099	.516	4.526	1	.033	3.000		

Administration

								95.0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1(a)	administration			.101	2	.951			
	administration(1)	.249	.786	.101	1	.751	1.283	.275	5.984
	administration(2)	.076	.551	.019	1	.890	1.079	.367	3.175
	Constant	1.050	.439	5.715	1	.017	2.857		

Commercial funding (protocol)

								95.0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1(a)	RECdrugfund	.671	.498	1.820	1	.177	1.957	.738	5.188
	Constant	.738	.367	4.048	1	.044	2.091		

Proposed sample size

								95.0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1(a)	RECss200c	.054	.493	.012	1	.913	1.056	.402	2.774
	Constant	1.099	.365	9.052	1	.003	3.000		

Completeness of the sample size calculation

								95.0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1(a)	sscompcode2			4.732	2	.094			
	sscompcode2(1)	1.674	.854	3.843	1	.050	5.333	1.000	28.435
	sscompcode2(2)	1.264	.657	3.697	1	.055	3.538	.976	12.832
	Constant	.000	.577	.000	1	1.000	1.000		

Allocation concealment

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECalloc	-.054	.493	.012	1	.913	.947	.361	2.489
	Constant	1.153	.331	12.117	1	.000	3.167		

Sequence generation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECseqgen	-.388	.519	.558	1	.455	.679	.245	1.877
	Constant	1.253	.303	17.089	1	.000	3.500		

Blinding

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECblind	.177	.491	.130	1	.719	1.194	.456	3.125
	Constant	1.036	.351	8.716	1	.003	2.818		

Number of comparisons

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numbercompgrp2	1.277	.538	5.638	1	.018	3.586	1.250	10.291
	Constant	.595	.311	3.647	1	.056	1.813		

Appendix 6.2.2: Threshold for completeness of reporting = 100%: adjusted for number of comparisons

Proportion of trials with fully reported comparisons		0: <100% 1: 100%

Purpose

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numbercompgrp2	1.305	.545	5.735	1	.017	3.687	1.267	10.727
	purpose	-.203	.612	.110	1	.740	.816	.246	2.709
	Constant	.740	.542	1.868	1	.172	2.097		

Administration

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numbercompgrp2	1.289	.544	5.622	1	.018	3.628	1.250	10.528
	administration			.090	2	.956			
	administration(1)	.097	.817	.014	1	.905	1.102	.222	5.466
	administration(2)	.171	.572	.090	1	.765	1.187	.387	3.644
	Constant	.482	.497	.941	1	.332	1.619		

Commercial funding (protocol)

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numbercompgrp2	1.306	.545	5.747	1	.017	3.693	1.269	10.748
	RECdrugfund	.723	.518	1.946	1	.163	2.061	.746	5.692
	Constant	.161	.436	.137	1	.711	1.175		

Proposed sample size

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECss200c	.176	.514	.118	1	.732	1.193	.436	3.264
	numbercompgrp2	1.295	.541	5.724	1	.017	3.649	1.264	10.539
	Constant	.490	.434	1.275	1	.259	1.632		

Completeness of the sample size calculation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numbercompgrp2	1.496	.582	6.602	1	.010	4.465	1.426	13.981
	sscompcode2			5.878	2	.053			
	sscompcode2(1)	2.035	.920	4.887	1	.027	7.650	1.260	46.467
	sscompcode2(2)	1.569	.723	4.708	1	.030	4.802	1.164	19.810
	Constant	-.884	.711	1.549	1	.213	.413		

Allocation concealment

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numbercompgrp2	1.334	.553	5.807	1	.016	3.795	1.283	11.226
	RECalloc	.242	.525	.213	1	.644	1.274	.456	3.562
	Constant	.462	.421	1.206	1	.272	1.588		

Sequence generation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numbercompgrp2	1.260	.539	5.464	1	.019	3.527	1.226	10.146
	RECseqgen	-.315	.538	.343	1	.558	.730	.254	2.095
	Constant	.703	.365	3.702	1	.054	2.019		

Blinding

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numbercompgrp2	1.406	.560	6.288	1	.012	4.078	1.359	12.232
	RECblind	.511	.530	.928	1	.335	1.667	.590	4.712
	Constant	.274	.452	.366	1	.545	1.315		

Appendix 6.2.3: Threshold for completeness of reporting = 66%

Proportion of trials with fully reported comparisons		0: ≤ 66% 1: > 66%

Purpose

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	purpose	.581	.612	.901	1	.342	1.788	.539	5.933
	Constant	1.099	.516	4.526	1	.033	3.000		

Administration

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	administration			.547	2	.761			
	administration(1)	.539	.893	.364	1	.546	1.714	.298	9.869
	administration(2)	.381	.603	.400	1	.527	1.464	.449	4.775
	Constant	1.253	.463	7.324	1	.007	3.500		

Commercial funding (protocol)

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECdrugfund	-.014	.569	.001	1	.980	.986	.323	3.009
	Constant	1.540	.450	11.725	1	.001	4.667		

Proposed sample size

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECss200c	.579	.556	1.081	1	.298	1.783	.599	5.306
	Constant	1.237	.379	10.669	1	.001	3.444		

Completeness of the sample size calculation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	sscompcode2			8.270	2	.016			
	sscompcode2(1)	2.140	.945	5.133	1	.023	8.500	1.335	54.127
	sscompcode2(2)	1.852	.691	7.179	1	.007	6.375	1.644	24.714
	Constant	.000	.577	.000	1	1.000	1.000		

Allocation concealment

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECalloc	1.045	.622	2.816	1	.093	2.842	.839	9.626
	Constant	1.153	.331	12.117	1	.000	3.167		

Sequence generation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECseqgen	-.072	.596	.014	1	.904	.931	.289	2.995
	Constant	1.553	.332	21.908	1	.000	4.727		

Blinding

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECblind	.468	.556	.711	1	.399	1.597	.538	4.746
	Constant	1.299	.376	11.938	1	.001	3.667		

Number of comparisons

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numbercompgrp2	.619	.566	1.194	1	.274	1.857	.612	5.637
	Constant	1.253	.359	12.207	1	.000	3.500		

Appendix 6.2.4: Threshold for completeness of reporting = 66%: adjusted for number of comparisons

Proportion of trials with fully reported comparisons		0: <=66% 1: >66%

Purpose

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numbercompgrp2	.553	.575	.926	1	.336	1.739	.564	5.364
	purpose	.483	.623	.602	1	.438	1.621	.478	5.490
	Constant	.923	.545	2.867	1	.090	2.516		

Administration

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numbercompgrp2	.630	.573	1.208	1	.272	1.878	.610	5.778
	administration			.567	2	.753			
	administration(1)	.466	.901	.268	1	.605	1.594	.272	9.324
	administration(2)	.430	.610	.499	1	.480	1.538	.466	5.080
	Constant	.954	.528	3.258	1	.071	2.595		

Commercial funding (protocol)

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numbercompgrp2	.619	.566	1.194	1	.274	1.857	.612	5.637
	RECdrugfund	-.015	.573	.001	1	.980	.986	.320	3.031
	Constant	1.262	.507	6.206	1	.013	3.532		

Proposed sample size

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numbercompgrp2	.687	.575	1.431	1	.232	1.988	.645	6.132
	RECss200c	.650	.565	1.324	1	.250	1.915	.633	5.796
	Constant	.891	.464	3.688	1	.055	2.437		

Completeness of the sample size calculation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numbercompgrp2	.852	.621	1.884	1	.170	2.345	.694	7.919
	sscompcode2			8.784	2	.012			
	sscompcode2(1)	2.306	.972	5.622	1	.018	10.030	1.491	67.449
	sscompcode2(2)	1.998	.721	7.678	1	.006	7.375	1.795	30.311
	Constant	-.499	.695	.516	1	.472	.607		

Allocation concealment

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numbercompgrp2	.912	.597	2.329	1	.127	2.488	.772	8.022
	RECalloc	1.272	.649	3.841	1	.050	3.567	1.000	12.722
	Constant	.657	.446	2.172	1	.141	1.929		

Sequence generation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numbercompgrp2	.617	.568	1.182	1	.277	1.854	.609	5.643
	RECseqgen	-.024	.602	.002	1	.968	.976	.300	3.175
	Constant	1.261	.412	9.377	1	.002	3.528		

Blinding

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numbercompgrp2	.774	.589	1.727	1	.189	2.168	.684	6.875
	RECblind	.650	.579	1.260	1	.262	1.916	.616	5.963
	Constant	.860	.487	3.123	1	.077	2.364		

Appendix 6.2.5: Threshold for completeness of reporting = 0%

Proportion of trials with fully reported comparisons		0: =0% 1: >0%

Purpose

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	purpose	.000	.845	.000	1	1.000	1.000	.191	5.241
	Constant	2.197	.745	8.690	1	.003	9.000		

Administration

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	administration			.164	2	.921			
	administration(1)	.486	1.205	.162	1	.687	1.625	.153	17.239
	administration(2)	.095	.773	.015	1	.902	1.100	.242	5.006
	Constant	2.079	.612	11.531	1	.001	8.000		

Commercial funding (protocol)

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECdrugfund	-.215	.743	.084	1	.772	.806	.188	3.461
	Constant	2.335	.605	14.918	1	.000	10.333		

Proposed sample size

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECss200c	.496	.707	.493	1	.483	1.643	.411	6.572
	Constant	1.946	.478	16.566	1	.000	7.000		

Completeness of the sample size calculation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	sscompcode2			3.264	2	.196			
	sscompcode2(1)	1.041	1.002	1.081	1	.298	2.833	.398	20.179
	sscompcode2(2)	1.522	.844	3.252	1	.071	4.583	.876	23.974
	Constant	1.099	.667	2.716	1	.099	3.000		

Allocation concealment

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECalloc	1.129	.832	1.841	1	.175	3.093	.605	15.801
	Constant	1.815	.408	19.838	1	.000	6.143		

Sequence generation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECseqgen	-.702	.715	.964	1	.326	.496	.122	2.011
	Constant	2.451	.466	27.653	1	.000	11.600		

Blinding

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECblind	.916	.742	1.527	1	.217	2.500	.584	10.696
	Constant	1.792	.441	16.511	1	.000	6.000		

Number of comparisons

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numbercompgrp2	-.767	.741	1.071	1	.301	.464	.109	1.985
	Constant	2.639	.598	19.501	1	.000	14.000		

Appendix 6.2.6: Threshold for completeness of reporting = 0%: adjusted for number of comparisons

Proportion of trials with fully reported comparisons		0: =0% 1: >0%

Purpose

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numbercompgrp2	-.789	.752	1.100	1	.294	.454	.104	1.984
	purpose	.154	.863	.032	1	.859	1.166	.215	6.334
	Constant	2.532	.840	9.092	1	.003	12.576		

Administration

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numbercompgrp2	-.809	.747	1.172	1	.279	.445	.103	1.927
	administration			.254	2	.881			
	administration(1)	.584	1.214	.231	1	.631	1.793	.166	19.347
	administration(2)	.041	.780	.003	1	.958	1.042	.226	4.809
	Constant	2.562	.795	10.381	1	.001	12.955		

Commercial funding (protocol)

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numbercompgrp2	-.768	.742	1.072	1	.300	.464	.108	1.985
	RECdrugfund	-.218	.748	.085	1	.771	.804	.186	3.481
	Constant	2.779	.778	12.768	1	.000	16.108		

Proposed sample size

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numbercompgrp2	-.731	.745	.964	1	.326	.481	.112	2.072
	RECss200c	.438	.713	.377	1	.539	1.549	.383	6.269
	Constant	2.397	.699	11.748	1	.001	10.986		

Completeness of the sample size calculation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numbercompgrp2	-.716	.755	.901	1	.343	.488	.111	2.145
	sscompcode2			3.026	2	.220			
	sscompcode2(1)	.982	1.011	.944	1	.331	2.669	.368	19.349
	sscompcode2(2)	1.480	.851	3.022	1	.082	4.393	.828	23.304
	Constant	1.548	.843	3.369	1	.066	4.702		

Allocation concealment

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numbercompgrp2	-.574	.759	.573	1	.449	.563	.127	2.492
	RECalloc	1.008	.847	1.415	1	.234	2.739	.521	14.418
	Constant	2.188	.668	10.714	1	.001	8.917		

Sequence generation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numbercompgrp2	-.835	.750	1.240	1	.265	.434	.100	1.887
	RECseqgen	-.779	.725	1.154	1	.283	.459	.111	1.900
	Constant	2.960	.699	17.936	1	.000	19.295		

Blinding

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numbercompgrp2	-.602	.760	.627	1	.428	.548	.124	2.429
	RECblind	.791	.758	1.088	1	.297	2.205	.499	9.747
	Constant	2.194	.702	9.769	1	.002	8.971		

Appendix 6.3: Univariate analyses for completeness of documentation of the sample size calculation

(0=incomplete; 1=partial or complete)

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	purposecode	.795	.575	1.909	1	.167	2.214	.717	6.836
	Constant	1.041	.475	4.810	1	.028	2.833		

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	administration			11.415	2	.003			
	administration(1)	.981	.732	1.796	1	.180	2.667	.635	11.194
	administration(2)	2.679	.807	11.033	1	.001	14.571	2.999	70.804
	Constant	.560	.362	2.391	1	.122	1.750		

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECdrugfundcode	1.388	.558	6.182	1	.013	4.006	1.342	11.963
	Constant	.898	.358	6.302	1	.012	2.455		

Appendix 6.4: Univariate analyses for adequacy of reporting of the power calculation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	purposecode	1.795	.511	12.360	1	.000	6.020	2.213	16.375
	Constant	-.629	.438	2.062	1	.151	.533		

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	administration			15.460	2	.000			
	administration(1)	.662	.606	1.195	1	.274	1.939	.591	6.355
	administration(2)	2.033	.521	15.229	1	.000	7.634	2.750	21.188
	Constant	-.305	.352	.752	1	.386	.737		

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECdrugfundcode	-.103	.436	.056	1	.813	.902	.383	2.122
	Constant	.773	.349	4.908	1	.027	2.167		

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	sscompcode2			7.348	2	.025			
	sscompcode2(1)	1.130	.697	2.624	1	.105	3.095	.789	12.144
	sscompcode2(2)	1.590	.588	7.314	1	.007	4.902	1.549	15.513
	Constant	-.511	.516	.979	1	.323	.600		

Appendix 6.5: Univariate analyses for journal type

Outcome: journal (0 = specialty; 1 = general)

Purpose

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	purposecode	2.303	1.051	4.804	1	.028	10.000	1.276	78.383
	Constant	-3.091	1.022	9.139	1	.003	.045		

Administration

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	administration			7.981	2	.018			
	administration (1)	2.590	1.147	5.104	1	.024	13.333	1.409	126.150
	administration (2)	2.965	1.054	7.909	1	.005	19.394	2.456	153.140
	Constant	-3.466	1.016	11.647	1	.001	.031		

Commercial funding (protocol)

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECdrugfundcode	-.090	.468	.037	1	.848	.914	.366	2.286
	Constant	-1.030	.368	7.811	1	.005	.357		

Proposed sample size

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECss200c	2.703	.779	12.029	1	.001	14.929	3.240	68.779
	Constant	-2.944	.725	16.472	1	.000	.053		

Completeness of the sample size calculation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	sscompcode2			2.906	2	.234			
	sscompcode2(1)	1.861	1.142	2.654	1	.103	6.429	.685	60.313
	sscompcode2(2)	1.781	1.068	2.783	1	.095	5.937	.732	48.138
	Constant	-2.708	1.033	6.875	1	.009	.067		

