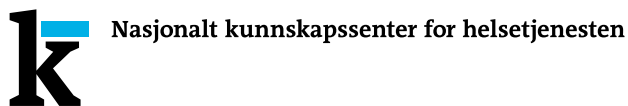


# Overview of systematic methodology reviews

of the design and conduct of randomized trials and systematic reviews of healthcare interventions

Rapport fra Kunnskapssenteret Nr 17–2006

Kunnskapsoppsummering



**Background:** Decisions about the design and reporting of randomized trials and systematic reviews should be informed by the best available methodological research. Ideally this evidence should be summarised in systematic reviews. This overview includes systematic methodology reviews relevant to conducting and reporting pragmatic randomized controlled trials (RCTs) and systematic reviews of RCTs. It was undertaken as part of the Practihc project (Pragmatic Randomized Controlled Trials in HealthCare) and in collaboration with ESCORT. This overview of reviews aims to support Practihc guidance for designing pragmatic randomised trials; to inform future revisions of the CONSORT guidelines for reporting randomised trials; to inform guidelines for conducting and reporting systematic reviews; and to inform decisions about priorities for Cochrane methodology reviews. **Methods:** Methodology reviews were compiled by searching the Cochrane Methodology Register, the Cochrane Database of Methodology Reviews, and UK NHS HTA Methodology Reviews. Two reviewers identified potentially relevant reviews. These were retrieved and the same two reviewers assessed (fortsetter på baksiden)

Nasjonalt kunnskapssenter for helsetjenesten  
Postboks 7004, St. Olavs plass  
N-0130 Oslo  
(+47) 23 25 50 00  
www.kunnskapssenteret.no  
Rapport: ISBN 82-8121-111-3 ISSN 1503-9544

nr 17-2006

Nasjonalt kunnskapssenter for helsetjenesten



(fortsettelsen fra forsiden)

the relevance. **Main results:** A total of twenty-eight methodology reviews were included covering sixteen topic areas. Thirty-one structured abstracts were prepared for the included reviews and a commentary was written for each topic area. **Conclusions:** There are relatively few systematic methodology reviews and many of the included methodology reviews found a paucity of empirical evidence. As a consequence, many decisions about the design and reporting of randomized trials and systematic reviews must be based on logical arguments, often with uncertainty about what empirical evidence is available, due to the lack of a systematic methodology review, or uncertainty about the impact of alternative decisions, due to the lack of empirical evidence. This uncertainty not only impacts on the use of resources for research, but it has important consequences for the availability of reliable evidence to inform decisions about health care.

---

# Table of Contents

INTRODUCTION.....	3
CONFLICTS OF INTEREST .....	4
Research funded by drug companies is more likely to have outcomes that .....	4
favour the sponsor's product.....	4
Financial conflicts of interest are widespread and can influence biomedical .....	5
research in important ways.....	5
A Practihc/ESCORT Commentary .....	5
COST OF HEALTHCARE TECHNOLOGIES .....	7
Assessment of costs of healthcare technologies in clinical trials .....	7
A Practihc/ESCORT Commentary .....	8
EVALUATING AREA-WIDE AND ORGANIZATION BASED INTERVENTIONS .....	10
Methods for evaluating area-wide and organisation-based interventions in health.....	10
and health care .....	10
A Practihc/ESCORT Commentary .....	11
EVALUATING NON-RANDOMIZED INTERVENTIONS .....	12
Differences in the results of randomized and non-randomized studies.....	12
Most quality assessment tools used for appraising non-randomized studies omit key quality domains .....	13
A Practihc/ESCORT Commentary .....	14
EVALUATING PATIENT FOCUSED OUTCOMES.....	15
Eight criteria for selecting patient-based outcome measures for clinical trials .....	15
A Practihc/ESCORT Commentary .....	15
FACTORS LIMITING QUALITY, NUMBER AND PROGRESS OF RCTS.....	17
Remediable factors limit the quality, number and progress of RCTs .....	17
A Practihc/ESCORT Commentary .....	18
IDENTIFYING TRIALS.....	19
Complex electronic search strategy retrieves most RCTs identified by handsearching.....	19
Expert MEDLINE searches yield half of relevant RCTs on average.....	20
A Practihc/ESCORT Commentary .....	21
INCREASING RESPONSE RATES TO POSTAL QUESTIONNAIRES .....	23
A number of strategies can increase or decrease responses to postal questionnaires.....	23
A Practihc/ESCORT Commentary .....	24
INFORMATION FRAMING.....	25
Information framing may influence physicians' decision-making but the effects of information framing are unstable.....	25
A Practihc/ESCORT Commentary .....	25
INFORMED CONSENT .....	27
Informed consent: more information results in more knowledge, but may also increase anxiety and reduce recruitment .....	27
Lack of empirical evidence of the process of informed consent in clinical trials and its impact on trial results .....	28
A Practihc/ESCORT Commentary .....	29
PLACEBO EFFECTS .....	30
Colours affect the perceived action of drugs and may influence effects.....	30
Non specific factors can affect the results in clinical trials .....	30
Placebo effects vary greatly, are frequently large, and may be misattributed to specific treatment effects .....	31
A Practihc/ESCORT Commentary .....	33
PUBLICATION BIAS .....	35
Trials with positive results are published sooner and more often than those with negative results .....	35
Published trials are generally larger than grey trials and may show an overall greater treatment effect .....	36
Studies reported as abstracts are published more frequently in full if they .....	36
have significant results.....	36
Studies with significant results are more widely disseminated than those .....	37
with non-significant results.....	37

---

A Practihc/ESCORT Commentary .....	39
QUALITY OF RANDOMIZED TRIALS.....	40
Major weaknesses in scales and checklists for assessing the methodological quality of RCTs.....	40
A Practihc/ESCORT Commentary .....	41
RANDOMIZATION .....	42
Differences in participants in randomized versus non-randomized studies may.....	42
influence effect estimates.....	42
The effect of allocating by patient preference rather than randomization is not known .....	43
Randomized versus non-randomized allocation often leads to a different estimate of effect .....	44
Non-randomized and randomized trials with inadequate concealment differ from those with adequate concealment.....	45
Estimates of effect from non-randomized studies may be valid if important confounding factors are controlled for .....	46
A Practihc/ESCORT Commentary .....	47
RECRUITMENT FOR TRIALS .....	49
A number of factors influence recruitment into clinical trials .....	49
Uncertainty about the effects of most strategies to improve recruitment into.....	50
health-care studies.....	50
Many barriers to participation in randomized controlled trials by patients and clinicians.....	51
A Practihc/ESCORT Commentary .....	52
STRATIFIED RANDOMIZATION .....	54
When, why and how to stratify in clinical trials.....	54
A Practihc/ESCORT Commentary .....	55

---

---

## INTRODUCTION

Decisions about the design and reporting of randomized trials and systematic reviews should be informed by the best available methodological research. Ideally this evidence should be summarised in systematic reviews. This overview of systematic methodology reviews was undertaken as part of Practihc (Pragmatic Randomized Controlled Trials in HealthCare), a European Commission-funded project aimed at providing training and support to researchers in developing countries who are interested in designing and conducting pragmatic randomized controlled trials of healthcare interventions in collaboration with ESCORT, a project to compile research relevant to the CONSORT guidelines. This overview of reviews aims to support Practihc guidance for designing pragmatic randomised trials; to inform future revisions of the CONSORT guidelines for reporting randomised trials; to inform guidelines for conducting and reporting systematic reviews; and to inform decisions about priorities for Cochrane methodology reviews.

The overview includes systematic methodology reviews relevant to conducting and reporting pragmatic randomized controlled trials (RCTs) and systematic reviews of RCTs.

Methodology reviews were compiled by searching the Cochrane Methodology Register (The Cochrane Library, Issue 2, 2003), the Cochrane Database of Methodology Reviews (Issue 2, 2003), and UK NHS HTA Methodology Reviews (August 2003). Two reviewers identified potentially relevant reviews from the search in the Cochrane Methodology Register and among the UK NHS HTA Methodology Reviews. Reviews that were potentially relevant were retrieved and the same two reviewers assessed the relevance of the articles. Disagreements were resolved by discussion, including a third reviewer, if needed.

For each included methodology review, a structured abstract and a commentary were prepared. Structured abstracts were written by a single author and then checked for accuracy by two reviewers. If a review covered several topics, a separate abstract was prepared for each topic. In some cases, one commentary was written to address several reviews on the same topic.

---

## CONFLICTS OF INTEREST

### **Research funded by drug companies is more likely to have outcomes that favour the sponsor's product**

Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research Outcome and quality: systematic review. *BMJ* 2003; 326:1167-1170.