Allocation concealment

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECalloccode	1.996	.526	14.375	1	.000	7.361	2.623	20.659
	Constant	-2.179	.431	25.580	1	.000	.113		

Sequence generation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECseqgencode	.511	.477	1.145	1	.285	1.667	.654	4.249
	Constant	-1.253	.283	19.530	1	.000	.286		

Blinding

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECblindcode	.232	.458	.257	1	.612	1.261	.514	3.094
	Constant	-1.213	.343	12.476	1	.000	.297		

Appendix 6.6: Univariate analyses for exclusions

Outcome: exclusions (0 = patients excluded; 1 = no apparent exclusions)
Note: this is based on all 103 trials

Design

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	designcode	.671	.647	1.075	1	.300	1.957	.550	6.957
	Constant	-.693	.612	1.281	1	.258	.500		

Purpose

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	purposecode	.212	.476	.199	1	.656	1.237	.486	3.146
	Constant	-.262	.421	.389	1	.533	.769		

Administration

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	centrenationcode			1.212	2	.546			
	centrenationcode(1)	-.051	.606	.007	1	.933	.950	.290	3.114
	centrenationcode(2)	.419	.447	.878	1	.349	1.520	.633	3.650
	Constant	-.305	.352	.752	1	.386	.737		

Commercial funding (protocol)

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECdrugfundcode	.181	.410	.194	1	.660	1.198	.536	2.675
	Constant	-.211	.326	.419	1	.517	.810		

Proposed sample size

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECssgroup			2.491	2	.288			
	RECssgroup(1)	-.307	.462	.441	1	.507	.736	.297	1.820
	RECssgroup(2)	.506	.558	.823	1	.364	1.659	.555	4.956
	Constant	-.065	.359	.032	1	.857	.938		

Completeness of the sample size calculation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	sscompcode2			1.793	2	.408			
	sscompcode2(1)	-.870	.688	1.599	1	.206	.419	.109	1.614
	sscompcode2(2)	-.281	.560	.252	1	.616	.755	.252	2.263
	Constant	.251	.504	.249	1	.618	1.286		

Allocation concealment

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECalloccode	.490	.401	1.489	1	.222	1.632	.743	3.584
	Constant	-.307	.263	1.362	1	.243	.735		

Sequence generation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECseqgencode	.417	.432	.934	1	.334	1.518	.651	3.539
	Constant	-.223	.237	.885	1	.347	.800		

Blinding

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECblindcode	-.339	.397	.731	1	.392	.712	.327	1.550
	Constant	.083	.289	.083	1	.773	1.087		

Appendix 6.7: Univariate analyses for commercial funding in protocol

Note: these analyses are based on all 103 trials

Commercial funding mentioned in protocol	0: no commercial funding mentioned 1: commercial funding mentioned

Purpose

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	purposecode	-.369	.508	.527	1	.468	.692	.256	1.871
	Constant	.827	.453	3.328	1	.068	2.286		

Administration

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	administration			11.797	2	.003			
	administration(1)	1.037	.620	2.796	1	.094	2.821	.837	9.509
	administration(2)	1.659	.484	11.736	1	.001	5.256	2.034	13.584
	Constant	-.431	.356	1.462	1	.227	.650		

Proposed sample size

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECss200code	.490	.411	1.419	1	.234	1.632	.729	3.655
	Constant	.288	.289	.993	1	.319	1.333		

Completeness of the sample size calculation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	sscompcode2			7.480	2	.024			
	sscompcode2(1)	1.407	.715	3.879	1	.049	4.086	1.007	16.579
	sscompcode2(2)	1.643	.602	7.452	1	.006	5.170	1.589	16.817
	Constant	-.788	.539	2.137	1	.144	.455		

Allocation concealment

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECalloccode	.040	.413	.009	1	.923	1.041	.463	2.338
	Constant	.520	.269	3.729	1	.053	1.682		

Sequence generation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECseqgencode	.953	.491	3.762	1	.052	2.592	.990	6.788
	Constant	.280	.238	1.380	1	.240	1.323		

Blinding

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECblindcode	1.256	.429	8.565	1	.003	3.512	1.514	8.144
	Constant	-.083	.289	.083	1	.773	.920		

Appendix 6.8: Univariate analyses for commercial funding in publication

Commercial funding mentioned in publication	0: no commercial funding mentioned 1: commercial funding mentioned

Sample size achieved (0: <=200; 1: >200)

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	pubss200code	.328	.409	.643	1	.422	1.388	.623	3.096
	Constant	.314	.302	1.080	1	.299	1.368		

Reporting of power calculation (0: inadequate; 1: adequate)

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	pubpowercalccode	-.164	.435	.142	1	.706	.848	.361	1.992
	Constant	.606	.359	2.853	1	.091	1.833		

Allocation concealment reported

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	pubacokcode	-.290	.551	.278	1	.598	.748	.254	2.202
	Constant	.542	.222	5.934	1	.015	1.719		

Sequence generation reported (0:

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	pubsgokcode	-.459	.419	1.202	1	.273	.632	.278	1.435
	Constant	.670	.262	6.536	1	.011	1.955		

Blinding (0: not reported as double blind; 1: reported as double blind)

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	pubblindcode	1.340	.430	9.721	1	.002	3.818	1.645	8.864
	Constant	-.167	.290	.333	1	.564	.846		

Appendix 7: Multivariate models

Appendix 7.1: Multivariate models for discrepancy in the identity of the primary outcome

Proportion of trials with a difference in the identity of a primary outcome	0: <100% 1: 100%

These analyses are based on the 97 trials with at least one identifiable primary outcome. Logistic regression (best model) based on significant covariates in univariate analysis.

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	RECdrugfundcode	1.091	.607	3.229	1	.072	2.978	.906	9.794
	RECss200code	.827	.622	1.768	1	.184	2.287	.676	7.745
	sscompcode2			7.250	2	.027			
	sscompcode2(1)	1.350	1.013	1.774	1	.183	3.856	.529	28.097
	sscompcode2(2)	2.556	.969	6.959	1	.008	12.886	1.929	86.077
	numoutgrp2	-3.290	.725	20.619	1	.000	.037	.009	.154
	Constant	2.539	1.131	5.040	1	.025	12.662		
Step 2(a)	RECdrugfundcode	1.017	.602	2.856	1	.091	2.766	.850	9.000
	sscompcode2			9.315	2	.009			
	sscompcode2(1)	1.600	1.002	2.550	1	.110	4.955	.695	35.321
	sscompcode2(2)	2.898	.964	9.029	1	.003	18.131	2.739	120.020
	numoutgrp2	-3.171	.701	20.471	1	.000	.042	.011	.166
	Constant	2.507	1.118	5.025	1	.025	12.262		

Proportion of trials with a difference in the identity of a primary outcome		0: ≤ 66% 1: > 66%

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numoutgrp2	-2.292	.634	13.090	1	.000	.101	.029	.350
	RECblindcode	.958	.575	2.772	1	.096	2.605	.844	8.042
	RECss200code	.984	.592	2.762	1	.097	2.674	.838	8.530
	sscompcode2			6.171	2	.046			
	sscompcode2(1)	.565	.871	.421	1	.516	1.760	.319	9.694
	sscompcode2(2)	1.749	.753	5.392	1	.020	5.747	1.314	25.148
	Constant	2.292	1.092	4.404	1	.036	9.895		

Proportion of trials with a difference in the identity of a primary outcome		0: =0% 1: >0%

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	administration			.015	2	.992			
	administration(1)	.080	1.143	.005	1	.944	1.084	.115	10.189
	administration(2)	-.079	1.428	.003	1	.956	.924	.056	15.189
	RECdrugfundcode	.022	.852	.001	1	.979	1.023	.193	5.431
	RECss200code	1.118	1.124	.990	1	.320	3.059	.338	27.689
	sscompcode2			10.185	2	.006			
	sscompcode2(1)	2.500	1.183	4.468	1	.035	12.184	1.200	123.742
	sscompcode2(2)	3.382	1.061	10.151	1	.001	29.419	3.674	235.546
	RECalloccode	.429	1.137	.142	1	.706	1.535	.165	14.249
	RECblindcode	2.153	.991	4.722	1	.030	8.614	1.235	60.080
	Constant	-1.870	.929	4.050	1	.044	.154		
Step 2(a)	RECdrugfundcode	.013	.807	.000	1	.988	1.013	.208	4.927
	RECss200code	1.079	.873	1.526	1	.217	2.942	.531	16.299
	sscompcode2			10.203	2	.006			
	sscompcode2(1)	2.469	1.139	4.703	1	.030	11.812	1.268	110.014
	sscompcode2(2)	3.375	1.059	10.162	1	.001	29.230	3.669	232.859
	RECalloccode	.414	.984	.177	1	.674	1.513	.220	10.417
	RECblindcode	2.140	.973	4.833	1	.028	8.500	1.261	57.280
	Constant	-1.843	.896	4.229	1	.040	.158		
Step 3(a)	RECss200code	1.079	.873	1.528	1	.216	2.941	.532	16.269
	sscompcode2			10.499	2	.005			
	sscompcode2(1)	2.475	1.080	5.251	1	.022	11.880	1.431	98.652
	sscompcode2(2)	3.377	1.049	10.372	1	.001	29.297	3.751	228.819
	RECalloccode	.415	.983	.178	1	.673	1.514	.221	10.389
	RECblindcode	2.144	.934	5.275	1	.022	8.537	1.369	53.216
	Constant	-1.840	.876	4.410	1	.036	.159		
Step 4(a)	RECss200code	1.175	.841	1.952	1	.162	3.238	.623	16.833
	sscompcode2			13.713	2	.001			
	sscompcode2(1)	2.620	1.033	6.437	1	.011	13.734	1.815	103.927
	sscompcode2(2)	3.564	.971	13.480	1	.000	35.313	5.267	236.736
	RECblindcode	2.162	.937	5.323	1	.021	8.685	1.384	54.489
	Constant	-1.873	.878	4.553	1	.033	.154		
Step 5(a)	sscompcode2			15.891	2	.000			
	sscompcode2(1)	2.706	1.011	7.161	1	.007	14.972	2.063	108.663
	sscompcode2(2)	3.789	.956	15.713	1	.000	44.193	6.789	287.664
	RECblindcode	2.139	.898	5.677	1	.017	8.488	1.461	49.302
	Constant	-1.585	.834	3.611	1	.057	.205		

Appendix 7.2: Multivariate model for completeness of reporting

Proportion of trials with fully reported comparisons		0: <100% 1: 100%

This analysis is based on the 90 trials with at least one identifiable primary outcome that is not global or involving more than 2 time points

Best model based on significant covariates in univariate analysis

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	numbercompgrp2	1.496	.582	6.602	1	.010	4.465	1.426	13.981
	sscompcode2			5.878	2	.053			
	sscompcode2(1)	2.035	.920	4.887	1	.027	7.650	1.260	46.467
	sscompcode2(2)	1.569	.723	4.708	1	.030	4.802	1.164	19.810
	Constant	-.884	.711	1.549	1	.213	.413		

Proportion of trials with fully reported comparisons		0: ≤ 66% 1: > 66%

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	sscompcode2			5.400	2	.067			
	sscompcode2(1)	1.974	1.009	3.829	1	.050	7.196	.997	51.956
	sscompcode2(2)	1.665	.775	4.620	1	.032	5.285	1.158	24.116
	RECalloc	.754	.712	1.120	1	.290	2.125	.526	8.581
	numbercompgrp2	.980	.634	2.385	1	.123	2.664	.768	9.237
	Constant	-.575	.703	.668	1	.414	.563		
Step 2(a)	sscompcode2			8.784	2	.012			
	sscompcode2(1)	2.306	.972	5.622	1	.018	10.030	1.491	67.449
	sscompcode2(2)	1.998	.721	7.678	1	.006	7.375	1.795	30.311
	numbercompgrp2	.852	.621	1.884	1	.170	2.345	.694	7.919
	Constant	-.499	.695	.516	1	.472	.607		
Step 3(a)	sscompcode2			8.270	2	.016			
	sscompcode2(1)	2.140	.945	5.133	1	.023	8.500	1.335	54.127
	sscompcode2(2)	1.852	.691	7.179	1	.007	6.375	1.644	24.714
	Constant	.000	.577	.000	1	1.000	1.000		

Proportion of trials with fully reported comparisons		0: = 0% 1: > 0%

A multivariate model was not compiled for the threshold of 0% for the outcome "completeness of reporting" as no variables were significant in univariate analysis.

Appendix 7.3: Multivariate models for completeness of documentation of the sample size calculation

Completeness of documentation of the sample size calculation		0: incomplete 1: complete or partial

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	purposecode	.150	.694	.047	1	.828	1.162	.298	4.530
	administration			6.956	2	.031			
	administration(1)	.764	.771	.981	1	.322	2.146	.474	9.722
	administration(2)	2.338	.889	6.920	1	.009	10.361	1.815	59.148
	RECdrugfundcode	.860	.654	1.728	1	.189	2.364	.655	8.523
	Constant	.162	.631	.066	1	.797	1.176		
Step 2(a)	administration			8.546	2	.014			
	administration(1)	.802	.751	1.140	1	.286	2.230	.512	9.712
	administration(2)	2.408	.830	8.421	1	.004	11.111	2.185	56.502
	RECdrugfundcode	.809	.609	1.763	1	.184	2.245	.680	7.410
	Constant	.264	.421	.394	1	.530	1.302		
Step 3(a)	administration			11.415	2	.003			
	administration(1)	.981	.732	1.796	1	.180	2.667	.635	11.194
	administration(2)	2.679	.807	11.033	1	.001	14.571	2.999	70.804
	Constant	.560	.362	2.391	1	.122	1.750		

Appendix 7.4: Multivariate model for adequacy of reporting of the power calculation

Adequacy of reporting of power calculation		0: inadequate 1: adequate

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	purposecode	1.261	.612	4.240	1	.039	3.528	1.063	11.712
	administration			4.948	2	.084			
	administration(1)	.286	.654	.191	1	.662	1.331	.369	4.799
	administration(2)	1.313	.612	4.603	1	.032	3.719	1.120	12.345
	sscompcode2			2.223	2	.329			
	sscompcode2(1)	.901	.848	1.129	1	.288	2.463	.467	12.978
	sscompcode2(2)	.969	.654	2.195	1	.138	2.636	.731	9.507
	Constant	-1.637	.706	5.383	1	.020	.195		
Step 2(a)	purposecode	1.179	.558	4.462	1	.035	3.250	1.089	9.698
	administration			8.560	2	.014			
	administration(1)	.517	.631	.672	1	.412	1.677	.487	5.777
	administration(2)	1.620	.559	8.408	1	.004	5.054	1.691	15.111
	Constant	-.978	.496	3.882	1	.049	.376		

Appendix 7.5: Multivariate model for journal type

Journal type		0: specialty 1: general

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	purposecode	.282	1.331	.045	1	.832	1.326	.098	18.012
	administration			1.779	2	.411			
	administration(1)	1.577	1.249	1.594	1	.207	4.840	.418	55.984
	administration(2)	.955	1.227	.606	1	.436	2.599	.235	28.768
	RECss200code	1.892	.915	4.275	1	.039	6.634	1.104	39.882
	RECalloccode	1.588	.579	7.524	1	.006	4.893	1.573	15.217
	Constant	-4.447	1.424	9.755	1	.002	.012		
Step 2(a)	administration			1.775	2	.412			
	administration(1)	1.578	1.247	1.600	1	.206	4.845	.420	55.843
	administration(2)	.967	1.226	.622	1	.430	2.630	.238	29.071
	RECss200code	1.991	.805	6.113	1	.013	7.322	1.511	35.488
	RECalloccode	1.597	.577	7.666	1	.006	4.938	1.594	15.295
	Constant	-4.280	1.154	13.748	1	.000	.014		
Step 3(a)	RECss200code	2.122	.679	9.777	1	.002	8.347	2.208	31.561
	RECalloccode	1.655	.560	8.743	1	.003	5.232	1.747	15.670
	Constant	-3.435	.695	24.459	1	.000	.032		