#### STRUCTURED ABSTRACT

*From The Cochrane Collaboration Methods Groups Newsletter 2004 and peer reviewed by Andy Oxman and Merrick Zwarenstein.*

#### **Background**

An increasing number of clinical trials are sponsored by the pharmaceutical industry. Results that are unfavourable, that is, trials that find a drug less clinically effective or less safe than other drugs used to treat the same condition, pose considerable financial risks to the pharmaceutical industry. Pressure to show that a drug causes a favourable outcome may therefore result in biases in the design, outcome and reporting of industry sponsored research.

#### **Question**

Is the funding of drug studies by the pharmaceutical industry associated with outcomes that are favourable to the funder and are the methods of trials funded by pharmaceutical companies different from the methods in trials with other sources of support?

#### **Search Strategy**

Studies were identified by searching MEDLINE (January 1966 to December 2002), EMBASE (January 1980 to December 2002) and the Cochrane Methodology Register. Searches were supplemented by contacting two e-mail discussion lists and content experts, scanning references in articles and the authors' personal files.

#### **Selection Criteria**

Studies that specifically stated that they analysed research sponsored by a pharmaceutical company, compared methodological quality or outcomes with studies with other sources of funding, and reported the results in quantitative terms.

#### **Data collection and analysis**

One of the authors did the initial selection of studies. Three of the authors confirmed each study for inclusion and independently extracted data. Disagreements were resolved by consensus. Studies that reported the effects of funding on the results of clinical trials were pooled if an odds ratio could be computed.

#### **Main results**

3351 references were screened, 103 were retrieved for more detailed evaluation, and 30 studies were included. Two studies of cohorts of clinical trials found that research funded by drug companies was less likely to be published than research funded by other sources. Sixteen studies investigated the relationship between funding source and the results of clinical trials and meta-analyses. Studies sponsored by pharmaceutical companies were more likely to have outcomes favouring the sponsor than studies funded by other sources (odds ratio 4.05; 95% confidence interval 2.98 to 5.51; 18 comparisons from 15 studies). None of the 13 studies that analysed methodological quality reported that studies funded by industry were of poorer methodological quality.

#### **Authors' conclusions**

The results of research are biased in favour of the products that are made by the company funding the research. Explanations include the selection of an inappropriate comparator to the product being investigated and publication bias.

---

## **Financial conflicts of interest are widespread and can influence biomedical research in important ways**

Bekelman JE, Li Y, Gross CP. Scope and impact of financial conflicts of interest in biomedical research: a systematic review. *JAMA* 2003; 289:454-465.

### **STRUCTURED ABSTRACT**

*From The Cochrane Collaboration Methods Groups Newsletter 2003 and peer reviewed by Andy Oxman and Max Petzold.*

#### **Background**

Conflicts of interest are a set of conditions in which professional judgment concerning a primary interest (such as a patient's welfare or the validity of research) tends to be unduly influenced by a secondary interest (such as financial gain). There is increasing awareness and concern about the potential impact of financial conflicts of interest on biomedical research.

#### **Question**

What is the extent and impact of financial conflicts of interest in biomedical research and how are they managed?

#### **Search strategy**

Studies were identified by searching MEDLINE (January 1980 to October 2002), science citation index, references of articles, letters, commentaries, editorials, books and by contacting experts.

#### **Selection criteria**

All English language studies containing original, quantitative data on financial relationships among industry, scientific investigators, and academic institutions. The impact of such relationships, or how these financial relationships are managed, were included.

#### **Data collection and analysis**

Two investigators reviewed citations and selected appropriate studies. One investigator extracted data from each of the included studies. Criteria used to appraise methodological quality varied according to study design. The main outcomes were the prevalence of specific types of industry relationships, the relation between industry sponsorship and research results or investigator behaviour, and the process for disclosure, review, and management of financial conflicts of interest. The investigators contacted authors for missing data. The results of studies of the association between industry sponsorship and original research results were pooled.

#### **Main results**

1664 citations were screened, 144 potentially eligible full articles were retrieved, and 37 studies met the inclusion criteria. Approximately a quarter of investigators have industry affiliations, and roughly two thirds of academic institutions hold equity in start-ups. Eight articles, which together evaluated 1140 original studies, assessed the relationship between industry sponsorship and original research results. Aggregating the results of these articles showed a statistically significant association between industry sponsorship and pro-industry conclusions (pooled Mantel-Haenszel odds ratio, 3.6; 95% confidence interval 2.6 to 4.9). Industry sponsorship was also associated with restrictions on publication and data sharing. The approach to managing financial conflicts varied substantially across academic institutions and peer reviewed journals.

#### **Authors' conclusions**

Financial relationships among industry, scientific investigators and academic institutions are widespread. Conflicts of interest arising from these ties can influence biomedical research in important ways.

### **A Practihc/ESCORT Commentary**

Andy Oxman

A commentary addressing the methodology reviews:

- Lexchin J, Bero LA, Djulbegovic B, Clark O. Pharmaceutical industry sponsorship and research outcome and quality: systematic review. *BMJ* 2003; 326:1167-1170.

- 
- Bekelman JE, Li Y, Gross CP. Scope and impact of financial conflicts of interest in biomedical research: a systematic review. *JAMA* 2003; 289:454-465.

### **Are the results valid?**

Lexchin found 30 studies published between 1986 and 2002 that compared the methodological quality or outcomes of research sponsored by the pharmaceutical industry with research that had other sources of support. Bekelman only included studies reported in English and found 37 studies published between 1985 and 2002 that contained original, quantitative data on financial relationships among industry, scientific investigators, and academic institutions and assessed the impact of such relationships or how these financial relationships were managed. They did not include 16 of the studies found by Lexchin. Both reviews are limited by the difficulty in locating relevant research. For example, Lexchin found only 13 of the 30 studies they included using MEDLINE and EMBASE. The remaining studies were found through personal contacts and scanning reference lists. Neither used explicit criteria to assess the methodological quality of the included studies.

### **What are the implications?**

Studies sponsored by pharmaceutical companies are four times more likely to have outcomes favouring the sponsor than studies funded by other sources. This may be due in part to bias. Lexchin did not find evidence that trials sponsored by pharmaceutical companies have lower methodological quality, but the criteria used to assess methodological quality in the studies they found did not include criteria such as selection of an inappropriate comparator.<sup>1,2</sup> Another explanation for industry-funded trials being more likely to have favourable results is publication bias. Both Bekelman and other evidence show that the interest of research sponsors can restrict the dissemination of research findings.<sup>3</sup> These findings, together with evidence of publication bias generally, provide strong support for repeated calls for obligatory prospective registration of trials.<sup>4</sup> They also provide strong support for the full disclosure of the nature and extent of relationships between investigators and sponsors in reports of trials. Trialists should ensure that there are no restrictions on reporting the results of trials before accepting funding from industry and they should make available all research results from completed trials in a comprehensive, publicly accessible registry. Governments, academic institutions and journals should help to ensure that this happens.

Systematic reviews of the effects of drugs should not be restricted to published trials and should consider inappropriate comparators and other sources of bias beyond those addressed by commonly used criteria for assessing the risk of bias.

### **What are the implications for future methodology research?**

These reviews warrant updating and could be improved by expanded search strategies and methodological assessments of the included studies. We need empirical methodological studies and methodology reviews of sources of misleading results such as inappropriate comparators, composite endpoints and inappropriate subgroup analyses, which are not captured by criteria used to assess the risk of bias in trials.

### **References**

1. Sackett DL, Oxman AD. HARLOT plc. an amalgamation of the world's two oldest professions. *BMJ* 2003; 327:1442-1445.
2. Montori VM, Jaeschke R, Schünemann HJ, Bhandari M, Brozek JL, Devereaux PJ, Guyatt GH. Users' guide to detecting misleading claims in clinical research reports. *BMJ* 2004; 329:1093-1096.
3. Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases. *Health Technology Assessment* 2000; 4(10).
4. Krleza-Jeric K, Chan AW, Dickersin K, Sim I, Grimshaw J, Gluud C. Principles for international registration of protocol information and results from human trials of health related interventions: Ottawa statement (part 1). *BMJ* 2005;330:956-958.



---

# COST OF HEALTHCARE TECHNOLOGIES

## Assessment of costs of healthcare technologies in clinical trials

Johnston K, Buxton MJ, Jones DR, Fitzpatrick R. Assessing the costs of healthcare technologies in clinical trials. *Health Technology Assessment* 1999;3(6).