Appendix 7.6: Multivariate model for commercial funding in protocol

Commercial funding mentioned in protocol	0: no commercial funding mentioned 1: commercial funding mentioned

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1(a)	administration			4.111	2	.128			
	administration(1)	.810	.663	1.496	1	.221	2.249	.614	8.242
	administration(2)	1.086	.544	3.980	1	.046	2.962	1.019	8.606
	RECseqgencode	.555	.564	.969	1	.325	1.743	.577	5.266
	RECblindcode	.952	.473	4.057	1	.044	2.592	1.026	6.548
	sscompcode2			1.640	2	.440			
	sscompcode2(1)	.924	.797	1.345	1	.246	2.520	.528	12.022
	sscompcode2(2)	.827	.697	1.406	1	.236	2.286	.583	8.968
	Constant	-1.453	.619	5.506	1	.019	.234		
Step 2(a)	administration			7.004	2	.030			
	administration(1)	.997	.641	2.420	1	.120	2.709	.772	9.507
	administration(2)	1.325	.511	6.739	1	.009	3.764	1.384	10.238
	RECseqgencode	.715	.531	1.812	1	.178	2.043	.722	5.784
	RECblindcode	.918	.460	3.981	1	.046	2.505	1.016	6.173
	Constant	-.911	.421	4.686	1	.030	.402		
Step 3(a)	administration			7.832	2	.020			
	administration(1)	.957	.636	2.265	1	.132	2.604	.749	9.060
	administration(2)	1.399	.504	7.692	1	.006	4.050	1.507	10.883
	RECblindcode	.958	.456	4.414	1	.036	2.608	1.066	6.376
	Constant	-.767	.401	3.654	1	.056	.464		

Appendix 8: Systematic review tables

Appendix 8.1: Characteristics of included studies

Study	Methods	Participants	Interventions	Outcomes	Notes
Ahmed 1996 (Ahmed & Nicholson 1996)	Case study	A single study submitted to an LREC and then 36 other LRECs Year: 1989-1990 Country: England	Local REC/s	Time to approval Expedited review v full committee	Objectives: to compare the practices of LRECs and the time they take to obtain ethical approval for a multi-centre study
al Shahi 1999 (Al-Shahi & Warlow 1999)	Case study	A single study submitted to an MREC then multiple LRECs (15) Year: 1998 Country: Scotland	Central then local REC/s	Time to receipt by EC to LREC meeting, initial LREC decision and final LREC approval Objections Resources	Objectives: not stated
Ashford 1987 (Ashford 1987)	Case study	A single study submitted to an independent ethics committee (organised by 3 pharmaceutical companies) then 25 LRECs Year: 1984-1986 Country: England	Central then local REC/s	Approval Conditions of approval	Objectives: not stated
Bennett 2001 (Bennett et al. 2001)	Case study	Single study (retrospective medical record review) submitted to 104+ IRBs in 1990 and a follow-up of this study submitted to 74+ IRBs in 1997 Year: 1990 and 1997 Country: USA	Local REC/s	% IRBs approving study	Objectives: not stated
Benster 1993 (Benster & Pollock 1992;Benster & Pollock 1993)	Case study	A single study submitted to 28 LRECs Year: 1991 Country: England	Local REC/s	Time to approval Conditions of approval	Objectives: not stated

Boyce 2002 (Boyce 2002b)	Case series (member of MREC reviewing his records)	353 new applications to a single MREC Year: between Oct 1997 and Nov 2000 Country: England	Central REC	Approval status (first meeting) Changes requested Other outcomes collected but not relevant to this review	Objectives: Individual did not have access to complete MREC records. It was not possible to determine how cases were selected
Burman 2003 (Burman et al. 2003)	Case series	2 trials submitted to 25 sites Year: not reported Country: USA and Canada	Local REC/s	Time to approval Resources Changes requested (specifically changes to the consent form and impact on readability)	Objectives: not stated
Christian 2002 (Christian, Goldberg, Killen, Abrams, McCabe, Mauer, & Wittes 2002)	Case series	20 studies submitted to a central IRB with expertise in specific disease Year: 2001 Country: USA	Central REC	Expedited review v full review Approval	Objectives: not stated Pilot of CIRB - limited data available.
Dal Re 1999 (Dal-Re, Espada, & Ortega 1999)	Case series	First 100 applications for multicentre drug trial protocols submitted by a pharmaceutical company (SKB) in particular year 15 protocols submitted to 41 RECs Year: 1995 Country: Spain	Local REC/s	Time to approval	Objective: to review the characteristics and performance of research ethics committees in Spain in the evaluation of multicentre clinical trial drug protocols
Dockerty 1992 (Dockerty & Elwood 1992)	Case study	Single study submitted to 14 RECs (approval obtained from 1 then submitted to remaining 13) Year: 1990 Country: New Zealand	Local REC/s	Time to approval Changes requested	Objectives: not states
Druml 1999 (Druml, Svolba, Singer, Bonkovsky, & Bauer 1999)	Case series	All studies submitted to ethics committee of the medical faculty of the University of Vienna	Local REC/s	Approval	Objectives: not stated

		Year: 1993-1997 Country: Austria			
Dunn 2000 (Dunn, Arscott, & Mann 2000b)	2 Case studies	One study submitted to 197 LRECs in 1996-1997 and a second study submitted to 1 MREC then 26 LRECs in 1999. Restricted to 21 LRECs that reviewed both studies. Year: 1996-7 and 1999 Country: England	Local REC/s Central then local REC/s	Number of copies of application Fee charged	Objectives: not stated
Faccini 1984 (Faccini, Bennett, & Reid 1982)	Case series	First 294 protocols submitted Year: commenced 1977 Country: 9 European countries	Central REC	Decision made	Objectives: not stated NOTE: ethical review committee established and funded by pharmaceutical companies
Garfield 1995 (Garfield 1995)	Case study	Single study submitted to an LREC and (following approval) then submitted to 13 other LRECs within a region Year: not stated Country: England	Local REC/s	Time to approval Queries Changes requested	Objectives: not stated
Gray 1975 (Gray 1975)	Case series	79 projects in selected medical specialties submitted to an institutional REC Year: 1969-1970 Country: USA	Local REC/s	Approval Type of objection Impact of REC on proposed research Impact of REC on conduct of research	Objectives: not stated (possibly declared in methodology paper)
Hotopf 1995 (Hotopf, Wessely, & Noah 1995)	Case study	Single study submitted to 6 LRECs Year: not reported Country: England	Local REC/s	Time to reply Approval Changes requested	Objectives: not stated
Humphreys 2003 (Humphreys, Trafton, & Wagner 2003)	Case study	Single study submitted to 8 LRECs Year: not reported Country: USA	Local REC/s	Resources	Objectives: not stated Published as letter.

Jamrozik 1999 (Jamrozik & Kolybaba 1999)	Case study	Single study submitted to a University EC then 29 IECs Year: 1997 Country: Australia	Local REC/s	Time to reply	Objectives: not stated
Keinonen 2001 (Keinonen, Nieminen, Saareks, Saano, & Ylitalo 2001)	Case series	All (666) protocols of clinical studies on medicinal products submitted to 2 University hospitals Year: 1992, 1994, 1996, 1998 Country: Finland	Local REC/s	Descriptive data (study phase, subjects, design, etc) Approval Questions raised / modifications requested	Objectives: to determine the amount and type of deficiencies and questions the ECs pointed out to investigators, and find out if the ECs had been notified of study completion (or not)
Larcombe 1999 (Larcombe & Mott 1999)	Case study	Single study submitted to MREC then 51 LRECs Year: 1998 Country: England	Central then local REC/s	Time to response (not defined) resources Changes requested	Objectives: not stated Note: new MREC system introduced half way through study
Lewis 2001 (Lewis, Tomkins, & Sampson 2001)	Case study	Single-centre study involving geographically dispersed subjects submitted to MREC then 53 LRECs Year: 1998 Country: Wales	Central then local REC/s	Time to approval Volume of paperwork Resources required Expedited review v full committee Changes requested	Objectives: to assess the process involved in obtaining ethical approval
Lux 2000 (Lux, Edwards, & Osborne 2000)	Case study	Single study submitted to MREC then 99 LRECs Year:1998-1999 Country: England	Central then local REC/s	Expedited review v full committee Time to approval (response time) Approval status	Objectives: not stated
McWilliams 2003 (McWilliams et al. 2003)	Case study	A single genetic epidemiology study submitted to 42 centres Year: not reported Country: USA	Local REC/s	Expedited review v full committee Time to approval Resources Changes requested	Objectives: to assess the burden imposed by (IRB) review on multicentre studies by submitting a common protocol to multiple IRBs. Participating centres were surveyed. Note: reports on a number of risk factors
Nafria 1998 (Nafria,	Case series	22 trials sponsored by a	Local REC/s	Time to approval	Objectives: to describe the time

Ferragud, & Navarro 1998)		single drug company submitted to 107 LRECs Year: 1993-1996 Country: Spain			required to obtain ethics approval for a clinical trial in Spain Note: also reports the results of a survey conducted to ascertain the fees charged for submission to an ethics committee Note: published in Spanish
Ortega 1995 (Ortega & Dal Re 1995a)	Case series	First 10 protocols submitted by a drug company to 26 clinical trials committees (same as RECs) Year: from July 1992 with approvals before July 1993 Country: Spain	Local REC/s	Time to approval Clarifications / modifications requested	Objectives: not stated Subgroups: number of REC mem bers, Frequency of meetings, Advance fee (yes/no)
Penn 1995 (Penn & Steer 1995)	Case study	1 trial submitted to 26 RECs Year: not reported Country: UK	Local REC/s	Time to approval (although not defined and possible problems) Approval status (overall)	Objectives: not stated
Redshaw 1996 (Redshaw, Harris, & Baum 1996)	Case study	A single study submitted to multiple LRECs Year: not reported Country: England	Local REC/s	Time to approval Approval Changes requested	Objectives: not stated
Sandhu 2001 (Sandhu & Okasha 2001)	Case study	Two studies submitted MREC then 225 and 137 LRECs Year: not reported Country: UK	Central then local REC/s	Resources	Objectives: not stated
Smith 1994 (Smith et al. 1994)	Case study	Single case-control study submitted to 15 IECs Year: not reported Country: Australia	Local REC/s	Approval status first meeting Resources	Objectives: not stated
Stair 2001 (Stair et al. 2001)	Case study	Single (collaborative group) study submitted to 44 IRBs	Local REC/s	Time to approval Modifications requested	Objectives: To describe IRB responses to one standard protocol and thereby gain

		Year: 1998 Country: USA and Canada			insight into the advantages and disadvantages of local IRB review
Tully 2000 (Tully, Ninis, Booy, & Viner 2000)	Case study	Single study submitted to MREC then 125 LRECs Year: 1998 Country: England	Central then local REC/s	Time to reply Time to approval Changes (number of)	Objectives: to assess the function of the new system of review by MRECs and to highlight areas where improvement is still needed Subgroup: Executive sub-committee v not
Watling 1993 (Watling & Dewhurst 1993)	Case study	Single study submitted to RCGP REC then 26 LRECs Year: not reported Country: England	Central then local REC/s	Time to approval Accepting central approval v requiring local review Approval Changes requested / queries	Objectives: not stated
Wise 1996 (Wise & Drury 1996)	Case series	First 100 general practice based multicentre research projects Year: 1984-1989 Country: UK Royal College of General Practitioners clinical research ethics committee	Central REC	Approval Amendments Reasons for non-approval Study success (started recruitment, recruitment, publication)	Objectives: to assess the outcome of 100 general practice based, multicentre research projects submitted to the ethics committee of the Royal College of General Practitioners by pharmaceutical companies or their agents between 1984 and 1989

Appendix 8.2: Characteristics of excluded studies

Study	Reason for exclusion
Bortolussi 2002 (Bortolussi & Nicholson 2002)	Reports on the results of audits of the procedures used by 6 RECs in Canada performed between 1992 and 2000
Bruinsma 1999 (Bruinsma, Venn, & Skene 1999)	Insufficient data reported. Reported experience with 2 cohort studies.
Crooks 1996 (Crooks, Colman, & Campbell 1996)	Reports on questionnaire sent to 60 physicians who had not entered patients or had withdrawn participation from 2 multi-centre trials. 17 reported reason for non-participation as difficulty obtaining local ethics approval
Dickersin 1992 (Dickersin, Min, & Meinert 1991)	A follow-up study of 737 studies approved by two RECs. Does not report outcomes of relevance to this review.
Easterbrook 1991 (Easterbrook, Berlin, Gopalan, & Matthews 1991)	A retrospective survey of 487 research projects approved by a REC. Does not report outcomes of relevance to this review.
Easterbrook 1992 (Easterbrook & Matthews 1992)	A follow-up study of all (720) research protocols approved by a REC. Does not report outcomes of relevance to this review.
Goldman 1982 (Goldman & Katz 1982)	3 deliberately flawed protocols were submitted to IRBs at 22 academic institutions in the USA. The authors expected to find that different IRBs reached similar judgements on identical protocols and individual IRBs would apply the same standards to each protocol. They found that IRBs were consistent in their non-approval of the protocols, were inconsistent in the reasons offered in support of similar decisions, and inconsistent in the application of ethical, methodological and informed consent standards.
Gray 1978 (Gray, Cooke, & Tannenbaum 1978)	Interviews with research investigators whose proposals had been reviewed, IRB members and subjects who had consented to take part in research approved by 61 (of 420) institutions with review committees between 1974 and 1975. The committees were all approved by the US Department of Health, Education and Welfare. The study investigated IRBs (composition, policies and procedures), modifications of research proposals, risks and benefits of approved research, selection of subjects in approved projects, informed consent, involvement of IRBs after initial review, the performance of IRBs and the attitudes of research subjects.
Hahn 2002 (Hahn, Williamson, & Hutton 2002)	A follow-up study of all 56 submissions to an LREC in the UK in a particular year (year not specified). Did not collect outcomes relevant to this review (study to investigate within-study selective reporting).
Holley 1998 (Holley & Foster 1998)	27 researchers were sent a questionnaire asking for their views on the substance of ethical review and their experiences of the process of ethical review.
Keinonen 2002 (Keinonen et al. 2002)	This study aimed to investigate the validity of clinical drug study notifications by the regulatory agency in Finland during the 1990s. The overlap with ethics committees was recognised but not evaluated.
Lardot 2002 (Lardot et al. 2002)	Reports on a central protocol review committee of a disease-specific European organisation that functions as an internal review of ideas in development, not as a scientific review as part of the ethics approval process.
Lynn 1994 (Lynn, Johnson, & Levine 1994)	Survey of 50 institutions not selected to participate in project. Opportunity "to examine how a number of teaching hospitals responded to informed consent issues in health services research".
McNay 2002 (McNay et al. 2002)	International, US-based trial where all sites (including non-US sites) required assurance of compliance with the Code of Federal Regulations with the Office of Human Research Protection of the US Department of Health and Human Services. Once available, sites could submit study for consideration for ethics approval.
Nicholl 2000 (Nicholl 2000)	Single study submitted to MREC then unspecified number of local RECs in the UK. Did not report any usable

	outcome data
Pelerin 1992 (Pelerin & Hall 1992)	Single epidemiological study submitted to a national ethics committee on condition that district committees be approached. An advertisement was placed in a relevant publication to which 47 RECs responded.
Pich 2003 (Pich, Carne, Arnaiz, Gomez, Trilla, & Rodes 2003)	Survey sent to the principal investigators of protocols submitted to the REC in 1997 Objective: to assess the outcome of all protocols submitted to the HCEC (Hospital Clinic ethics committee) during 1997
Silverman 2001 (Silverman, Hull, & Sugarman 2001)	Single study submitted to 16 IRBs in the USA. Objective: to determine the extent of variability among IRBs on their approved research practices within the context of a specific multi-centre trial. Specifically reported the various practices relating to informed consent.
Stone 1997 (Stone & Blogg 1997)	Insufficient data reported. Report of 1 site's experience with 1 multi-centre trial and their local REC.
Takayanagi 2001 (Takayanagi et al. 2001)	Experience of a single ethics committee in Tokyo (The University of Tokyo Hospital) that introduced a prior review system for clinical trials in that institution.
Tappin 1992 (Tappin & Cockburn 1992)	Study that aligns dates of REC approval with incidence of HIV in each district.
Thornquist 2002 (Thornquist et al. 2002)	Proposes using single-study cooperative agreements to facilitate approval of multi-centre trials, specifically to facilitate approval of post-hoc sub-studies conducted on the large datasets collected as a result of multi-centre clinical trials. Does not report outcome data.
While 1995 (While 1996)	43 district ethics committees were asked to indicate whether a proposed study required formal ethical approval.