### STRUCTURED ABSTRACT

*Prepared by Signe Flottorp and peer reviewed by Max Petzold and Joel Gagnier.*

#### **Background**

When economic evaluations are conducted alongside clinical trials, comprehensive information on resource-use quantities can be collected. Estimation of cost data at the individual level has the advantage that it allows statistical analysis of costs to be performed, but it might overburden the trial data collection process.

#### **Question**

What are the methodological issues concerning the collection of resource-use data for costing purposes and its analyses?

#### **Search strategy**

The process by which relevant articles were identified consisted of ten key stages:

- definition of inclusion criteria and design of initial search strategy.
- searches of an in-house bibliographic database.
- manual searches of key journals.
- refinement of search strategy.
- electronic searches for key articles.
- review of key papers and identification of key methodological issues.
- electronic searches for articles on specific methodological issues.
- citation searches using articles on specific methodological issues.
- reference lists of identified articles.
- articles identified by experts.

Electronic searches were conducted in MEDLINE, EMBASE, Healthstar and the Health Economic Evaluations Database (Office of Health Economics) and were limited to English language articles for the period 1986-1996.

#### **Selection criteria**

Articles were included that:

- reviewed methods of costing alongside clinical trials or reviews of a single methodological issue.
- conducted an empirical analysis of an economic evaluation alongside a trial that raises new methodological issues.
- presented guidelines on how to perform economic evaluations.
- conducted empirical analyses of specific methodological issues.
- presented guidelines for authors for publishing evaluations.

#### **Data collection and analysis**

Methodological issues in the retrieved articles were identified and structured into four categories: study design, data collection, data analysis and presentation of results. Details of retrieved articles were entered on Pro-cite with keywords assigned for the source and the type of the article. In developing the review, structured comments from relevant experts were sought with the aim of identifying further issues and opinions. It is not clearly stated how the information in the articles was used to come to the conclusions or to make the recommendations.

#### **Main results**

Current methodological issues where there is general agreement is distinguished from those of disagreement. Issues of disagreement can be further divided into two types:

- those that reflect legitimate differences in values and perspectives.
- those that are amenable to further elucidation by empirical research.

---

The recommendations that arise from the review are targeted at three main groups: 1) investigators; 2) funding bodies; and 3) those responsible for ensuring high standards in reporting of studies.

### **Authors' conclusions**

Methodological issues on which there is general agreement include:

- identifying perspective of study.
- measuring units of resource use, and applying appropriate unit cost.
- measurement of health service cost of the whole intervention.
- analysis of uncertainty.
- transparency in methods and results.

Issues remaining open because of legitimate differences in values or perspectives include:

- which perspective to adopt.
- whether to base decisions on economic welfare theory.
- which approach to analysis to adopt: estimation, hypothesis testing or decision analysis.

Methodological issues requiring further empirical study include:

- exploring optimal sampling approaches.
- questions surrounding multi-centre clinical trials.
- testing the validity and reliability of resource-use data collection methods.
- handling missing and censored data.
- methods used to generalise the results.
- development of a common reporting format for economic evaluations.

## **A Practihc/ESCORT Commentary**

Merrick Zwarenstein

A commentary addressing the methodology review:

Johnston K, Buxton MJ, Jones DR, Fitzpatrick R. Assessing the costs of healthcare technologies in clinical trials. *Health Technology Assessment* 1999;3(6).

### **Are the results valid?**

This is a well-conducted review, which gathered issues raised by conducting a costing study alongside a clinical trial. The most recent article in this review is from 1997, and it now needs to be updated.

### **What are the implications?**

A randomized trial is an ideal opportunity to also collect information on the resources used in providing an intervention, the better for decision-makers to choose among alternative options.

Key questions for investigators doing cost studies are: which costs should be collected; in how much detail; from what perspective; from how many subjects?

The rationale for these choices should be explicit. Existing information on costs should be found and used. Results should be reported in such a way that the study can be repeated. Data should be deposited in accessible archives if such exist, and should be shared.

The reviewers suggested that trial funding bodies should:

- be prescriptive about the perspective to be adopted in costing studies (or require that investigators justify their chosen perspective).
- ensure that decisions about resource-use data collection take place.
- recognise that this may require the analysis of existing data, or a pilot study.
- establish archives of data collection instruments and data sets.
- require researchers to deposit their work in these archives.
- encourage testing of methodological alternatives within trials.
- actively commission research into alternative methods of cost estimation.

---

The review supports the CONSORT guidelines.

**Implications for future methodology research**

The decision aid developed for this review and other identified methodological issues should be explicitly tested in future trials which incorporate costing studies.

---

# EVALUATING AREA-WIDE AND ORGANIZATION BASED INTERVENTIONS

## Methods for evaluating area-wide and organisation-based interventions in health and health care

Ukoumunne OC, Gulliford MC, Chinn S, Sterne JAC, Burney PGJ. Methods for evaluating area-wide and organisation-based interventions in health and health care: a systematic review. *Health Technology Assessment* 1999;3(5).

### STRUCTURED ABSTRACT

*Prepared by Signe Flottorp and peer reviewed by Max Petzold and Dave Sackett.*

#### **Background**

Health care interventions are often implemented at the level of geographical area or health service organisational unit, for clusters of individuals. Evaluation of cluster-based interventions presents a number of difficulties that are not always addressed in an optimal manner.

#### **Question**

What methods should be used for evaluating cluster-based interventions?

#### **Search strategy**

Studies were identified by searching MEDLINE (1966-1996), EMBASE, ERIC, Science Citation Index and the Social Science Citation Index databases, by handsearching *Statistics in Medicine* for the period January 1992 to July 1997, and some other relevant journals for the period July 1996 to November 1998. Other sources of knowledge were considered including books, newsletters, conference papers and personal communication from key informants.

#### **Selection criteria**

The search was restricted to English language. The search process identified a very large number of potential papers. The titles and abstracts were inspected to evaluate their relevance to the focus of the review.

#### **Data collection and analysis**

The papers were assessed against conventional epidemiological and statistical principles by two reviewers. The number of papers included in the review is not stated. A narrative review was drafted in which the methods proposed in the most relevant and valid papers were recommended for adoption. The general approach was classified as one of "best evidence synthesis".

#### **Main results**

The main methodological findings of the review were synthesised into a 12-point checklist for investigators:

- recognise the cluster as the unit of intervention or allocation.
- justify the use of the cluster as the unit of intervention or allocation.
- include a sufficient number of clusters.
- randomize clusters wherever possible.
- in non-randomized studies include a control.
- in single group studies include repeated measurements over time.
- allow for clustering when estimating the required sample size.
- consider the use of pairing or stratification of clusters where appropriate.
- consider the different approaches to repeated measurements in prospective evaluations.
- allow for clustering at the time of analysis.
- allow for confounding at both individual and cluster level.
- include estimates of intraclass correlations and components of variance in published reports.

#### **Authors' conclusions**

Over the last few years a considerable amount of research has been carried out to address the methodological problems that are encountered in area- and organisation-based evaluations. These methods are now sufficiently

---

---

accessible to allow implementation in the context of much healthcare evaluation. Further research work is needed: to aid the design of quasi-experimental cluster-based studies; to provide intraclass correlations and components of variance for a range of outcomes and different types of clustering; to provide analytical methods for different types of data; and to permit meta-analyses of the results of cluster-based studies.

## **A Practihc/ESCORT Commentary**

Merrick Zwarenstein

A commentary addressing the methodology review:

Ukoumunne OC, Gulliford MC, Chinn S, Sterne JAC, Burney PGJ. Methods for evaluating area-wide and organisation-based interventions in health and health care: a systematic review. *Health Technology Assessment* 1999;3(5).

This review is a scan of methodological issues in the conduct of area and organization based intervention studies. The authors reviewed current books and articles on the conduct of cluster trials, and synthesized these into a checklist for researchers to use in their study design phase.

### **Are the results valid?**

This is a well conducted review, which also included time series studies, controlled before after designs, and some case studies. Weaknesses in the non-randomized designs are discussed.

The review is up to date for its publication date of 1999 but must be updated now as several books and many relevant articles have since been published, for example, the CONSORT article on cluster randomized trials, which incorporates many of the points made in this review.

### **What are the implications?**

Trialists should be alert to the clustered structure of their study at the time of allocation, of intervention and of analysis, and ensure that the statistical techniques they use are appropriate.

Evaluators who are contemplating a non-randomized design should consider this option very carefully, and be cautious about the causal conclusions they draw from these weaker designs.

Intraclass correlation coefficients are a rare and vital item of information for clustered studies; these should be reported fully. If possible a sustainable way to collect these should be established, to update and extend the important collection in this review.