Appendix 8.3: Data tables: systematic review

Table 72: Impact on clinical research: Changes requested following first meeting of REC (Central REC)

Study	Nature of queries
Boyce 2002	Of the 339 applications conditionally approved, deferred or rejected at first meeting: - 287 (85%) queries related to information leaflet (such as inadequate information, jargon, poor clarity, need for proofreading) - 171 (50%) queries related to study design (such as use of placebo, stopping current treatment, choice of comparator drug)
Wise 1996	Of the 100 studies (applications) 82 were eventually approved of which 45 required amendment and resubmission, with an average of 1.5 amendment items per protocol. For 15 resubmissions chairman's approval was given (subject to later endorsement of the parent committee). Reasons for amendment (number of studies (% of requests for amendment)) were: - safety aspect or inappropriate drug dose: 15 (33%) - remuneration considerations: 14 (31%) - logistics or cost of pathology: 9 (20%) - inadequate information sheet: 8 (18%) - statistical clarification or modification: 4 (9%) - imprecise diagnostic criteria: 4 (9%) - age limit considerations: 3 (7%) - pregnancy safety aspects: 2 (4%) - no run-in phase: 1 (2%) - other: 3 (7%)

Table 73: Impact on clinical research: Changes requested following first meeting of REC (Local REC/s)

Study	Nature of queries
Burman 2003	Median of 46.5 changes per consent form (IQR 11-62) - small changes (involving <1 sentence): median 26 (IQR 8-45) - large changes (1 or more contiguous sentences): median 10 (IQR 5-21) Analysed consent forms in detail and found the IRB changes made consent forms less readable
Dal Re 1999	Type and number of queries for phase II/phase III trialss Protocol related issues: - study design: 0/3 - selection criteria 3/0 - study procedures 1/5 - statistics: 2/1 - other: 1/1 Informed consent: - wording of the patient information sheet: 5/3 - consent form sign-off: 3/1 Insurance cover: 1/3 Medication: 5/0 others: 3/5 TOTAL: 24/22
Garfield 1995	Median number of changes requested 1.5 (range 0-4)
Hotopf 1995	5 of the 6 RECs requested changes.
Keinonen 2001	Site 1 - no complementary questions = 158 (55%) - study protocol only = 14 - + subject information + consent form = 109 - other = 4 Site 2 - no complementary questions = 298 (78%) - study protocol only = 22 - + subject information + consent form = 48 - other = 13
McWilliams 2003	Mean number of changes requested: - expedited review = 5.7 - full review = 8.6
Ortega 1995	There were 13 requests for clarification or modification for 6 of the 10 protocols by 8 RECs. These related to: - Design / methodology Re: objectives/endpoint 6 items in 4 protocols by 20 RECs Re: sample size: 1 item in 1 protocol by 5 RECs Re: other: 1 item in 1 protocol by 4 RECs - Additional information on investigational drug: 1 item in 1 protocol by 4 RECs - Miscellaneous: 4 items in 4 protocols by 4 RECs
Stair 2001	The 44 IRBs requested an aberage of 3.5 changes, categorised as: - logistics and supervision: 20 (45%), most commonly the use of a different application form (32%) - research protocol: 19 (43%), most commonly the inclusion/exclusion criteria (9%) and issues to do with patient recruitment (9%) - consent form: 40 (91%): 30 (68%)grammer, syntax, readability; 15 (34%) risk information about the intervention; 12 (27%) risk information not part of trial.

Table 74: Impact on clinical research: Changes requested following first meeting of REC (Central then local REC)

Study	Nature of queries
Tully 2000	52 RECs (42%) approved without change. - 64% of RECs with an Executive subcommittee v 53% of those without gave uncontested approval (ie approved without change) (RR 1.4, 95% CI 0.9-2.1, p=0.23) Of the RECs asking for amendments 67% asked for non-local amendments. Non-local changes requested were: - changes to information sheets: 22 (18%) - changes to consent procedures: 12 (10%) - changes to questionnaire: 9 (7%) - changes to methods of recruiting subjects: 9 (7%) - changes to protocol (5 (4%) - confidentiality issues: 4 (3%) Local changes requested were: - local staff to be notified: 11 (9%) - ethnic mix to be considered: 10 (8%) - local investigator to be identified: 6 (5%) - information sheets to be on locally headed paper: 4 (3%) Various other reasons accounted for 2% or less each.
Watling 1993	Responses from 26 local RECs: Accepted central approval: 7 (27%) Required local review 19 (73%) - changes to patient information leaflet 4 (21%) - toxicity/safety of trial drug 3 (16%) - efficacy of trial drug/comparator 2 - need for chest radiograph 2 - suitability of investigator 1 - confidentiality 1 - indemnity 1 - financial 1 - none 8

Table 75: Time to accrue patients: Local REC/s

Study	Details
Bennett 2001	Authors estimate that the IRB process resulted in a delay in conducting and reporting the first study by 6 years, and the follow-up study by 3 years.

Table 76: Time to accrue patients: Central then local REC

Study	Details
Tully 2000	Authors state that delay receiving ethics approval "had a significant effect on study commencement and recruitment" and the 17% of patients referred at the time of reporting "were not recruited because ethical approval had not been granted by the relevant research ethics committee".

Table 77: Ethics committee decision: Time to approval (Central REC)

Study	Definition	Results	Notes
Boyce 2002	Time from first meeting to approval	Median 64 days (range 7-386)	Information only available for 266 of the 353 applications
	Time from conditional approval to approval	Median 62 days (range 21-408)	for the 122 applications deferred or rejected at the first meeting

Table 78: Ethics committee decision: Time to approval (Local REC/s)

Study	Definition	Results	Notes
Ahmed 1996	Time from submission to: i) receiving approval ii) chairman's approval iii) full committee approval	i) Range 6 - 208 days ii) Mean 35 days; range 6-70 days iii) Mean 77 days; range 18-208 days One third of RECs were unable to approve the study within 3 months, and 3 RECs took longer than 6 months	Submission date described as date application sent. No description of "receiving approval".
Benster 1993	Time from request for application form to obtaining written ethical approval i) all RECs ii) RECs at teaching hospitals iii) RECs at non-teaching hospitals	i) mean 11.9 weeks ii) <4 to 16 weeks iii) <4 to >20 weeks	
Burman 2003	Time (days) from the date the centrally approved protocol and consent form was sent from the CDC to local sites, to the date the locally approved consent form was submitted back to the CDC for review	Median 104.5 days (range 31-346) (similar for both studies)	"CDC = Center for Disease Control "Could not determine how much of the time required to obtain local approval was needed for the local IRB review versus the time spent by the local study site to prepare the IRB submission and submit the completed review back to the CDC
Dal Re 1999	Time from submission to arrival of REC's decision form i) all 100 applications ii) for applications without queries raised iii) for applications with queries raised	i) Mean 87 (SD 54) days Median 70 days Range 23-238 days ii) Mean 78 (SD 46) days Median 69 days Range 23-238 days iii) Mean 95 (SD 61) days Median 75 days Range 24-236 days	Submission date not defined. Also reports for the following subgroups: - Evaluation fee (Y or N) - Number of REC members (>=13 or <13) - high v lower volume RECs
	Time from	i) Mean 64 (SD 51) days	

	submission to approval i) all 100 applications ii) for applications without queries raised iii) for applications with queries raised	Median 46 days Range 1-231 days ii) Mean 56 (SD 42) days Median 42 days Range 1-231 days iii) Mean 78 (SD 61) days Median 50 days Range 10-229 days	
	Time from approval to arrival i) all 100 applications ii) for applications without queries raised iii) for applications with queries raised	i) Mean 21 (SD 20) days Median 14 days Range 1-104 days ii) Mean 22 (SD 20) days Median 15 days Range 1-102 days iii) Mean 17 (SD 21) days Median 10 days Range 1-104 days	
Dockerty 1992	Time to approval (not defined)	Range 2 - 22 weeks	
Garfield 1995	Time from application to final approval (not defined)	Median 9 weeks; range 2-14 weeks	
	Time from application to REC meeting	Median 4 weeks; range 0.3-8.6 weeks	
	Time from REC meeting to reply	Median 0.4 weeks (3 days); range 0.1-3 weeks	
Hotopf 1995	Time to reply (not defined	mean 47 days; range 15-125 days	
Jamrozik 1999	Time from initial application to receipt of decision from REC	Public Hospitals (n=18): median 6.5 weeks (range 2-11 weeks, 75th centile 8.6 weeks) Private Hospitals (n=11): median 5 weeks (range 1-52 weeks, 75th centile 11 weeks)	
McWilliams 2003	Time to obtain approval (not defined)	Expedited review: Mean 32.3 days (range 9-72 days) Full review: Mean 81.9 days (range 13-252 days)	Authors stated that "Days to approval, an indicator of the difficulty of review, correlated with the number of changes requested when both review types were combined (p=0.004) and with full review (p=0.01) when the data were stratified by review type. No other significant correlations were observed."
Ortega 1995	time from submission (by investigator) to approval (not defined)	Mean 45 (SD 30) days Median 40.5 days Range 7-110 days	Also reports for the following subgroups: - number of members (<9 or >= 9)

			- Frequency of meetings (monthly or variable) - Advance fee (yes or no) - trial phase (II or III)
	time from submission to receipt of approval in investigator's office	Mean 58 (SD 35) days Median 52 days Range 9-136 days	
Penn 1995	Time to approval (not defined)	Mean 13 weeks Median 9 weeks Range 3-30 weeks	"A major cause of delay was difficulty in obtaining the application forms. Acquisition took an average of 9 weeks, and in one case it took 38 weeks."
Redshaw 1996	Time from submission to approval (not defined) i) all applications ii) teaching districts iii) other districts	i) Mean 109 days; range 22-298 days ii) Mean 102 days; range 36-149 days iii) Mean 111 days; range 22-298 days	
Stair 2001	Time from submission to final IRB approval (no terms were defined)	Median 38 days (IQR 26-62 days)	

Table 79: Ethics committee decision: Time to approval (Central then local REC)

Study	Definition	Results	Notes
al Shahi 1999	Time from receipt of application by an REC to final REC approval. Date of receipt defined as next working day after its postage by first-class mail.	Median 39 days (range 21-209 days)	
	Time from receipt of application by an REC to first REC meeting. Date of receipt defined as next working day after its postage by first-class mail.	Median 28 days (range 14-97 days)	The authors report that the expected time-scale for MREC-approved applications is that an REC meeting should be called within 2 weeks of receipt of an application, and a decision communicated to the applicant within 5 working days of a subcommittee meeting.
Lewis 2001	Time for approval (not defined)	Median 39 days (range 12-132 days) - Executive Committees (fast track) median 34 days - No executive median 42 days 7 RECs responded within the recommended 21 days	
Lux 2000	Response time: number of days between arrival of the submission and the date on which written confirmation of the committee's decision was typed	In all sites (n=99) Median 28 days (5th, 95th centiles 4, 73 days) 33 (33%) responded within 21 days.	The authors report that the relevant guidelines suggest an upper limit of 21 days or less for response time.
		In sites with an Executive subcommittee (n=44): Median 30 days (5th, 95th centiles 4, 85 days) 14 (32%) responded within 21 days	
		In sites without an Executive subcommittee (n=55): Median 25 days (5th, 95th centiles 7, 64 days) 19 (35%) responded within 21 days.	
Tully 2000	Response time (not defined - refers to guidelines)	In all sites (n=125) - 39 (31%) replied within 21 days In sites with an Executive subcommittee (n=50): - 40% replied within 21 days In sites without an Executive subcommittee (n=75):	The authors report that the relevant guidelines suggest an upper limit of 21 days or less for response time. There was no

		- 25% replied within 21 days	statistically significant difference in response times between RECs with and without an Executive.
	Time to approval (not defined)	In sites with an Executive subcommittee (n=50): - Median 28.5 days In sites without an Executive subcommittee (n=75): - Median 46 days	RECs with an Executive subcommittee were significantly quicker to give approval than those without (p=0.0002)
Watling 1993	Time from initial contact to approval (from the time the REC received the application) i) for committees accepting central approval ii) for committees requiring full local submission	i) mean 54.7 days; range 14-84 days ii) mean 91 days; range 43-168 days	The time period reported only includes time attributable to the process of review. Time spent preparing or submitting data and postal time excluded.

Table 80: Approval status at first meeting (Central REC)

Study	Total	approved no change	conditional approval	deferred	rejected	Other
Boyce 2002	353	14 (4%)	217 (62%)	103 (29%)	19 (5%)	
Christian 2002	20	0	19 (2 of which were substantive revisions)			1 still being reviewed at time of publication. Assumed that these were decisions made at first meeting: time period not stated.