---

# EVALUATING NON-RANDOMIZED INTERVENTIONS

## Differences in the results of randomized and non-randomized studies

Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, Petticrew M, Altman DG. Review of empirical comparisons of the results of randomized and non-randomized studies. Chapter 3 in: Evaluating non-randomized intervention studies. *Health Technology Assessment* 2003; 7:(27).

### STRUCTURED ABSTRACT

*Prepared by Signe Flottorp and peer reviewed by Max Petzold and Dave Sackett.*

#### **Background**

In the absence of randomized controlled trials, healthcare practitioners and policy-makers rely on non-randomized studies to provide evidence of the effectiveness of healthcare interventions. However, there is controversy over the validity of non-randomized evidence, related to the existence and magnitude of selection bias. This HTA report aims to consider methods and related evidence for evaluating bias in non-randomized interventions studies. It contains three systematic reviews and new empirical investigations. Only one of the systematic reviews is presented in this abstract.

#### **Question**

What is the empirical evidence of the importance of randomisation per se, provided by meta-epidemiological comparisons of randomized and non-randomized studies?

#### **Search strategy**

Reviews were identified from a search of electronic databases including MEDLINE, EMBASE and PsycLit up to December 1999; from handsearches in relevant journals and from contact with experts in the field. Additionally, the searches carried out for other sections of the project were screened to identify suitable papers.

#### **Selection criteria**

Reviews were eligible for inclusion if:

- they compared quantitative results between RCTs of an intervention and non-randomized studies of the same intervention.
- they had accumulated, through some systematic search, results from several of these comparisons across healthcare interventions.

#### **Data collection and analysis**

The content, results and conclusions from each of the identified reviews were noted. In addition, the methodology of each review was critically assessed for potential weaknesses. Aspects considered were as follows:

- was the identification of included studies unlikely to be biased?
- did the RCTs and non-randomized studies recruit similar participants, use similar interventions and measure similar outcomes?
- were the RCTs and non-randomized studies shown to use similar study methodology in all respects other than the allocation mechanism?
- were sensible, objective criteria used to determine differences or equivalence of study findings?

#### **Main results**

Eight studies were identified that compared the results of randomized and non-randomized studies across multiple interventions using meta-epidemiological techniques. Each comparison reported multiple comparisons of results of randomized and non-randomized studies. Although there was overlap in the comparisons included in these reviews, they reached different conclusions concerning the likely validity of the non-randomized data. Five of the eight reviews concluded that there are differences between the results of randomized and non-randomized studies in many but not all clinical areas, but without there being a consistent pattern indicating systematic bias. One of the eight reviews found an overestimation of effects in all areas studied. The final two concluded that the results of randomized and non-randomized studies were "remarkably similar". Two reviews considered the relative variability of randomized and non-randomized results; one concluded that RCTs were

---

more consistent and the other that they were less consistent. The two studies that investigated the impact of case-mix adjustment both noted that adjustment did not necessarily reduce discordance between randomized and non-randomized findings.

Three commonly stated studies were excluded from the review because they did not consider whether there are differences in results of RCTs and non-randomized studies of the same interventions.

### **Authors' conclusions**

The only robust conclusion that can be drawn is that in some circumstances the results of randomized and non-randomized studies differ, but it cannot be proved that differences are not due to other confounding factors. Since the conclusions of the eight reviews are divergent, and as all the reviews have weaknesses, it is difficult to draw conclusions concerning the importance of randomisation from these investigations. The frequency, size and direction of the biases cannot be judged reliably from the information presented. These investigations also raise concerns regarding the usefulness of meta-epidemiological investigations where there is (or could be) variability in the direction of bias.

## **Most quality assessment tools used for appraising non-randomized studies omit key quality domains**

Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovich C, Song F, Petticrew M, Altman DG. Evaluation of checklists and scales for assessing quality of non-randomized studies. Chapter 4 in: Evaluating non-randomized intervention studies. *Health Technology Assessment* 2003; 7(27).

### **STRUCTURED ABSTRACT**

*Prepared by Signe Flottorp and peer reviewed by Max Petzold and Dave Sackett.*

### **Background**

Regardless of the study designs available, the validity of any estimate of effectiveness is conditional on the quality of the studies upon which that estimate is based. The formal assessment of methodological quality should be routine practice in systematic reviews. This is even more important for non-randomized studies since the largely observational nature of non-randomized studies leads to a much higher susceptibility to bias.

### **Question**

Do the quality assurance tools available for non-randomized intervention studies meet criteria developed by a group of methodologists?

### **Search strategy**

An extensive and comprehensive literature search up to December 1999 was carried out. This included searching a wide range of electronic databases, supplemented with searches of registers of methodological research, citation searches for key papers, handsearching of key journals, scanning of reference lists of all retrieved papers and contact with experts. Owing to the nature of the searches, and the poor indexing of the studies, it was deemed necessary to strike a balance between strategies that were less likely to miss any relevant papers, yet retrieved a "manageable" number of citations.

### **Selection criteria**

To be considered a quality assessment tool, a list of criteria that could be (or had been) used to assess the methodological quality of primary studies was required. These tools could exist either as individual publications or within the context of systematic reviews, such as methodological reviews that had used some form of tool. The tool must have been (or must have the potential to be) applied to non-randomized studies of intended effect. The tools included were not necessarily designed to measure methodological quality of non-randomized studies.

### **Data collection and analysis**

Items related to the following areas were recorded for each tool: descriptive information; tool development; tool content. A taxonomy of 12 quality domains covering the major aspects of study quality was constructed a priori using a modified Delphi process amongst review team members. In some cases additional items were added to accommodate all authors' items.

The data from each study were tabulated and synthesised in a qualitative manner. A primary selection criterion was adopted to reduce the number of tools that were discussed. A "good" quality assessment tool was deemed to

---

be one that included pre-specified items from at least five of six internal validity domains. The tools that covered at least three of four core internal validity criteria were considered to be the “best” tools. These criteria related to assessment of allocation method, attempt to achieve comparability by design, identification of important prognostic factors and adjustment of difference in case mix. Members of the project team assessed the practical use of the tools by using each tool at least twice, on one of three non-randomized studies.

### **Main results**

194 tools that could be used to assess the quality of non-randomized studies were identified. Overall the tools were poorly developed: the majority did not provide a means of assessing the internal validity of non-randomized studies and almost no attention was paid to the principles of scale development and evaluation. However, 14 tools were identified that included items related to each of the pre-specified core internal validity criteria. Six of the 14 tools were considered potentially suitable for use as quality assessment tools in systematic reviews.

### **Authors' conclusions**

Most quality assessment tools used for appraising non-randomized studies omit key quality domains. Future research should include: further appraisal of selected quality criteria; consideration of creation of a new tool or revision of an existing one; more detailed examination of tool “usability”.

## **A Practihc/ESCORT Commentary**

Max Petzold

A commentary addressing chapter 3 and 4 of the methodology review:

Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, Petticrew M, Altman DG. Review of empirical comparisons of the results of randomized and non-randomized studies. Evaluating non-randomized intervention studies. *Health Technology Assessment* 2003; 7:(27).

Results of non-randomized studies sometimes, but not always, differ from results of randomized studies of the same intervention. Two methodology reviews evaluating non-randomized intervention studies are presented. Both reviews are based on literature search up to December 1999. The controversy over the validity of non-randomized evidence is treated with a review of empirical comparisons of the results of randomized and non-randomized studies and with an evaluation of checklists and scales for assessing the quality of such studies.

### **Are the results valid?**

Inclusion criteria, focusing on the quality assessment in the included reviews, are clearly stated in the studies. Both reviews conducted searches in relevant electronic databases and handsearches in journals combined with further contacts with experts in the field. Both studies were limited to "a balance between strategies that were less likely to miss any relevant papers, yet retrieved a manageable number of citations", but the searches seem to be both extensive and comprehensive. A general update for 2000-2005 of the reviews would be valuable.

### **What are the implications?**

There are divergent results from the eight included reviews comparing results of randomized and non-randomized studies, and it is difficult to draw conclusions concerning the importance of randomisation from these investigations. Meta-epidemiological investigations may be of limited usefulness where there is (or could be) variability in the direction of bias. Six tools potentially suitable for quality assessment in systematic reviews of non-randomized studies were found.

What are the implications for future methodology research?

Further research should cover a more detailed examination of tool "usability", for example tools may be more or less useful according to field or type of intervention.



---

# EVALUATING PATIENT FOCUSED OUTCOMES

## Eight criteria for selecting patient-based outcome measures for clinical trials

Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment* 1998; 2(14).

### STRUCTURED ABSTRACT

Prepared by Andy Oxman and peer reviewed by Eduardo Bergel and Elizabeth Paulsen.