Table 81: Approval status at first meeting (Local REC/s)

Study	Total	approved no change	conditional approval	deferred	rejected	Other
Benster 1993	28	17	11	0	0	
Dal Re 1999	100	59	38	0	3	
Dockerty 1992	14	10	4			
Druml 1999	1531 applications over 5 years	not reported	not reported	664 (43.4%)	not reported	
Keinonen 2001	Site 1: 195 Site 2: 381	130 (46%) 234 (61%)	Approved with comments: 28 (10%) 63 (17%) Request for Modifications: 98 (34%) 55 (14%)	Pending 29 (10%) 6 (2%)	0 23 (6%)	
Redshaw 1996	24	14	6		3	
Smith 1994	15	4	- 6 requested changes to consent forms and/or information letters - 5 requested changes to protocol (unclear if these overlap)		1	
Stair 2001	44	4 (9%) plus 4 (9%) approved with only minor	26 (59%) returned to applicant once for revision			

		consent form changes requested	7 (16%) returned twice 2 (5%) returned 3 times 1 (2%) returned 4 times			

Table 82: Approval status at first meeting (Central then local REC/s)

Study	Total	approved no change	conditional approval	deferred	rejected	Other
Ashford 1987	21	est 8	est 10 (not clear if changes listed relate to overlapping studies)		3	
Larcombe 1999	51	23	12 (minor changes to letters or consent forms)	16		
Lux 2000	99 - fast track n=44 - standard n=55	82 (83%) - 35 (80%) - 47 (85%) Note: Number (%) of submissions approved after first review by REC)	included with approved no change	all those not approved at first REC meeting		

Table 83: Final approval status (Central REC)

Study	Total	approved	not approval	deferred	rejected
Boyce 2002	353	330 (93%)	9 (3%) not pursued by researchers	0	1 (0.3%)
Faccini 1984	294	37 without amendment 243 with amendment			14

Table 84: Final approval status (Local REC/s)

Study	Total	approved	not approval	deferred	rejected
Bennett 2001	In 1990: 104 Follow-up study in 1997: 74	87% 77%	not reported	not reported	not reported
Druml 1999	1531	not reported	not reported	117 (7.6%)	19 (1.5%)
Penn 1995	26	22	4 - 2 local REC demand for changes to protocol - 1 no response from REC - 1 trial abandoned before issues resolved		

Table 85: Final approval status (Central then local REC)

Study	Total	approved	not approval	deferred	rejected
Larcombe 1999	51	50			1

Table 86: Resource issues

Study	Details
Ahmed 1996	Estimate cost to researchers of completing applications to be GBP25.5 per district, GBP900 for the study. This includes cost of photocopying, postage, telephone, travel, time of research worker
al Shahi 1999	Median number of copies of each application required 10 (range 1-18) Paper: 5,789 pages weighing 26.9 kg Photocopying and printing: GBP231.56 Postage: GBP77.15 Person-hours: salaried secretary, research fellow and research projects coordinator
Burman 2003	Study sites estimated that the process of obtaining initial local approval required a median of 30 hours of staff time (range 10-48 hours)
Dal Re 1999	Mean number of copies of complete application required 6 (range 1-16, medican 4) and 9 of the protocol alone (range 2-23, median 10)
Dunn 2000	Median number of copies of each application required for Study 1 was 13 (range 1-34) and for Study 2 was 4 (range 1-16). Number of pages: Study 1 = about 1000 sides of A4; Study 2 = mean 450 sides of A4 Fees charged by RECs: Study 1 = 14 of 20 LRECs (70%) did not charge and the maximum amount charged was GBP500; Study 2 = 11 of 21 LRECs (52.4%) did not charge and the maximum amount charged was GBP940. Note: Study 2 had previously been seen by an MREC before submission to each local REC
Humphreys 2003	Estimated cost of each IRB action (initial review, continuing review, amendments, adverse event reports) to estimate the money spent on IRB review after the "home" IRB had approved the study. Est $US56,191 in 2001 dollars): incl $US16,951 for coordinating centre personnel, space, and supply costs.
Larcombe 1999	Estimate: - researcher's time: 2 weeks: GBP1025 - photocopying and postage: GBP400
Lewis 2001	Paper: 53 local RECs required 1-24 copies of the 62 page application, a median of 620 pages (range 62 to 1488 pages). Total pages photocopied to apply to all local RECs = 24,552. Resources: approximately 2 months of a clinical research fellow with secretarial support
Lux 2000	Local RECs with an Executive subcommittee (fast track) required a significantly lower number of copies of protocols and documents than those without. Median number of document copies required: - fast track: 3 (5th, 95th centiles 2, 13) - standard: 11 (5th, 95th centiles 1, 15)
McWilliams 2003	Mean preparation time: expedited review = mean 5.8 hours (range not stated); full review = 8.6 hours (range 2-40 hours).
Redshaw 1996	Estimated cost minimum GBP4,000 including 280 hours recorded work time, telephone calls, postage and photocopying but no overheads
Sandhu 2001	2 studies submitted to 225 and 137 local RECs, each application averaged 47 pages. Paper: estimated 109,000 sheets of paper. Estimated that 59,000 sheets could have been saved if each REC had a subcommittee of 3 people.
Smith 1994	Estimated cost over $AUD20,000 and more than one year to:

	"complete and generate multiple copies of the different application forms, questionnaires, protocols, consent forms and subject information letters and responding to IEC requests for changes".
Tully 2000	Total number of pages used: 105,888 (1,103 applications of 96 pages each). Total cost of application GBP6,132.90 - photocopying: GBP2,950 - postage: GBP1,200 - paper: GBP1,982.90

Appendix 9: SSAS invitations and data collection forms

Appendix 9.1: REC baseline form

Trial ID:

HREC Baseline Information

HREC name: _____

1. How many members does your HREC have?	
2. How often does the HREC meet	☐ Every 2nd week ☐ Every month ☐ Every 2nd month ☐ Other *(Specify _____)*
3. How long, on average, do HREC meetings last for?	_____ hours
4. How many research projects did your committee review in 2002?	
a. How many of these were clinical drug trials?	
b. How many of these clinical drug trials were multi-centre?	
5. Does your HREC have access to a committee formed to give you specialist advice on the scientific / technical aspects of clinical drug trials?	Yes ☐ No ☐
a. If yes, i. How often does the scientific committee meet?	
ii. How many people are on this committee?	
iii. How long, on average, do meetings run for?	_____ hours
6. Does you committee use experts outside your institution?	Yes ☐ No ☐
a. If yes, how many hours of expertise would you access each year?	_____ person-hours
7. How much time, on average, would the HREC Secretariat spend preparing a multi-centre drug trial for an HREC meeting? Please consider all tasks that might be involved such as photocopying, organising consultants, etc.	_____ person-hours
8. Do you assign HREC members to look at applications in more detail? a. If yes, please describe:	Yes ☐ No ☐
9. How much time, on average, would each committee member spend preparing for a meeting?	_____ person-hours
10. Are there any other issues that contribute to the workload of your HREC that you might like to mention? *(Please continue on reverse if required)*	

Completed by: _____ Date completed: ____/____/_____ (dd / mm / yyyy)

When complete please return to: The Secretariat, Shared Scientific Assessment Scheme, Health Ethics Branch, NSW Health Department, 73 Miller Street North Sydney NSW 2060

Appendix 9.2: SSAS Evaluation Form for RECs

Trial ID:

SSAS Evaluation Form for HRECs

Section 1: *To be completed by Sponsor*

1. Trial name:	
2. Sponsor:	
3. HREC name:	

Section 2: *To be completed by HREC Executive Officer*

4. Date submission received by HREC	____ / ____ / ____ (dd / mm / yyyy)
5. Date of HREC meeting at which the trial was first considered	____ / ____ / ____ (dd / mm / yyyy)

Section 3: *To be completed following discussion by the HREC (see Evaluation Protocol, p7)*

6. In the general opinion of the HREC, did the SSAS Final Report reduce the overall time taken to consider the trial at the meeting?	Yes ☐ No ☐ Uncertain ☐
Please comment.	
7. In the general opinion of the HREC, did the SSAS Final Report improve the committee's confidence in their decision?	Yes ☐ No ☐ Uncertain ☐
Please comment.	
8. Was the information provided by the SSAS Final Report useful?	1 Very useful; 2 Reasonably Useful; 3 Indifferent; 4 Not useful
Please comment.	

Section 4: *To be completed by HREC Executive Officer or Chair*

9. Decision made at the meeting
   - θ Approved
   - ف Approved with changes
   - θ Rejected
   - ف Decision pending (subject to clarification)

10. If changes requested, please summarise. Alternatively, a copy of the letter or approval can be attached to and sent to the SSAS Secretariat with this form.

11. How much time did you spend preparing this application for this meeting?	____ hours
Are there any issues that may have caused a delay in the review of this trial by your HREC (eg indemnity issues, legal issues)? Please describe.	

Thank you for participating in this Evaluation.

When complete please return to: The Secretariat, Shared Scientific Assessment Scheme, Health Ethics Branch, NSW Health Department, 73 Miller Street North Sydney NSW 2060

Appendix 9.3: SSAS Evaluation Form for Sponsors

Trial ID:	

SSAS Evaluation Form for Sponsors

1. Trial name:	
2. Sponsor name:	
3. On a scale of 1 to 10 please rate how satisfied are you with your experience with the SSAS regarding this trial? 1 Very satisfied 2 3 4 5 6 7 8 9 10 Very dissatisfied	
4. What aspects are you happy with?	
5. What aspects are you not happy with?	
6. Are there any issues that may have caused a delay in the review of the trial by HRECs (eg indemnity issues, legal issues)? Please describe.	

Completed by: _____ Date completed: ___/___/_____ (dd mm yyyy)

Thank you for participating in this Evaluation.

When complete please return to: The Secretariat, Shared Scientific Assessment Scheme, ***Research & Development Policy Branch, NSW Health Department, 73 Miller Street North Sydney NSW 2060***

Appendix 9.4: REC Log for Sponsors

Trial ID:

HREC Log for Sponsors

Please complete a separate log form for each trial. More than one log form can be completed for each trial if required.

HREC name	Date application sent to HREC	Date of actionable approval document	Are there any particular issues that influenced the time taken by this HREC to approve this trial?
	__/__/____ dd mm yyyy	__/__/____ dd mm yyyy	
	__/__/____ dd mm yyyy	__/__/____ dd mm yyyy	
	__/__/____ dd mm yyyy	__/__/____ dd mm yyyy	
	__/__/____ dd mm yyyy	__/__/____ dd mm yyyy	
	__/__/____ dd mm yyyy	__/__/____ dd mm yyyy	
	__/__/____ dd mm yyyy	__/__/____ dd mm yyyy	

When complete please return to: The Secretariat, Shared Scientific Assessment Scheme, ***Research & Development Policy Branch, NSW Health Department, 73 Miller Street North Sydney NSW 2060***

Appendix 9.5: SSAS End of Study Form: RECs

SSAS End of Study Form: HRECs

The SSAS Scheme has reached the end of its one-year pilot. We would like to ask those who have used the scheme, as well as those who chose not to use the scheme, to provide us with information that will help us to determine whether or not the scheme should continue and, if so, if there are ways in which it could be improved.

Participation in SSAS Pilot

1. HREC name (optional)

2. Did your HREC participate in the SSAS pilot? (That is, did your HREC review any trials that had been submitted to the SSAS)? Yes ☐ No ☐ Uncertain ☐

   a. If no, please tell us why

   b. If yes, did your HREC:
      - ❑ Insist sponsors/investigators use the SSAS
      - ❑ Suggest (but not insist) sponsors / investigators use the SSAS
      - ❑ Other (please specify)

Resources and Support

3. On a scale of 1 to 10 please rate how satisfied you are with the advice and support provided by NSW Health regarding the SSAS

   1 (Very dissatisfied) 2 3 4 5 6 7 8 9 10 (Very satisfied)

4. If you think improvements could be made, please specify

5. Did you feel adequately informed about the SSAS and its processes? Yes ☐ No ☐

   a. If no, in what areas did you find it inadequate?

6. Was the Manual useful? Yes ☐ No ☐ Uncertain ☐

7. Was the Web site useful? Yes ☐ No ☐ Uncertain ☐

When complete please return to: The Secretariat, Shared Scientific Assessment Scheme Health Ethics Branch, NSW Health Department, 73 Miller Street North Sydney NSW 2060

The Final Report

Please think back to the trials you have seen during the past year which have been through the SSAS process, and for which an SSAS Final Report was available.

8 Is there anything about the Final Report that you would like to see changed?

- ❑ No. All of the information required was included in the Final Report
- ❑ Yes. I would like to suggest the following changes:

The scheme

9 Would you be prepared to replace your current method of scientific assessment with the SSAS evaluation?

Yes	No	Undecided
☐	☐	☐

a If yes, please tell us why

b If no or undecided, please tell us why

10 Do you think the scheme should continue?

Yes	No	Uncertain
☐	☐	☐

11 If the scheme continues, what form should it take? (please tick all that apply)

- ❑ Continue in its current form
- ❑ Continue in an expanded form
- ❑ Other (*please specify*)

12 If you feel the scheme should continue in an expanded form, in what way do you think it should expand (tick all that apply):

- ❑ The SSAS should consider all multi-centre randomised clinical drug trials
- ❑ The SSAS should consider all multi-centre randomised trials
- ❑ The SSAS should consider all multi-centre clinical drug trials
- ❑ The SSAS should consider all multi-centre clinical research
- ❑ Other ways the SSAS should consider expanding its role:

Name of person completing this form: _____

Date: _____ / _____ / _____

Thank you for participating in this Evaluation.

When complete please return to: The Secretariat, Shared Scientific Assessment Scheme
Health Ethics Branch, NSW Health Department, 73 Miller Street North Sydney NSW 2060

Appendix 9.6: SSAS End of Study Form: Applicants

SSAS End of Study Form: Applicants

The SSAS Scheme has reached the end of its one-year pilot. We would like to ask those who have used the scheme, as well as those who chose not to use the scheme, to provide us with information that will help us to determine whether or not the scheme should continue and, if so, if there are ways in which it could be improved.

Participation in SSAS Pilot

1 Applicant name (optional)

2. Did you submit any trials for consideration by the SSAS as part of the pilot? Yes ☐ No ☐ Uncertain ☐

a If not, please tell us why

b If you participated in the pilot, how did this come about?

- ❑ At least one HREC insisted that we use the SSAS
- ❑ We chose to use the SSAS without prompting by an HREC
- ❑ Other (please specify)

Resources and Support

3. On a scale of 1 to 10 please rate how satisfied you are with the advice and support provided by NSW Health regarding the SSAS

1	2	3	4	5	6	7	8	9	10
Very dissatisfied									Very satisfied

4. If you think improvements could be made, please specify

5. Did you feel adequately informed about the SSAS and its processes? Yes ☐ No ☐

a If no, in what areas did you find it inadequate?

6. Was the Manual useful? Yes ☐ No ☐ Uncertain ☐

7. Was the Web site useful? Yes ☐ No ☐ Uncertain ☐

When complete please return to: The Secretariat, Shared Scientific Assessment Scheme Health Ethics Branch, NSW Health Department, 73 Miller Street North Sydney NSW 2060

The Final Report

Please think back to the trials you submitted to the SSAS during the past year and for which an SSAS Final Report was available.

8 Is there anything about the Final Report that you would like to see changed?

- ❑ No. All of the information required was included in the Final Report
- ❑ Yes. I would like to suggest the following changes:

The scheme

9 Do you think the scheme should continue? Yes ☐ No ☐ Undecided ☐

- a Please tell us why you think this way:

10. If the scheme continues, what form should it take? (please tick all that apply)

- ❑ Continue in its current form
- ❑ Continue in an expanded form
- ❑ Other (*please specify*)

11. If you feel the scheme should continue in an expanded form, in what way do you think it should expand (tick all that apply):

- ❑ The SSAS should consider all multi-centre randomised clinical drug trials
- ❑ The SSAS should consider all multi-centre randomised trials
- ❑ The SSAS should consider all multi-centre clinical drug trials
- ❑ The SSAS should consider all multi-centre clinical research
- ❑ Other ways the SSAS should consider expanding its role:

12 Are there any other issues you would like to raise?

Name of person completing this form: _____

Date: ____/____/_____

Thank you for participating in this Evaluation.

When complete please return to: The Secretariat, Shared Scientific Assessment Scheme
Health Ethics Branch, NSW Health Department, 73 Miller Street North Sydney NSW 2060

Appendix 9.7: SSAS End of Study Form: SSAS members

SSAS End of Study Form: SSAS

The SSAS Scheme has reached the end of its one-year pilot. We would like to ask those who have used the scheme, as well as those who chose not to use the scheme, to provide us with information that will help us to determine whether or not the scheme should continue and, if so, if there are ways in which it could be improved.

Resources and Support

1. On a scale of 1 to 10 please rate how satisfied you are with the advice and support provided by the SSAS Secretariat at NSW Health regarding the SSAS process

1	2	3	4	5	6	7	8	9	10
Very dissatisfied									Very satisfied

2. If you think improvements could be made, please specify

3. How much time, on average, did you spend preparing for each SSAS meeting (reading applications, etc)
_____ hours

4. Were the meetings (including teleconferences) held at an acceptable time?

Yes	No
☐	☐

The Final Report

Please think back to the trials you have seen during the past year which have been through the SSAS process, and for which an SSAS Final Report was available.

5. Is there anything about the Final Report that you would like to see changed?

- ❑ No. All of the information required was included in the Final Report
- ❑ Yes. I would like to suggest the following changes

When complete please return to: The Secretariat, Shared Scientific Assessment Scheme Health Ethics Branch, NSW Health Department, 73 Miller Street North Sydney NSW 2060

The scheme

6. Do you think the scheme should continue?

Yes	No	Undecided
☐	☐	☐

a. Please tell us why you think this way

7. If the scheme continues, what form should it take? (please tick all that apply)

- ❑ Continue in its current form
- ❑ Continue in an expanded form
- ❑ Other (*please specify*)

8. If you feel the scheme should continue in an expanded form, in what way do you think it should expand (tick all that apply):

- ❑ The SSAS should consider all multi-centre randomised clinical drug trials
- ❑ The SSAS should consider all multi-centre randomised trials
- ❑ The SSAS should consider all multi-centre clinical drug trials
- ❑ The SSAS should consider all multi-centre clinical research
- ❑ Other ways the SSAS should consider expanding its role:

9. If the SSAS continues, are there any changes to the process that should be made?

10. Are there any other issues you would like to raise?

Name of person completing this form: _____

Date: _____ / _____ / _____

Thank you for participating in this Evaluation.

When complete please return to: The Secretariat, Shared Scientific Assessment Scheme
Health Ethics Branch, NSW Health Department, 73 Miller Street North Sydney NSW 2060

Appendix 10: SSAS invitations to participate and consent forms

NHMRC Clinical Trials Centre
The University of Sydney

Locked Bag 77
Camperdown NSW 2050
Tel: +61-2 9562 5000
Fax: +61-2 9565 1863
E-mail: enquiry@ctc.usyd.edu.au

NSW HEALTH

SSAC Secretariat
Health Ethics Branch
NSW Health Department
73 Miller Street
North Sydney NSW 2060

Telephone (02) 9391 9854

<<current date>>

Name and
address of
chair of REC

Dear *<<name of chair>>*

Re: Evaluation of the Shared Scientific Assessment Scheme (SSAS)

In February 2003 NSW Health will commence a one-year pilot of a Shared Scientific Assessment Scheme (SSAS) for multi-centre clinical drug trials. At the end of the pilot a decision will be made as to whether or not to continue the work of the SSAS and, if so, if the remit should be expanded to include other multi-centre research. To aid NSW Health in its decision it will be necessary to evaluate the SSAS during the pilot phase. An information sheet on the evaluation has been included with this letter as well as the complete protocol should you like to read further. The protocol includes the data collection forms to be used.

As your REC may participate in the SSAS Pilot we would like to ask your REC to participate in the Evaluation of the SSAS Pilot. It is anticipated that workload involved in the Evaluation would be minimal.

Please read the REC Information Sheet and, if you agree to take part, complete the enclosed Agreement to Participate form. This should then be returned to: The Secretariat, Shared Scientific Assessment Scheme, Health Ethics Branch, NSW Health Department, 73 Miller Street North Sydney NSW 2060.

If you require any further information, have any questions or would like to discuss the evaluation further please contact the SSAS Secretariat at NSW Health by telephoning (02) 9391 9854.

Yours Sincerely,

Davina Ghersi
Research Fellow, NHMRC Clinical Trials Centre
SSAS Evaluator

NHMRC Clinical Trials Centre
The University of Sydney

Locked Bag 77
Camperdown NSW 2050
Tel: +61-2 9562 5000
Fax: +61-2 9565 1863
E-mail: enquiry@ctc.usyd.edu.au

NSW HEALTH

SSAC Secretariat
Health Ethics Branch
NSW Health Department
73 Miller Street
North Sydney NSW 2060

Telephone (02) 9391 9854

REC Agreement to Participate

in the Evaluation of the NSW Health Shared Scientific Assessment Scheme

This is to confirm that _____
(name of Sponsor)

gives permission for the trial _____
(title of trial)

to be used in the Evaluation of the NSW Health Shared Scientific Assessment Scheme.

Name: _____

Signature: _____ Date: ____/____/_____
On behalf of the Committee

When complete please return to: The Secretariat, Shared Scientific Assessment Scheme, Health Ethics Branch, NSW Health Department, 73 Miller Street North Sydney NSW 2060

Evaluation of the NSW Health Shared Scientific Assessment Scheme (SSAS)

Information Sheet for RECs

The National Statement on Ethical Conduct in Research Involving Humans released by the NHMRC in 1999 makes it possible for RECs to "accept a scientific/technical assessment of the research by another institution or organization". This clause has resulted in various models for sharing scientific or ethical review being implemented across the country.

As your REC may participate in the SSAS Pilot we would like to ask your REC to participate in the Evaluation of that pilot. The Evaluation will enable NSW Health to decide at the end of the pilot whether or not to continue the work of the SSAS and, if so, if the remit should be expanded to include other multi-centre research.

If shared scientific review is effective it could improve the quality of clinical trials research, reduce the workload of RECs by reducing unnecessary duplication of effort, and reduce the time it takes for clinical trials to start recruiting. If not effective it could actually increase the burden on researchers and RECs by adding another level of bureaucracy.

The aim of the study is to evaluate the NSW Health Shared Scientific Assessment Scheme (SSAS) by assessing its influence on i) Health Research Ethics Committees and ii) Multi-centre clinical trials. If you agree to take part you will be asked to:

1. Complete an REC Baseline Information form

2. Complete an SSAS Evaluation Form for RECs for each eligible trial submitted to your REC. The form will be sent to you by the Sponsor when they submit the trial for consideration by your REC

Individual trials, the results of individual trials, Sponsors and RECs will not be identified or identifiable in any report or publication produced as a result of the SSAS Evaluation.

Participation in the Evaluation is voluntary. If your REC agrees to take part, please complete the attached *Agreement to Participate* form and send to: The Secretariat, Shared Scientific Assessment Scheme, Health Ethics Branch, NSW Health Department, 73 Miller Street North Sydney NSW 2060. RECs may withdraw from the Evaluation at any time without penalty or prejudice.

The Evaluation is being conducted by Davina Ghersi, Research Fellow at the NHMRC Clinical Trials Centre and will form the basis for the degree of Doctor of Philosophy in Medicine at the University of Sydney under the Supervision of Professor John Simes. All aspects of the study, including results, will be strictly confidential and only Davina Ghersi and the Secretariat of the SSAS will have access to information on participants except as required by law. A report of the study may be submitted for publication, but individual participants (trials, investigators or RECs) will not be identifiable in such a report.

If you require any further information, have any questions or would like to discuss the Evaluation further please contact the SSAS Secretariat at NSW Health by telephoning (02) 9391 9854.

NHMRC Clinical Trials Centre
The University of Sydney

Locked Bag 77
Camperdown NSW 2050
Tel: +61-2 9562 5000
Fax: +61-2 9565 1863
E-mail: enquiry@ctc.usyd.edu.au

NSW HEALTH

SSAC Secretariat
Health Ethics Branch
NSW Health Department
73 Miller Street
North Sydney NSW 2060

Telephone (02) 9391 9854

<<current date>>

Name and address of
Sponsor

Dear *<<name of Sponsor representative>>*

Re: Evaluation of the Shared Scientific Assessment Scheme (SSAS)

In February 2003 NSW Health will commence a one-year pilot of a Shared Scientific Assessment Scheme (SSAS) for multi-centre clinical drug trials. At the end of the pilot a decision will be made as to whether or not to continue the work of the SSAS and, if so, if the remit should be expanded to include other multi-centre research. To aid NSW Health in its decision it will be necessary to evaluate the SSAS during the pilot phase. An information sheet on the evaluation has been included with this letter as well as the complete protocol should you like to read further. The protocol includes the data collection forms to be used.

As a Sponsor submitting a trial for consideration by the SSAS Pilot you are being asked for permission for your trial to be used in the Evaluation of the SSAS Pilot. It is anticipated that workload involved in the Evaluation would be minimal.

Please read the Sponsor Information Sheet and, if you agree to take part, complete the enclosed Agreement to Participate form. This should then be returned to: The Secretariat, Shared Scientific Assessment Scheme, Health Ethics Branch, NSW Health Department, 73 Miller Street North Sydney NSW 2060.

If you require any further information, have any questions or would like to discuss the evaluation further please contact the SSAS Secretariat at NSW Health by telephoning (02) 9391 9854.

Yours Sincerely,

Davina Ghersi
Research Fellow, NHMRC Clinical Trials Centre
SSAS Evaluator

Evaluation of the NSW Health Shared Scientific Assessment Scheme (SSAS)

Information Sheet for Sponsors

As a Sponsor submitting a trial for consideration by the SSAS Pilot you are being asked for permission for your trial to be used in the Evaluation of the SSAS Pilot. The Evaluation will enable NSW Health to decide at the end of the pilot whether or not to continue the work of the SSAS and, if so, if the remit should be expanded to include other multi-centre research.

If shared scientific review is effective it could improve the quality of clinical trials research, reduce the workload of RECs by reducing unnecessary duplication of effort, and reduce the time it takes for clinical trials to start recruiting. If not effective it could actually increase the burden on researchers and RECs by adding another level of bureaucracy.

The aim of the study is to evaluate the NSW Health Shared Scientific Assessment Scheme (SSAS) by assessing its influence on i) Health Research Ethics Committees and ii) Multi-centre clinical trials. If you agree to take part you will be asked to:

3. Distribute an SSAS Evaluation Form along with the Final Report from the SSAS with your application to each REC
4. Keep a log of SSES Evaluation Forms to make it possible to track each trial through each REC
5. Complete an SSAS Evaluation Form for Sponsors for each eligible trial submitted to the SSAS

Individual trials, the results of individual trials, Sponsors and RECs will not be identified or identifiable in any report or publication produced as a result of the SSAS Evaluation.

Participation in the Evaluation is voluntary. If you agree to take part, please complete the attached *Agreement to Participate* form and submit with your completed SSAS Application to: The Secretariat, Shared Scientific Assessment Scheme, Health Ethics Branch, NSW Health Department, 73 Miller Street North Sydney NSW 2060. Sponsors may withdraw from the Evaluation at any time without penalty or prejudice.

The Evaluation is being conducted by Davina Ghersi, Research Fellow at the NHMRC Clinical Trials Centre and will form the basis for the degree of Doctor of Philosophy in Medicine at the University of Sydney under the Supervision of Professor John Simes. All aspects of the study, including results, will be strictly confidential and only Davina Ghersi and the Secretariat of the SSAS will have access to information on participants except as required by law. A report of the study may be submitted for publication, but individual participants (trials, investigators or RECs) will not be identifiable in such a report.

If you require any further information, have any questions or would like to discuss the Evaluation further please contact the SSAS Secretariat by telephoning (02) 9391 9854.

NHMRC Clinical Trials Centre
The University of Sydney

Locked Bag 77
Camperdown NSW 2050
Tel: +61-2 9562 5000
Fax: +61-2 9565 1863
E-mail: enquiry@ctc.usyd.edu.au

NSW HEALTH

SSAC Secretariat
Health Ethics Branch
NSW Health Department
73 Miller Street
North Sydney NSW 2060

Telephone (02) 9391 9854

Sponsor Agreement to Participate

in the Evaluation of the NSW Health Shared Scientific Assessment Scheme

This is to confirm that _____
(name of Sponsor)

gives permission for the trial _____
(title of trial)

to be used in the Evaluation of the NSW Health Shared Scientific Assessment Scheme.

Name: _____

Signature: _____ Date: ____/____/_____
On behalf of the Sponsor

When complete please return to: The Secretariat, Shared Scientific Assessment Scheme, Research & Development Policy Branch, NSW Health Department, 73 Miller Street North Sydney NSW 2060

2001, *Institutional Review Board Guidebook. Office for Human Research Protections. US Department of Health and Human Services.*

Ahmed, A. H. & Nicholson, K. G. 1996, "Delays and diversity in the practice of local research ethics committees", *Journal of Medical Ethics*, vol. 22, no. 5, pp. 263-266.

Al-Shahi, R. & Warlow, C. P. 1999, "Ethical review of a multicentre study in Scotland: a weighty problem", *Journal of the Royal College of Physicians of London*, vol. 33, no. 6, pp. 549-552.

Alberti, K. G. 2000, "Multicentre research ethics committees: has the cure been worse than the disease? No, but idiosyncracies and obstructions to good research must be removed.", *BMJ*, vol. 320, no. 7243, pp. 1157-1158.

Alexander, J. I. 1999, "Multicentre research ethics committees. [letter; comment.]", *Journal of the Royal Society of Medicine*, vol. 92, no. 12, p. 662.

Altman, D. G. 1991, "Randomisation", *BMJ*, vol. 302, no. 6791, pp. 1481-1482.

Arnau, C., Cobo, E., Ribera, J. M., Cardellach, F., Selva, A., & Urrutia, A. 2003, "Efecto de la revision estadistica en la calidad de los manuscritos publicados en Medicina Clinica: estudio aleatorizado.", *Medicina Clinica*, vol. 121, no. 18, pp. 690-694.

Ashford, J. J. 1987, "Recent experience of ethics committee review of a multicentre research project", *British Journal of Clinical Pharmacology*, vol. 23, no. 3, pp. 373-374.

Assmann, S. F., Pocock, S. J., Enos, L. E., & Kasten, L. E. 2000, "Subgroup analysis and other (mis)uses of baseline data in clinical trials.", *Lancet*, vol. 355, no. 9209, pp. 1064-1069.

Australian Clinical Trials Registry. Number of trials registered. Australian Clinical Trials Registry Newsletter [Issue No. 1]. 2005.
Ref Type: Magazine Article

Australian Government. Therapeutic Goods Act No.84 of 7997. 1984.
Ref Type: Bill/Resolution

Bardy, A. H. 1998, "Bias in reporting clinical trials", *British Journal of Clinical Pharmacology*, vol. 46, no. 2, pp. 147-150.

Bennett, C. L., Sipler, A. M., Parada, J. P., Goetz, M. B., DeHovitz, J. A., & Weinstein, R. A. 2001, "Variations in institutional review board decisions for HIV quality of care studies: a potential source of study bias", *Journal of Acquired Immune Deficiency Syndrome*, vol. 26, no. 4, pp. 390-391.

Benster, R. & Pollock, A. 1992, "Ethics and multicentre research projects.", *BMJ*, vol. 304, no. 6843, p. 1696.

Benster, R. & Pollock, A. M. 1993, "Guidelines for local research ethics committees: distinguishing between patient and population research in the Multicentre Research Project", *Public Health*, vol. 107, no. 1, pp. 3-7.

Berlin, J. & Ghersi, D. 2005, "Preventing Publication Bias: Registries and Prospective Meta-Analysis," in *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*, H. R. Rothstein, A. J. Sutton, & M. Borenstein, eds., John Wiley & Sons Ltd, Chichester, West Sussex, England.