#### Background

'Patient-based outcome measures' are questionnaires or interviews used to assess constructs such as health-related quality of life, subjective health status and functional status from the patient's perspective. There is a diverse array of such measures that can be used in clinical trials.

#### Question

What criteria should be used to evaluate and select patient-based outcome measures for use in a clinical trial?

#### Search strategy

MEDLINE, EMBASE, CINAHL, PsychLIT, Sociofile and the Health Economics Research Group's database were searched up to 1996. The reference lists of identified articles were checked. Nine journals were hand searched from between 1990 and 1994 to 1996.

#### Selection criteria

Any article that focused on any methodological aspect of patient-based outcome measures, including methods of evaluating them, psychometric evaluation, evaluation of feasibility, principles of selection, use in trials, utility methodology, comparative studies of different measures, and validation studies of prominent measures.

#### Data collection and analysis

A first draft based on a qualitative synthesis of the included articles was produced by the first author and critiqued by the other three authors. A revised version was submitted to 10 experts. The authors discussed the feedback and prepared the final version.

#### Main results

5621 abstracts were identified as potentially relevant, 391 key references were selected as useful to the review and a further 22 references were incorporated into the final version as a result of comments from external experts and referees.

Seven major types of instruments were identified: disease-specific, site-specific, dimension-specific, generic, summary item, individualised and utility. Concepts, definitions and theories were generally not clearly or consistently used. There are advantages and disadvantages to each of the different types of instrument.

Eight criteria can be used to evaluate and select patient-based outcome measures for clinical trials: appropriateness, reliability, validity, responsiveness, precision, interpretability, acceptability and feasibility.

#### Authors' conclusions

Investigators should choose patient-based outcome measures for trials based on these criteria. Developers of instruments should make evidence available under the same headings.

## A Practihc/ESCORT Commentary

Eduardo Bergel

A commentary addressing the methodology review:

---

Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. *Health Technology Assessment* 1998; 2(14).

### **Are the results valid?**

This review is based on a systematic and comprehensive review of the literature. However, there was no quality assessment of the articles, and a significant proportion of the author's conclusions are based on expert opinions or the authors' views. Not all articles were reviewed using the same procedure. Due to "time constraints" the authors reviewed 199 articles, but only "skim read" another 43 articles, based on date of publication. This is a qualitative review. The methods used to synthesise the findings of the included studies were not clearly described. This article did not attempt to produce a report that directly links evidence and conclusions, but to produce a qualitative summary of available evidence and experts' opinions. Therefore it is unclear whether the conclusions are supported by the evidence.

### **What are the implications?**

The lack of consistency in the definition of patient based outcomes reported in the literature makes it difficult to summarize the evidence from trials in clear, explicit and unambiguous terms. An improvement in reporting might be achieved if investigators make their choice of patient-based outcome measures for trials in terms of the eight criteria identified in this review (appropriateness, reliability, validity, responsiveness, precision, interpretability, acceptability and feasibility) and clearly state this in their published reports.

Other implications include the use of more appropriate and valid instruments in clinical trials and the highlighting of instruments that require further assessment of their reliability and validity.

### **What are the implications for future methodology research?**

There are substantial gaps in knowledge of how to capture patients' perception of illness and outcomes of interventions within clinical trials. Research should focus in the comparison of different types of measures completed by the same patient within a trial especially with regard to responsiveness. This could be done as a methodological component to a clinical trial or as stand-alone studies. Consensus-type processes should also be used to evaluate the instruments. Developers of instruments need to make evidence available that addresses the criteria identified in this review.

### **References**

1. Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. *BMJ*. 2002;324(7351):1417.
2. Patel KK, Veenstra DL, Patrick DL. A review of selected patient-generated outcome measures and their application in clinical trials. *Value Health*. 2003;6(5):595-603.
3. Wiklund I. Assessment of patient-reported outcomes in clinical trials: the example of health-related quality of life. *Fundamental Clinical Pharmacology*. 2004;18(3):351-363.

---

# FACTORS LIMITING QUALITY, NUMBER AND PROGRESS OF RCTS

## Remediable factors limit the quality, number and progress of RCTs

Prescott RJ, Counsell CE, Gillespie WJ, Grant AM, Russell IT, Kiauka S et al. Factors that limit the quality, number and progress of randomized controlled trials. *Health Technology Assessment* 1999;3(20).

### STRUCTURED ABSTRACT

*Prepared by Signe Flottorp and peer reviewed by Merrick Zwarenstein and Eduardo Bergel.*

#### Background

The randomized controlled trial (RCT) is the most powerful research tool for evaluating health technologies. However, for most therapeutic activities reliable information from RCTs is not available.

#### Question

What are the factors that limit the quality, number and progress of RCTs?

#### Search strategy

Specific search strategies were developed for each of the three electronic databases searched: MEDLINE, EMBASE and CINAHL, covering the period 1986-96. The tables of contents of all volumes of Controlled Clinical Trials for the years 1989-96 were scanned. The bibliographies of included articles were reviewed.

#### Selection criteria

The review systematically searched for all factors limiting the quality, number and progress of RCTs. Any RCT comparing approaches to trial design, conduct, analysis or reporting were included. Other articles such as surveys or case studies of RCTs were read and categorised for relevance.

#### Data collection and analysis

Each full text article was assessed for its relevance in the eight different categories of the review. The eight categories were: design, analysis, cost, other obstacles, clinician participation, patient participation, reporting and conduct. Multiple entries for an article were possible. The relevance of the articles was scored on a scale from 0 (irrelevant) to 3 (highly relevant). Priority was given to the articles scoring highest on the relevant keywords within each category. The levels of evidence are stated in relation to the recommendations made.

#### Main results

A total of 638 articles were given a relevance score of 2 or 3 in at least one category. The report presents several recommendations regarding the following issues:

Wide patient inclusion criteria improve representativeness and recruitment, narrow criteria may be appropriate for expensive or hazardous technologies. Outcomes should be clinically and socially relevant. Surrogate and intermediate outcomes may mislead. Randomisation should be secured centrally, and analysis should be as allocated. Bias can be prevented by blinding treatments (when possible), by clear treatment protocols, by high follow up rates, and blind outcome assessment. Power estimates should be reported, with sensitivity analyses and small trials regarded as hypothesis forming.

Participation is improved in trials that investigate important questions with minimal extra work for clinicians and little impact on provision of care. Half of all trials have problems with recruitment, but no strategies for reducing this have been tested. Too much quality control may be harmful to recruitment and costly. There is little or no evidence on the best structure for running a trial, nor on steering or data monitoring committees. Commercial trials should be independently monitored. Primary and secondary outcomes, and subgroups for analysis should be pre-specified. Analysis should be by intention to treat. The CONSORT statement should improve reporting of RCTs.

#### Authors' conclusions

The evidence to guide many aspects of the design, conduct and analysis of RCTs is not always being applied.

---

## A Practihc/ESCORT Commentary

Eduardo Bergel

A commentary addressing the methodology review:

Prescott RJ, Counsell CE, Gillespie WJ, Grant AM, Russell IT, Kiauka S et al. Factors that limit the quality, number and progress of randomized controlled trials. *Health Technology Assessment* 1999;3(20).

### **Are the results valid?**

The authors attempt to find and summarize a very large body of evidence. Single page articles and comments were excluded without review. Non-English articles were excluded. Study quality was assessed, but without a pre-specified criteria. This is a narrative review; no quantitative synthesis of the evidence was attempted. The review is divided into chapters, by topic, and there is substantial heterogeneity between chapters in the strategy that was used to summarize the evidence. There is also heterogeneity in the link between supporting data and conclusions. In significant proportions of the paper, conclusions are not based on data or analysis but on the reviewer's views and opinions. Significant proportions of this review are out of date because data collection was from 1986-1996.

### **What are the implications?**

Recommendations cover many topic areas, including the importance of the research question, the interaction between trialist and clinicians, methodological training of trialist, trial planning, run-in periods, sample size estimations, uses and potential pitfalls of interim analysis and subgroup analysis, recruitment strategies, ethical issues, randomisation procedures, outcome assessment, follow-up strategies, data quality, and data analysis and reporting. The review also covers less common topics like trial administration (i.e. in multicenter trials), refereeing, economic evaluations, and recommendations for journal editors and funding agencies.

### **What are the implications for future methodology research?**

Main topic areas where research is needed are recruitment strategies, patients' understanding of trial procedures, impact of quality-controlled procedures, bias in commercially sponsored trials, impact of run-in periods and subgroup analysis. It is also suggested that the effect of the CONSORT guidelines on the quality of reporting RCTs should continue to be assessed.