Bland, J. M., Jones, D. R., Bennett, S., Cook, D. G., Haines, A. P., & MacFarlane, A. J. 1985, "Is the clinical trial evidence about new drugs statistically adequate?", *British Journal of Clinical Pharmacology*, vol. 19, no. 2, pp. 155-160.

Bortolussi, R. & Nicholson, D. 2002, "Auditing of clinical research ethics in a children's and women's academic hospital", *Clin Invest Med*, vol. 25, no. 3, pp. 83-88.

Bouchy, C. Memoranda from Claude Bouchy, CEO for Lilly in Germany to Lilly headquarters on the coding of suicides. 13-11-1990.
Ref Type: Unpublished Work

Boyce, M. 2002a, "Observational study of 353 applications to London multicentre research ethics committee 1997-2000", *BMJ*, vol. 325, no. 7372, p. 1081.

Boyce, M. 2002b, "Observational study of 353 applications to London multicentre research ethics committee 1997-2000", *BMJ*, vol. 325, no. 7372, p. 1081.

Bruinsma, F. J., Venn, A. J., & Skene, L. L. 1999, "We need a better system for multicentre research.", *Medical Journal of Australia*, vol. 171, no. 1, p. 52.

Burman, W. J., Breese, P., Weis, S., Bock, N., Bernardo, J., & Vernon, A. 2003, "The effects of local review on informed consent documents from a multicenter clinical trials consortium", *Controlled Clinical Trials*, vol. 24, no. 245, p. 255.

Burman, W. J., Reves, R. R., Cohn, D. L., & Schooley, R. T. 2001, "Breaking the Camel's Back: Multicenter Clinical Trials and Local Institutional Review Boards", *Annals of Internal Medicine*, vol. 134, no. 2, pp. 152-157.

Chalmers, I., Adams, M., Dickersin, K., Hetherington, J., Tarnow-Mordi, W., Meinert, C., Tonascia, S., & Chalmers, T. C. 1990, "A cohort study of summary reports of controlled trials", *JAMA*, vol. 263, no. 10, pp. 1401-1405.

Chalmers, T. C., Smith, H., Jr., Blackburn, B., Silverman, B., Schroeder, B., Reitman, D., & Ambroz, A. 1981, "A method for assessing the quality of a randomized control trial", *Controlled Clinical Trials*, vol. 2, no. 1, pp. 31-49.

Chan AW 2003, *Outcome reporting bias in randomised trials: implications for systematic reviews*, DPhil, Hertford College, Oxford University.

Chan AW & Altman, D. G. 2003, *Outcome reporting bias in recent randomised trials.*

Chan AW & Altman, D. G. 2005, "Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors", *BMJ*, vol. 330, pp. 753-759.

Chan AW, Upshur, R., Singh, J. A., Ghersi, D., Chapuis, F., & Altman, D. G. 2006, "Access to research protocols for methodological review: waiving confidentiality for the greater good", *BMJ* pp. accepted 03-06.

Chan, A. W., Hrobjartsson A, Haahr, M. T., Gotzsche, P. C., & Altman, D. G. 2004a, "Empirical Evidence for Selective Reporting of Outcomes in Randomized Trials: Comparison of Protocols to Published Articles", *JAMA*, vol. 291, no. 20, pp. 2457-2465.

Chan, A. W., Krelzja-Jeric, K., Schmid, I., & Altman, D. G. 2004b, "Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research", *CMAJ*, vol. 171, no. 7, pp. 735-740.

Christian, M. C., Goldberg, J. L., Killen, J., Abrams, J. S., McCabe, M. S., Mauer, J. K., & Wittes, R. E. 2002, "A central institutional review board for multi-institutional trials.", *New England Journal of Medicine*, vol. 346, no. 18, pp. 1405-1408.

Clarke, M. J., Alderson, P., & Chalmers, I. 2002, "Discussion sections in reports of controlled trials published in general medical journals", *JAMA*, vol. 287, pp. 2799-2901.

Commonwealth Department of Health and Ageing 1999, *Report of the review of the role and functioning of institutional ethics committees*, Commonwealth of Australia, Canberra.

Commonwealth Department of Health and Ageing 2001a, *Access to unapproved therapeutic goods - clinical trials in Australia*, Therapeutic Goods Administration.

Commonwealth Department of Health and Ageing 2001b, *Human Research Ethics Committees and the Therapeutic Goods Legislation (Appendix 2)*, Commonwealth of Australia, Canberra, ACT, Appendix 2.

CONSORT Group 1996, "Improving the quality of reporting of randomised controlled trials: the CONSORT statement.", *JAMA*, vol. 276, pp. 637-639.

CONSORT Group. CONSORT web site: Statistical Guidelines. 21-1-2005. Ref Type: Generic

Cookson, J. B. 1992, "Auditing a research ethics committee", *Journal of the Royal College of Physicians of London*, vol. 26, no. 2, pp. 181-183.

Crooks, S. W., Colman, S. B., & Campbell, I. A. 1996, "Costs and getting ethical approval deter doctors from participating in multicentre trials", *BMJ*, vol. 312, no. 7047, p. 1669.

Dal Re, R., Espada, J., & Ortega, R. 1999, "Performance of research ethics committees in Spain. A prospective study of 100 applications for clinical trial protocols on medicines.", *Journal of Medical Ethics*, vol. 25, no. 3, pp. 268-273.

Dal-Re, R., Espada, J., & Ortega, R. 1999, "Performance of research ethics committees in Spain. A prospective study of 100 applications for clinical trial protocols on medicines", *Journal of Medical Ethics*, vol. 25, pp. 268-273.

De Angelis, C., Drazen, J., Frizelle, F., Haug, C., Hoey, J., & Horton, R. 2004, "Clinical trial registration: a statement from the International Committee of Medical Journal Editors.", *Annals of Internal Medicine*, vol. 141, pp. 477-478.

Demicheli, V. & Di Pietrantonj, C. 2003, "Peer review for improving the quality of grant applications", *Cochrane Database of Methodology Reviews* no. 1.

Detsky, A. S., Naylor, C. D., O'Rourke, K., McGeer, A. J., & L'Abbe, K. A. 1992, "Incorporating variations in the quality of individual randomized trials into meta-analysis", *Journal of Clinical Epidemiology*, vol. 45, no. 3, pp. 255-265.

Dickersin, K. 1990, "The existence of publication bias and risk factors for its occurrence", *JAMA*, vol. 263, no. 10, pp. 1385-1389.

Dickersin, K. & Min, Y. I. 1993, "NIH clinical trials and publication bias", *Online Journal of Current Clinical Trials*, vol. Doc No 50, p. [4967 words; 53 para.

Dickersin, K., Min, Y. I., & Meinert, C. 1991, "Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards", *JAMA*, vol. 267, no. 3, pp. 374-378.

Dockerty, J. D. & Elwood, M. 1992, "The difficulties and costs of ethical assessment of a national research project", *New Zealand Medical Journal*, vol. 105, no. 939, pp. 311-313.

Druml, C., Svolba, G., Singer, E. A., Bonkovsky, F. O., & Bauer, P. 1999, "Twenty years of the ethics committee at the medical faculty of the University of Vienna. An interim report", *Wiener Klinische Wochenschrift*, vol. 111, no. 24, pp. 1019-1026.

Dunn, N. R., Arscott, A., & Mann, R. D. 2000a, "Costs of seeking ethics approval before and after the introduction of multicentre research ethics committees", *Journal of the Royal Society of Medicine*, vol. 93, no. 10, pp. 511-512.

Dunn, N. R., Arscott, A., & Mann, R. D. 2000b, "Costs of seeking ethics approval before and after the introduction of multicentre research ethics committees", *Journal of the Royal Society of Medicine*, vol. 93, no. 10, pp. 511-512.

Easterbrook, P. & Matthews, D. R. 1992, "Fate of research studies", *Journal of the Royal Society of Medicine*, vol. 85, no. 2, pp. 71-76.

Easterbrook, P. J., Berlin, J. A., Gopalan, R., & Matthews, D. R. 1991, "Publication bias in clinical research", *Lancet*, vol. 337, no. 8746, pp. 867-872.

European Commission. 2001, "Directive 2001/20/EC of the European Parliament and of the Council of 4 April 2001 on the approximation of the laws, regulations and administrative provisions of the Member States relating to the implementation of good clinical practice in the conduct of clinical trials on medicinal products for human use.", *Official Journal of the European Communities* no. 1-5-2001, L 121/34-44.

Faccini, J. M., Bennett, P. N., & Reid, J. L. 1982, "European ethical review committee: the experience of an international ethics committee reviewing protocols for drug trials", *British Medical Journal*, vol. 289, pp. 1052-1054.

Fieller, N. 2003, *New ICH "Points to Consider" when analysing and presenting clinical trial data in European licensing applications: Multiplicity.*

Fisher, M. 1989, *Guide to Clinical Preventive Services. An Assessment of the Effectiveness of 169 Interventions. Report of the U.S. Preventive Services Task Force.* Williams and Wilkins.

Fletcher, R. H. & Fletcher, S. W. 1999, "The effectiveness of journal peer review," in *Peer review in health sciences*, Godlee F & T. Jefferson, eds., BMJ Books, London, pp. 62-75.

Freemantle, N., Calvert, M., Wood, J., Eastaugh, J., & Griffin, C. 2003, "Composite in Randomized Trials: Greater Precision But With Greater Uncertainty?", *JAMA*, vol. 289, pp. 2554-2559.

Freiman, J. A., Chalmers, T. C., Smith, H., Jr., & Kuebler, R. R. 1978, "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", *New England Journal of Medicine*, vol. 299, no. 13, pp. 690-694.

Friedman, L. M., Furberg, C. D., & DeMets, D. L. 1985b, in *Fundamentals of Clinical Trials.*, 2nd edn, PSG Publishing Company, Inc..

Friedman, L. M., Furberg, C. D., & DeMets, D. L. 1985a, in *Fundamentals of Clinical Trials.*, 2nd edn, PSG Publishing Company, Inc..

Garfield, P. 1995, "Cross district comparison of applications to research ethics committees", *BMJ*, vol. 311, no. 7006, pp. 660-661.

Ghersi, D. & Dickersin, K. 2004, "Impact of shared scientific or ethical review of multi-centre clinical research on the quality of clinical research and the clinical-research process. (Protocol for a Cochrane Methodology Review)", *The Cochrane Library* no. 1.

Godlee F 2001, "Publishing study protocols: making them visible will improve registration, reporting and recruitment", *BMC News and Views*, vol. 2, no. 4.

Goldman, J. & Katz, M. D. 1982, "Inconsistency and institutional review boards", *JAMA*, vol. 248, no. 2, pp. 197-202.

Gray, B. H. 1975, "An assessment of institutional review committees in human experimentation", *Medical Care*, vol. 13, no. 4, pp. 318-328.

Gray, B. H., Cooke, R. A., & Tannenbaum, A. S. 1978, "Research involving human subjects", *Science*, vol. 201, no. 4361, pp. 1094-1101.

Grimes, D. A. & Schulz, K. F. 2005, "Surrogate end points in clinical research: hazardous to your health.", *Obstetrics & Gynecology*, vol. 105, no. 5 Pt 1, pp. 1114-1118.

Hahn, S., Williamson, P. R., & Hutton, J. L. 2002, "Investigation of within-study selective reporting in clinical research: follow-up of applications submitted to a local research ethics committee", *Journal of Evaluation in Clinical Practice*, vol. 8, no. 3, pp. 353-359.

Halpern, S. D., Karlawish, J. H. T., & Berlin, J. A. 2002, "The continuing unethical conduct of underpowered clinical trials", *JAMA*, vol. 288, no. 3, pp. 358-362.

Healy, D. I. 2002, "Conflicting interests in Toronto: anatomy of a controversy at the interface of academia and industry", *Perspectives in Biology & Medicine*, vol. 45, no. 2, pp. 250-263.

Hemminki, A. & Kellokumpu-Lehtinen, P. 2006, "Harmful impact of EU clinical trials directive", *BMJ*, vol. 332, no. 7540, pp. 501-502.

Hendrick, J. 2001, "Legal aspects of clinical ethics committees", *Journal of Medical Ethics*, vol. 27 Suppl 1, p. i50-i53.

Holley, S. & Foster, C. 1998, "Ethical review of multi-centre research: a survey of local research ethics committees in the south Thames region", *Journal of the Royal College of Physicians of London*, vol. 32, no. 3, pp. 238-241.

Hopewell, S. & Clarke, M. 2005, "Abstracts presented at the American Society of Clinical Oncology conference: how completely are trials reported?", *Clinical Trials*, vol. 2, no. 3, pp. 265-268.

Hopewell, S., McDonald, S., Clarke, M., & Egger, M. 2002, "Grey literature in meta-analyses of randomized trials of health care interventions.", *The Cochrane Database of Methodology Reviews* no. 4.

Hotopf, M., Wessely, S., & Noah, N. 1995, "Are ethical committees reliable?", *Journal of the Royal Society of Medicine*, vol. 88, no. 1, pp. 31-33.

Hsueh, H., Chen, J. J., & Kodell, R. L. 2003, "Comparison of methods for estimating the number of true null hypotheses in multiplicity testing", *Journal of Biopharmaceutical Statistics*, vol. 13, no. 4, pp. 675-689.

Humphreys, K., Trafton, J., & Wagner, T. 2003, "The cost of institutional review board procedures in multicenter observational research", *Annals of Internal Medicine*, vol. 139, no. 1, p. 77.

International Conference on Harmonisation 1996, *Guideline for Good Clinical Practice (E6)*.

International Conference on Harmonisation 1998, *ICH Harmonised Tripartite Guideline: Statistical Principles for Clinical Trials (E9)*.

Jadad, A. 1998, *Randomised Controlled Trials*, 1 edn, BMJ Books, London.

Jadad, A. R., Moore, R. A., Carroll, D., Jenkinson, C., Reynolds, D. J., Gavaghan, D. J., & McQuay, H. J. 1996, "Assessing the quality of reports of randomized clinical trials: is blinding necessary?", *Controlled Clinical Trials*, vol. 17, no. 1, pp. 1-12.

Jamrozik, K. & Kolybaba, M. 1999, "Are ethics committees retarding the improvement of health services in Australia?", *Medical Journal of Australia*, vol. 170, no. 1, pp. 26-28.

Jefferson, T., Rudin, M., Brodney Folse, S., & Davidoff, F. 2006, "Editorial peer review for improving the quality of reports of biomedical studies", *Cochrane Database of Methodology Reviews* no. 1.

Jones, G. & Abbasi, K. 2004, "Trial protocols at the BMJ", *BMJ*, vol. 329, no. 7479, p. 1360.

Keinonen, T., Nieminen, S., Saareks, V., Miettinen, P., Saano, V., & Ylitalo, P. 2002, "The quality and characteristics of clinical drug study notifications reviewed by the regulatory agency in Finland", *Controlled Clinical Trials*, vol. 23, no. 1, pp. 42-51.

Keinonen, T., Nieminen, S., Saareks, V., Saano, V., & Ylitalo, P. 2001, "Acceptability and profile of the clinical drug trials underway in Finnish university hospitals in the 1990s: applications reviewed by ethics committees", *Methods & Findings in Experimental & Clinical Pharmacology*, vol. 23, no. 7, pp. 415-423.

Lancet Editorial 2000, "Resisting smoke and spin", *The Lancet*, vol. 355, no. 9211, p. 1197.

Larcombe, I. & Mott, M. 1999, "Multicentre research ethics committees: have they helped?", *Journal of the Royal Society of Medicine*, vol. 92, no. 10, pp. 500-501.

Lardot, C., Steward, W., van Glabbeke, M., & Armand, J. P. 2002, "Scientific review of EORTC trials: the functioning of the New Treatment Committee and Protocol REview Committee", *European Journal of Cancer*, vol. 38, no. Suppl 4, p. S24-S30.

Lemonick, M. D. & Goldstein, A. 2002, "At your own risk. Some patients join clinical trials out of desperation, others to help medicine advance. Who is to blame if they get sick--or even die?", *Time*, vol. 159, no. 16, pp. 46-56.

Levine, R. J. 2001, "Institutional Review Boards: a crisis in confidence", *Annals of Internal Medicine*, vol. 134, no. 2, pp. 161-163.

Lewis, J. C., Tomkins, S., & Sampson, J. R. 2001, "Ethical approval for research involving geographically dispersed subjects: unsuitability of teh UK MREC/LREC system and relevance to uncommon genetic disorders", *Journal of Medical Ethics*, vol. 27, no. 5, pp. 347-351.

Lexchin, J., Bero, L. A., Djulbegovic, B., & Clark, O. 2003, "Pharmaceutical industry sponsorship and research outcome and quality: systematic review", *BMJ*, vol. 326, no. 1167, p. 1176.

Lux, A. L., Edwards, S. W., & Osborne, J. P. 2000, "Responses of local research ethics committees to a study with approval from a multicentre research ethics committee. [see comments]", *BMJ*, vol. 320, no. 7243, pp. 1182-1183.

Lynn, J., Johnson, J., & Levine, R. J. 1994, "The ethical conduct of health services research: a case study of 55 institutions' applications to the SUPPORT project", *Clinical Research*, vol. 42, no. 1, pp. 3-10.

McAuley, L., Pham.B., Tugwell, P., & Moher, D. 2001, "Does the inclusion of grey literature influence estimates of intervention effectiveness reported in meta-analyses?", *Lancet*, vol. 356, pp. 1228-1231.

McNay, L. A., Tavel, J. A., Oseekey, K., McDermott, C. M., Mollerup, D., & Bebchuk, J. D. 2002, "Regulatory approvals in a large multinational clinical trial: the ESPRIT experience", *Controlled Clinical Trials*, vol. 23, no. 1, pp. 59-66.

McWilliams, R., Hoover-Fong, J., Hamosh, A., Beck, S., Beaty, T., & Cutting, G. 2003, "Problematic variation in local institutional review of multicenter genetic epidemiology study", *JAMA*, vol. 290, no. 3, pp. 360-366.

Meinert, C. 1986a, "Interim data analyses for treatment monitoring," in *Clinical trials: design, conduct and analysis. Vol. 8 of Monographs in epidemiology and biostatistics*, Oxford University Press, New York, pp. 208-216.

Meinert, C. L. 1986b, in *Clinical Trials: Design, Conduct and Analysis.*, Oxford University Press, Oxford.

Melander, H., Ahlqvist-Rastad, J., Meijer, G., & Beermann, B. 2003, "Evidence b(i)ased medicine – selective reporting from studies sponsored by pharmaceutical industry", *BMJ*, vol. 326, p. 1171.

Moher, D., Dulberg, C. S., & Wells, G. A. 1994, "Statistical power, sample size, and their reporting in randomized controlled trials", *JAMA*, vol. 272, no. 2, pp. 122-124.

Moher, D., Jadad, A. R., Nichol, G., Penman, M., Tugwell, P., & Walsh, S. 1995, "Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists", *Controlled Clinical Trials*, vol. 16, no. 1, pp. 62-73.

Nafria, E. C., Ferragud, A., & Navarro, R. E. 1998, "Difficulties in the approval of clinical trials in Spain. An obstacle race.", *Revista Clinica Espanola*, vol. 198, no. 5, pp. 272-277.

NHMRC 2001, *National statement on ethical conduct in research involving humans*, Commonwealth of Australia.

NHMRC 2002, *Human Research Ethics Handbook*, **© Commonwealth of Australia 2001**.

NHMRC, A. H. E. C. 1995, *Report on compensation, insurance and indemnity arrangements for institutional ethics committees*, Commonwealth of Australia, Canberra, ACT.

Nicholl, J. 2000, "The ethics of research ethics committees", *BMJ*, vol. 320, no. 29 April, p. 1217.

Ortega, R. & Dal Re, R. 1995b, "Clinical trials committees: how long is the protocol review and approval process in Spain? A prospective study", *Irb: a Review of Human Subjects Research*, vol. 17, no. 4, pp. 6-9.

Ortega, R. & Dal Re, R. 1995a, "Clinical trials committees: how long is the protocol review and approval process in Spain? A prospective study", *Irb: a Review of Human Subjects Research*, vol. 17, no. 4, pp. 6-9.

Pelerin, M. & Hall, S. M. 1992, "Ethics and multicentre research projectss", *BMJ*, vol. 304, p. 1696.

Penn, Z. J. & Steer, P. J. 1995, "Local research ethics committees: hindrance or help", *British Journal of Obstetrics & Gynaecology*, vol. 17, no. 4, pp. 6-9.

Peto, R., Collins, R., & Gray, R. 1993, "Large-scale randomized evidence: large, simple trials and overviews of trials. [Review]", *Annals of the New York Academy of Sciences*, vol. 703, pp. 314-340.

Peto, R., Collins, R., & Gray, R. 1995, "Large-scale randomized evidence: large, simple trials and overviews of trials", *Journal of Clinical Epidemiology*, vol. 48, no. 1, pp. 23-40.

Pich, J., Carne, X., Arnaiz, J. A., Gomez, B., Trilla, A., & Rodes, J. 2003, "Role of a research ethics committee in follow-up and publication of results.[comment]", *Lancet*, vol. 361, no. 9362, pp. 1015-1016.

Pierce, E. 1997, "Are research ethics committees behaving unethically?: Committees are now being expected to do everything [letter]", *BMJ*, vol. 314, no. 7081, pp. 676-673.

Pildal J, Chan AW, Hrobjartsson A, Forfang E, Altman, D. G., & Gotzsche, P. C. 2005, "Comparison of descriptions of allocation concealment in trial protocols and the published reports: cohort study", *BMJ*, vol. 330, no. 7499, p. 1049.

Pocock, S. J. 1997, "Clinical trials with multiple outcomes: a statistical perspective on their design, analysis and interpretation", *Controlled Clinical Trials*, vol. 18, pp. 530-545.

Pocock, S. J. 2002, "Statistical reporting of trials: how to spot those damned lies?", *ADR in Cardiology*.

Pocock, S. J., Hughes, M. D., & Lee, R. J. 1987a, "Statistical problems in the reporting of clinical trials. A survey of three medical journals", *New England Journal of Medicine*, vol. 317, no. 7, pp. 426-432.

Pocock, S. J., Hughes, M. D., & Lee, R. J. 1987b, "Statistical problems in the reporting of clinical trials. A survey of three medical journals", *New England Journal of Medicine*, vol. 317, no. 7, pp. 426-432.

Proschan, M. A. & Waclawiw, M. A. 2000, "Practical guidelines for multiplicity adjustment in clinical trials.", *Controlled Clinical Trials*, vol. 21, no. 6, pp. 527-539.

Quality of Care & Health Outcomes Committee 1995, *Guidelines for the Development and Implementation of Clinical Practice Guidelines.*, Australian Government Publishing Service, Canberra, Australia.

Raju, T. N., Langenberg, P., & Sen, A. 1993, "Suspended judgment. Treatment effect size in clinical trials: an example from surfactant trials", *Controlled Clinical Trials*, vol. 14, no. 6, pp. 467-470.

Redshaw, M. E., Harris, A., & Baum, J. D. 1996, "Research ethics committee audit: differences between committees", *Journal of Medical Ethics*, vol. 22, pp. 78-82.

Roland, M. & Torgerson, D. J. 1998, "Understanding controlled trials: What are pragmatic trials?", *BMJ*, vol. 316, no. 7127, p. 285.

Ross, W. M. 1994, "Ethics approval.", *Journal of the Royal College of Physicians of London*, vol. 28, no. 1, pp. 87-88.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. 2005, *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments* John Wiley and Sons Ltd, Chichester, West Sussex, England.

Sandhu, J. & Okasha, M. 2001, "Evolutionary ethics - a continuing frustration?", *British Journal of General Practice*, vol. 51, no. 467, p. 495.

Savulescu, J., Chalmers, I., & Blunt, J. 1996, "Are research ethics committees behaving unethically? Some suggestions for improving performance and accountability", *BMJ*, vol. 313, no. 7069, pp. 1390-1393.

Schulz, K. F., Chalmers, I., Grimes, D. A., & Altman, D. G. 1994, "Assessing the quality of randomization from reports of controlled trials published in obstetrics and gynecology journals", *JAMA*, vol. 272, no. 2, pp. 125-128.

Schulz, K. F., Chalmers, I., Hayes, R. J., & Altman, D. G. 1995, "Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials", *JAMA*, vol. 273, no. 5, pp. 408-412.

Silagy, C. A. & Jewell, D. 1994, "Review of 39 years of randomized controlled trials in the British Journal of General Practice.", *British Journal of General Practice*, vol. 44, no. 385, pp. 359-363.

Silverman, H., Hull, S. C., & Sugarman, J. 2001, "Variability among institutional review boards' decisions within the context of a multicenter trial", *Critical Care Medicine*, vol. 29, no. 2, pp. 235-241.

Silverman, W. A. 1999, "Realism and fairness in clinical trials", *Paediatric and Perinatal Epidemiology*, vol. 13, pp. 386-391.

Simes, R. J. 1986, "Publication bias: the case for an international registry of clinical trials", *Journal of Clinical Oncology*, vol. 4, no. 10, pp. 1529-1541.

Simes, R. J. 1998, "Controlled clinical trials," in *Handbook of Public Health Methods*, C. Kerr, R. Taylor, & G. Heard, eds., McGraw Hill Australia, Sydney, pp. 120-136.

Smith, G. D., Clemens, J., Crede, W., Harvey, M., & Gracely, E. J. 1987, "Impact of multiple comparisons in randomized clinical trials", *The American Journal of Medicine*, vol. 83, pp. 545-550.

Smith, M. A., Jalaludin, B., Leeder, S. R., & Smith, W. T. 1994, "Isn't one institutional ethics committee's approval enough?", *Medical Journal of Australia*, vol. 160, no. 10, p. 662.

Soares, H. P., Daniels, S., Kumar, A., Clarke, M., Scot, C., Swann, S., & Djulbegovic, B. 2004, "Bad reporting does not mean bad methods for randomised trials: observational study of randomised controlled trials performed by the Radiation Therapy Oncology Group", *BMJ*, vol. 328, pp. 22-25.

Stair, T. O., Reed, C. R., Radeos, M. S., Koski, G., & Camargo, C. A. 2001, "Variation in institutional review board responses to a standard protocol for a multicenter clinical trial", *Academic Emergency Medicine*, vol. 8, no. 6, pp. 636-641.

Stern, J. & Simes, R. J. 1997, "Publication bias: evidence of delayed publication in a cohort study of clinical research projects.", *BMJ*, vol. 315, pp. 640-645.

Sterne, J. & Davey Smith, G. 2001, "Sifting the evidence - what's wrong with significance tests?", *BMJ*, vol. 322, pp. 226-231.

Stone, P. G. & Blogg, C. F. 1997, "Local research ethics committees. National research ethics committee is needed.", *BMJ*, vol. 315, no. 7099, pp. 60-61.

Takayanagi, T., Yamada, Y., Ito, K., Nakamura, H., Sato, H., Kimura, K., Omata, M., & Iga, T. 2001, "The roles of pharmacists on prior review of clinical trials", *Yakugaku Zasshi*, vol. 121, no. 9, pp. 683-690.

Tannock, I. F. 1996, "False-positive results in clinical trials: multiple significance tests and the problem of unreported comparisons", *Journal of the National Cancer Institute*, vol. 88, no. 3-4, pp. 206-207.

Tappin, D. M. & Cockburn, F. 1992, "Ethics and ethics committees: HIV serosurveillance in Scotland", *Journal of Medical Ethics*, vol. 18, no. 1, pp. 43-46.

The European Agency for the Evaluation of Medicinal Products 2002, *Points to consider on multiplicity issues in clinical trials.*

The Ottawa Group 2006, *Ottawa Statement on Trial Registration.*

Thornquist, M. D., Edelstein, C., Goodman, G. E., & Omenn, G. S. 2002, "Streamlining IRB review in multisite trials through single-study IRB Cooperative Agreements: experience of the Beta-Carotene and Retinol Efficacy Trial (CARET)", *Controlled Clinical Trials*, vol. 23, no. 1, pp. 80-86.

Tonkin, A. & Jureidini, J. N. 2005, "Wishful thinking: antidepressant drugs in childhood depression", *British Journal of Psychiatry*, vol. 187, pp. 304-305.

Tully, J., Ninis, N., Booy, R., & Viner, R. 2000, "The new system of review by multicentre research ethics committees: prospective study", *BMJ*, vol. 320, no. 29 April, pp. 1179-1182.

U.S Food and Drug Administration 1998, *Guidance for Industry: Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products*, US Department of Health and Human Services.

U.S Food and Drug Administration 2002, *Code of Federal Regulations - Title 21 - Food and Drugs. Part 56 Institutional Review Boards.*, US Department of Health and Human Services..

UK National Health Service 1997, "Multi-Centre Research Ethics Committees", *R&D News*, vol. 10.

US Dept Health and Human Services 1998, *Institutional Review Boards: a time for reform*, Office of the Inspector General, US Department of Health and Human Services.

Van Luijn, H. E., Musschenga, A. W., Keus, R. B., Robinson, W. M., & Aaronson, N. K. 2002, "Assessment of the risk/benefit ratio of phase II cancer clinical trials by Institutional Review Board (IRB) members", *Annals of Oncology*, vol. 13, no. 8, pp. 1307-1313.

Watling, M. I. & Dewhurst, J. K. 1995, "Local research ethics committees: central ethics committee might have to face hostile locals", *BMJ*, vol. 311, no. 7019, pp. 1571-1572.

Watling, M. I. & Dewhurst, J. K. 1993, "Current experience of central versus local ethics approval in multicentre studies.", *Journal of the Royal College of Physicians of London*, vol. 27, no. 4, pp. 399-402.

Wessely, S. & Kerwin, R. 2004, "Suicide risk and the SSRIs", *JAMA*, vol. 292, no. 3, pp. 379-381.

While, A. E. 1996, "Research ethics committees at work: the experience of one multi-location study", *Journal of Medical Ethics*, vol. 22, pp. 352-355.

Williamson, P. R. & Gamble, C. 2005, "Identification and impact of outcome selection bias in meta-analysis.", *Statistics in Medicine*, vol. 24, no. 10, pp. 1547-1561.

Wise, P. & Drury, M. 1996, "Pharmaceutical trials in general practice: the first 100 protocols. An audit by the clinical research ethics committee of the Royal College of General Practitioners.", *BMJ*, vol. 313, pp. 1245-8.

Wolf, L. E., Croughan, M., & Lo, B. 2002, "The challenges of IRB review and human subjects protections in practice-based research", *Medical Care*, vol. 40, no. 6, pp. 521-529.

World Health Organization 2006, *International Clinical Trials Registry Platform Scientific Advisory Group. Report of Meeting, 17-18 November 2005, Geneva, Switzerland.*

World Medical Association. Declaration of Helsinki: ethical principles for medical research involving humans. 2000.
Ref Type: Bill/Resolution