---

## IDENTIFYING TRIALS

### Complex electronic search strategy retrieves most RCTs identified by handsearching

Hopewell S, Clarke M, Lefebvre C, Scherer R. Handsearching versus electronic searching to identify reports of randomized Trials. *The Cochrane Database of Methodology Reviews* 2002, Issue 4. Art. No.: MR000001. DOI: 10.1002/14651858.MR000001.

#### STRUCTURED ABSTRACT

*Prepared by Merrick Zwarenstein and peer reviewed by Max Petzold and Eduardo Bergel.*

#### Background

Searching is the key stage in ensuring that a systematic review provides an up to date, comprehensive and least biased summary of the effects of an intervention. While neither electronic nor hand searching has perfect yield, the relative merits are unknown.

#### Question

What are the effects of hand versus three kinds of electronic search (simple, complex and the Cochrane Highly Sensitive Search Strategy- HSSS) on the number of RCT reports retrieved?

#### Search strategy

Studies were sought from the Cochrane Methodology Register (The Cochrane Library, Issue 2, 2002), MEDLINE (1966 to Week 1 July 2002), EMBASE (1980 to Week 25 2002), Allied and Complementary Medicine Database (AMED) (1985 to June 2002), Biological Abstracts (BIOSIS) (1985 to June 2002), Cumulative Index to Nursing and Allied Health Literature (CINAHL) (1982 to June 2002), Library and Information Science Abstracts (LISA) (1969 to July 2002) and Psychological Abstracts (PsycINFO) (1972 to May 2002). Studies were also sought during the handsearching of selected journals, which is being carried out by the UK Cochrane Centre for all studies relevant to the methodology of systematic reviews. The abstracts presented at all Cochrane Colloquia (1993 to 2001), Systematic Reviews Symposia (1998 to 2002) and Society for Clinical Trials Meetings (1980 to 2001) (as published in *Controlled Clinical Trials*) have also been handsearched as part of this activity. Researchers who may have carried out relevant studies were also contacted.

#### Selection criteria

Studies that compared handsearching with searching one or more electronic databases to identify reports of randomized trials were included. Excluded were studies that used electronic searches to identify known RCTs.

#### Data collection and analysis

Two reviewers independently screened abstracts for inclusion then reviewed the full reports and assessed methodological quality. Disagreements were resolved by consensus. Two reviewers independently abstracted data with a third resolving disagreements. The reviewers contacted investigators to obtain missing information where required. Data were combined where studies appeared similar in design.

Meta-analysis of homogenous studies produced the outcome: proportion of trials retrieved by each search strategy over the total trials retrieved by both. Subgroup analysis [database searched, simple versus complex versus Cochrane HSSS, English versus non English reports and the type of report (full article, abstract or letter)] was performed.

#### Main results

Thirty-four studies were included, 27 as full reports, six abstracts and one unpublished report. Nine eligible studies await assessment.

Half of the studies conducted appropriate handsearching (including 11 in which handsearching was independently duplicated), with the remainder unclear. Of 34, 29 carried out appropriate electronic searches, with the remainder unclear. In 28 studies, comparable methods were used for judging eligibility of reports found by hand and electronically; for six comparability is unclear.

---

Handsearching identified 92% to 100% of the total number of reports of randomized trials found in the various comparisons in this review. Electronic searching of MEDLINE alone retrieved 55% of the total, EMBASE 49% and PsycINFO 67%. The retrieval rate of the electronic database varied depending on the complexity of the search. The Cochrane Highly Sensitive Search Strategy (HSSS) identified 80% of the total number of reports of randomized trials found. Searches categorised as 'complex' (including the Cochrane HSSS) found 65% and 'simple' found 42%. The retrieval rate for an electronic search was higher when both searches were restricted to English language journals; 62% versus 39% for journals published in languages other than English. When the search was restricted to full reports of randomized trials, the retrieval rate for the electronic search improved: a complex search strategy (including the Cochrane HSSS) retrieved 82% of the total number of such reports of randomized trials. Only one study compared the time taken per randomized trial retrieved: 17.20 minutes for handsearching versus 1.24 minutes for electronic searching.

### **Authors' conclusions**

Handsearching still has a valuable role to play in identifying reports of randomized trials for inclusion in systematic reviews of health care interventions, particularly in identifying trials reported as abstracts, letters and those published in languages other than English, together with all reports published in journals not indexed in electronic databases. However, where time and resources are limited, searching an electronic database using a complex search (or the Cochrane HSSS) will identify the majority of trials published as full reports in English language journals, provided, of course, that the relevant journals have been indexed in the database.

## **Expert MEDLINE searches yield half of relevant RCTs on average**

Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews.  
*BMJ* 1994; 309: 1286-1291.

### **STRUCTURED ABSTRACT**

*Prepared by Merrick Zwarenstein and peer reviewed by Eduardo Bergel and Elizabeth Paulsen.*

### **Background**

Unbiased and complete identification of studies is particularly important to the results obtained in a systematic review or meta-analysis. Even when studies are published, they may be difficult to find. The validity of MEDLINE as a source for retrieving randomized trials is unknown.

### **Question**

This publication incorporates two studies. The first is a study of a database of controlled and randomized trials of ophthalmologic treatments to estimate the proportion of published reports of trials in that database that cannot be readily identified within MEDLINE. This first study also asks a second question: Does searching using truncated text words improve the sensitivity for the search?

The second study within this publication is a systematic review of all studies (including the ophthalmologic database study reported in the first part of this paper) that compare the sensitivity and precision of MEDLINE-only searches versus wider searches and hand searches.

### **Search strategy**

The ophthalmologic database study was based upon a list of relevant trials obtained through non-standardised MEDLINE searches over a period of several years, and handsearching of 66 eye journals for the year 1988. This database was compared with the yield obtained from two standardised formal searches of MEDLINE, one search incorporating and one not incorporating truncated text words derived from an analysis of the articles obtained from the database. The systematic review was based upon a formal search of MEDLINE and EMBASE plus follow up of citations in the articles, a longstanding ad hoc search of MEDLINE, and a meeting of investigators in November 1992 which revealed some additional studies. The review includes the first above-mentioned study, also reported in detail in this same paper.

### **Selection criteria (for the systematic review)**

Studies were included if they compared a MEDLINE search with any one of three kinds of gold-standard collections of RCTs: all publications including those not indexed in MEDLINE, publications in journals indexed in MEDLINE, and publications in selected MEDLINE journals.

### **Data collection and analysis**

---

The sensitivity and precision of the MEDLINE search compared to the gold standard was extracted from the publications or obtained from the authors if additional data or clarification were needed. Results were combined by adding numerators and denominators to give weighted means.

### **Main results**

Sixteen studies were identified, including the one reported in the paper. Information useful to this review could be obtained for 15 of these. Eight studies used a gold standard of all publications whether indexed in MEDLINE or not. On average, these found that a MEDLINE search, even when done by a trained searcher, yielded an average of 51% of the RCTs (range: 17%-82%). Eight studies were restricted to journals indexed in the database. These found that the MEDLINE search yielded an average of 77% of RCTs (range: 32%-91%). Six studies investigated selected MEDLINE journals and found an average yield of 63% (range: 46%-88%). Eleven studies investigated the precision of MEDLINE searches revealing a wide range. Some found that thousands of citations would need to be retrieved to achieve satisfactory sensitivity while others needed relatively few citations to do so. The median precision was 33% (range: 2%-82%).

### **Authors' conclusions**

Indexing terms available for searching MEDLINE for randomized clinical trials have improved, but sensitivity still remains unsatisfactory. A mechanism is needed to "register" reports of trials, preferably by retrospective tagging of MEDLINE entries. Trials published before 1966 and in journals not indexed by MEDLINE should be incorporated into the system. As a result of this paper and work by the Cochrane Collaboration and others, this effort has now been implemented.

## **A Practihc/ESCORT Commentary**

Andy Oxman

A commentary addressing the methodology reviews:

- Hopewell S, Clarke M, Lefebvre C, Scherer R. Handsearching versus electronic searching to identify reports of randomized Trials. *The Cochrane Database of Methodology Reviews* 2002, Issue 4. Art. No.: MR000001. DOI: 10.1002/14651858.MR000001.
- Dickersin K, Scherer R, Lefebvre C. Identifying relevant studies for systematic reviews. *BMJ* 1994; 309: 1286-1291.

### **Are the results valid?**

The first review (Hopewell et al.) summarises evaluations of the sensitivity of MEDLINE searches for identifying reports of randomized trials compared to one of three types of gold standards: randomized trials published in any journal, whether indexed in MEDLINE or not; those published in any journal indexed in MEDLINE; or those published in a selected group of journals indexed in MEDLINE. A thorough search for relevant studies was conducted and it is unlikely that there was biased selection of studies for inclusion in the review. However, the included studies were published between 1985 and 1994 and may not reflect subsequent improvements in the indexing of randomized trials by MEDLINE. The second review (Dickersin et al.) summarises evaluations of the sensitivity of search strategies, including the Cochrane highly sensitive search strategy, using one or more electronic databases, including MEDLINE, compared to handsearching to identify reports of randomized trials.

### **What are the implications?**

The sensitivity of MEDLINE searches can, and has been improved by better indexing and better use of terms in the titles and abstracts of articles, as recommended by CONSORT. Improved strategies for identifying trials in MEDLINE can also help and such strategies have been developed by the Cochrane Collaboration and others.<sup>1,2</sup> It is uncertain to what extent these developments have improved the sensitivity of MEDLINE searches compared with the results of this review.

Hopewell et al. found that sensitivity was highest when precision was at or below 35%, and decreased as precision increased. Searches for trials need to find a balance between sensitivity and precision.

Other ways of improving the sensitivity of searches include using multiple databases, such as EMBASE and The Cochrane Central Register of Controlled Trials (CENTRAL), to identify trials published in journals not indexed

---

by MEDLINE, particularly reports of trials written in languages other than English. Searching registers of trials, such as those found in the metaRegister of Controlled Trials ([www.controlled-trials.com/mrct/](http://www.controlled-trials.com/mrct/)), may be useful for identifying unpublished trials.

Handsearching may be useful for identifying trials reported as abstracts or letters, published in languages other than English, or published in journals not indexed in electronic databases. Dickersin et al. found an overall difference in retrieval of 10% for complex, highly sensitive electronic search strategies compared with hand searching. However, handsearching is likely to be most efficient and practical when it is done to develop and maintain databases of reports of trials, as is being done by the Cochrane Collaboration.

#### **What are the implications for future methodology research?**

The results of the first review (Hopewell et al.) are likely out of date in light of more recent developments. A Cochrane methodology review of the sensitivity of MEDLINE search strategies is currently underway.<sup>3</sup> Some trials are missed by both complex, highly sensitive electronic searches and handsearching, particularly unpublished trials. Future research should assess the impact of trials that are not identified using alternative search strategies on the results of reviews.

#### **References**

1. Alderson P, Green S, Higgins JPT, editors. Highly sensitive search strategies for identifying reports of randomized controlled trials in MEDLINE. Cochrane Reviewers' Handbook 4.2.2 [updated December 2003]; Appendix 5b. In: The Cochrane Library, Issue 2, 2005. Chichester, UK: John Wiley & Sons, Ltd.
2. Haynes RB, et al. Summary of Enhancements for Clinical Queries for MEDLINE for Studies. [www.nlm.nih.gov/pubs/techbull/jf04/cq\\_info.html](http://www.nlm.nih.gov/pubs/techbull/jf04/cq_info.html) (accessed July 5, 2005). Revised January, 2005.
3. McDonald S, Crumley E, Eisinga A, Villanueva E. Search strategies to identify reports of randomized trials in MEDLINE. (Protocol) *The Cochrane Database of Methodology Reviews* 2004, Issue 2. Art. No.: MR000018. DOI: 10.1002/14651858.MR000018.



---

# INCREASING RESPONSE RATES TO POSTAL QUESTIONNAIRES

## A number of strategies can increase or decrease responses to postal questionnaires

Edwards P, Roberts I, Clarke M, DiGiuseppi C, Pratap S, Wentz R, Kwan I, Cooper R. Methods to influence response to postal questionnaires. *The Cochrane Database of Methodology Reviews* 2001, Issue 3. Art. No.: MR000008. DOI: 10.1002/14651858.MR000008.pub2. [Date of most recent update 30 June 2003]

### STRUCTURED ABSTRACT

*Prepared by Dave Sackett and peer reviewed by Max Petzold and Elizabeth Paulsen.*

#### **Background**

Postal questionnaires are widely used for data collection in epidemiological studies but non-response reduces the effective sample size and can introduce bias. Finding ways to increase response to postal questionnaires would improve the quality of health research.

#### **Question**

Which strategies increase response to postal questionnaires?

#### **Search strategy**

The authors searched 14 electronic databases; manually searched references of relevant trials and reviews; searched all issues of *Public Opinion Quarterly* and the *American Journal of Epidemiology*; contacted authors of trials or reviews to ask about unpublished trials.

#### **Selection criteria**

Randomized controlled trials of methods to influence response to postal questionnaires.

#### **Data collection and analysis**

Two reviewers independently extracted data from eligible reports using a standard proforma, with disagreements resolved by a third reviewer. Data were extracted on the trial participants, interventions, numbers randomized, and allocation concealment. For each strategy, pooled odds ratios and 95% confidence intervals were estimated in a random effects model. Selection bias was assessed with Egger's weighted regression, Begg's rank correlation test, and funnel plot. Heterogeneity among odds ratios was assessed with Chi-square at 5% level of significance.

#### **Main results**

372 eligible trials were found evaluating 98 different strategies. For 62 of these, the combined trials included over 1,000 participants. The results for individual incentives often displayed substantial heterogeneity.

The most effective strategies were monetary incentives (\$1 doubled the response rate, with diminishing marginal benefit from more money) and recorded delivery (doubled the odds of response).

Other effective strategies were pre-notification that a questionnaire was coming, shorter questionnaires (response to a 1-page questionnaire was twice that for a 3-pager), personalised questionnaires (signed by hand), university (vs. government or commercial) sponsorship, 'user-friendly' questionnaires, use of coloured (rather than blue or black) ink, non-monetary incentives (e.g., key ring, offer of results), enclosing the incentive with the questionnaire (rather than waiting for its return), personalized questionnaires, an assurance of confidentiality, use of stamped (rather than franked) envelopes, first class outward mailing, follow-up contact, and providing a second copy of the questionnaire at follow-up.

Response rates were reduced when the questionnaire included questions of a sensitive nature, when they began with the most general questions, and when participants were offered the opportunity to opt out of the survey.

Strategies that generated small, non-statistically significant increases in response were sending it from a more senior or well-known person, from an ethnically-unidentifiable person, coloured (vs. white) questionnaires, brown (vs. white) envelopes, including the subject's name or number on the questionnaire, presenting the questionnaire as a booklet rather than stapled pages, sticking to factual (vs. attitudinal) questions, providing first-

---

class return postage, sending the questionnaire to the subject's workplace rather than their home, and requesting an explanation if participants did not wish to respond.

Strategies that generated small, non-statistically significant decreases in response were the use of commemorative stamps, pre-contact by telephone rather than letter, stressing the benefit of the survey to its sponsor, giving a deadline for responding, or providing instructions for completing the questionnaire.

A strategy with no apparent effect was permitting "don't know" responses.

#### **Authors' conclusions**

A number of strategies can increase or decrease responses to postal questionnaires.

## **A Practihc/ESCORT Commentary**

Joel J. Gagnier

A commentary addressing the methodology review:

Edwards P, Roberts I, Clarke M, DiGuseppi C, Pratap S, Wentz R, Kwan I, Cooper R. Methods to influence response to postal questionnaires. *The Cochrane Database of Methodology Reviews* 2001, Issue 3. Art. No.: MR000008. DOI: 10.1002/14651858.MR000008.pub2. [Date of most recent update 30 June 2003]

The above review included 372 trials exploring methods influencing responses to postal questionnaires. The review outlined several strategies that increase or decrease responses to postal questionnaires.

#### **Are the results valid?**

This review performed comprehensive searches in a wide variety of databases, selection criteria for inclusion of studies were stated and appropriate, two reviewers scored the quality of the trials on a scale used by Schulz (1995), primary findings were combined using pooled odds ratio (testing for selection bias and heterogeneity), and the conclusions are supported by the data. Overall, this review provides valid methods and results. This is the most recent published methodological review on the topic.

#### **What are the implications?**

Investigators should consider using nominal monetary incentives and recorded delivery to improve response rates to postal questionnaires. Other strategies to consider that may improve response rates include: pre-notification, shorter questionnaires, personalized questionnaires, university sponsorship, 'user-friendly' questionnaires, use of colored ink, non-monetary incentives, enclosing the incentive with the questionnaire, personalized questionnaires, an assurance of confidentiality, use of stamped envelopes, first class outward mailing, follow-up contact, and providing a second copy of the questionnaire at follow-up. Investigators should not ask questions of a sensitive nature or begin postal questionnaires with the most general questions, as this will likely decrease response rates.

These findings provide investigators with specific suggestions towards improving responses to postal questionnaires.

#### **What are the implications for future methodology research?**

Future research could clarify the strategies that produced only small effects on responses to postal questionnaires. Additionally, the effects of combination strategies on response rates would aid investigators in designing and implementing postal questionnaires. Finally, cost savings analyses associated with the use (or not) of the above strategies would provide additional rationale for their use.

---

## INFORMATION FRAMING

### **Information framing may influence physicians' decision-making but the effects of information framing are unstable**

McGettigan P, Sly K, O'Connell D, Hill S, Henry D. The effects of information framing on the practices of physicians. *Journal of General Internal Medicine* 1999; 14:633-642.

#### STRUCTURED ABSTRACT

*From The Cochrane Collaboration Methods Groups Newsletter 2000 and peer reviewed by Max Petzold and Signe Flottorp.*

#### **Background**

The presentation format of clinical trial results, or the "frame", may influence perceptions about the worth of a treatment. The extent and consistency of this influence are unclear.

#### **Question**

What are the effects of information framing on the practices of physicians?

#### **Search strategy**

Relevant studies were retrieved using bibliographic databases and electronic searches (MEDLINE, PSYCLIT, CINAHL, CANCERLIT, The Cochrane Library).

#### **Selection criteria**

Specific study inclusion criteria do not appear to have been pre-defined and were determined following a preliminary review of potentially relevant studies. The strategy was refined by assimilation of additional key words in an iterative process.

#### **Data collection and analysis**

Information was extracted in relation to study design, frame type, parameter assessed, assessment scale, clinical setting, intervention, results, and factors modifying the frame effect. It was the author's intention to carry out a meta-analysis of compatible data sets but this was not possible because of inter-study variability in the parameters assessed and in the methods of analysing and reporting the outcomes.

#### **Main results**

Twelve articles reported randomized trials investigating the effect of framing on doctors' opinions or intended practices. Methodological shortcomings were numerous. Seven papers investigated the effect of presenting clinical trial results in terms of relative risk reduction, absolute risk reduction or the number needed to treat; four papers assessed gain/loss (positive/negative) terms; one paper assessed verbal/numeric terms. In simple clinical scenarios, doctors viewed results expressed in relative risk reduction or gain terms most positively. Factors that reduced the impact of framing included the risk of causing harm, pre-existing prejudices about treatments, the type of decision, the therapeutic yield, clinical experience and cost. No study investigated the effect of framing on actual clinical practice.

#### **Authors' conclusions**

While a framing effect may exist, particularly when results are presented in terms of proportional or absolute measures of gain or loss, it appears highly susceptible to modification, and even neutralisation, by other factors that influence doctors' decision making. Its effects on clinical practice are unknown.

### **A Practihc/ESCORT Commentary**

Signe Flottorp

A commentary addressing the methodology review:

---

McGettigan P, Sly K, O'Connell D, Hill S, Henry D. The effects of information framing on the practices of physicians. *Journal of General Internal Medicine* 1999; 14:633-642.

**Are the results valid?**

The review was well conducted, but the methodological shortcomings in the published studies greatly limited the capacity of the review to quantify the extent of the influence of information framing. Important shortcomings reported were lack of control groups, suboptimal techniques for allocating subjects to comparison groups and within-subject comparisons (response conditioning). The studies included only examined physicians' opinions or their intended clinical practice. The reviewers hence concluded that the effect of framing on actual clinical practice is undeterminable.

**What are the implications?**

Framing of information may influence decision-making. Based on this review we cannot tell to what extent the decisions may be affected, but authors should consider these issues when reporting the results of randomized trials or systematic reviews of trials.

**What are the implications for future methodological research?**

Future research should use rigorous design with adequate randomisation and control groups. Studies should be performed in a real clinical setting, testing actual clinical practice. To better understand how and to what extent different framings of information might influence clinical decision-making, it is also important to perform studies that evaluate the importance of framing, and its place relative to other factors affecting clinical decision-making.

---

## INFORMED CONSENT

### **Informed consent: more information results in more knowledge, but may also increase anxiety and reduce recruitment**

Edwards SJL, Lilford RJ, Braunholtz DA, Jackson JC, Hewison J, Thornton J. Ethical issues in the design and conduct of randomized controlled trials. *Health Technology Assessment* 1998;2(15).

#### **STRUCTURED ABSTRACT**

*Prepared by Yoon Loke and peer reviewed by Shaun Treweek and Merrick Zwarenstein.*

This abstract covers the question of which method of seeking consent from patients for participation in a trial is best, one of several sections in the HTA report but the only one in which a systematic review was conducted.

#### **Background**

Participants in clinical trials need to give voluntary informed consent prior to enrolment in the study. The method by which participants are informed of their options may have an effect on the recruitment rates.

#### **Question**

Do different methods of obtaining informed consent affect recruitment rates, as well as patients' knowledge, attitudes and anxiety level?

#### **Search Strategy**

MEDLINE, Psychlit and BIDS science and social science citation index were searched. Reference lists, handsearching and personal contact were also used. In addition, library databases in Birmingham and Edinburgh (HTA 93/43/02) were handsearched.

#### **Selection Criteria**

Ethics of clinical trials and empirical data

#### **Data collection and analysis**

One reviewer went through all the retrieved studies; half the studies were independently assessed by another reviewer. Both reviewers used a quality of data checklist; discrepancies were resolved through group discussion.

#### **Main Results**

Fourteen studies were identified; all except three were randomized trials. Half of the studies were based on hypothetical scenarios, while the other half looked at patients who were offered entry into "real" trials.

Eleven studies looked at recruitment rates, eight assessed understanding and seven evaluated psychological outcomes.

Of the seven studies looking at the effects of quantity of information, five found that provision of more information was associated with a lower or unchanged recruitment rate. Only two of the studies showed statistically significant reductions in the recruitment rate.

There is only limited data on participants' comprehension of trial concepts, but their understanding does appear to be enhanced through the provision of more information. Three studies looked at information about adverse effects with discrepant results – it may be that there is a threshold whereby excessive information has a negative impact, rather than a positive effect on understanding. Four studies evaluated anxiety levels – two showed that less information was associated with less anxiety, while the other two had neutral or opposite results.

#### **Authors' conclusions**

Giving participants more information usually results in greater knowledge, provided that there isn't an excessive amount of information. However, provision of additional information may provoke an increase in anxiety levels, and reduce the recruitment rate.

Other versions of this review:

---

Edwards SJ, Lilford RJ, Thornton J, Hewison J. Informed consent for clinical trials: in search of the "best" method. *Social Science and Medicine* 1998; 47:1825-1840.

## **Lack of empirical evidence of the process of informed consent in clinical trials and its impact on trial results**

Verheggen FWSM, van Wijmen FCB. Informed consent in clinical trials. *Health Policy* 1998;36:131-153.

### **STRUCTURED ABSTRACT**

*Prepared by Eduardo Bergel and peer reviewed by Shaun Treweek and Elizabeth Paulsen.*

#### **Background**

Informed consent is a key element for the protection of subjects in the context of a clinical trial. The process of informed consent is poorly understood.

#### **Questions**

- 1) What is the impact of the type of information disclosure on patient comprehension of the consent process?
- 2) What is the impact of the type of decision-making process on patient willingness to participate in a trial (real or hypothetical)?
- 3) What are the variables that affect accrual (recruitment) in clinical trials?
- 4) What is the impact of patient's refusal to enter in a trial on study results?

#### **Search strategy**

Based on early reviews, the authors searched in medical, legal and social science journals (list of journals not provided). A MEDLINE search was also performed using "informed consent", "clinical trial" and "experimentation" as keywords.

#### **Selection criteria**

Articles focusing on empirical research on informed consent published between 1979 and 1995. Studies of informed consent in psychiatric research were excluded. Non-English articles, letters and one-page news items were also excluded. Abstracts of articles from MEDLINE were reviewed for content before retrieval (criteria not stated). Additional selection criteria were applied to retrieved articles: Quantitative empirical research methods used in the analyses, research focus on informed consent in clinical trials, results and conclusions had to focus upon either motivation for patient participation, evaluation of information disclosure or patient compliance and satisfaction. It was not stated if articles were evaluated by more than one reviewer.

#### **Data collection and analysis**

Not stated. This paper does not include a quantitative analysis.

#### **Main results**

The authors identified about 150 articles, and 96 fulfilled the inclusion criteria and were included in the review. In relation to question 1, there is some evidence showing that research subjects often do not understand major elements to which they have consent. Memory or recall does not necessarily imply comprehension. Lengthy and complicated informed consent forms may obstruct the process. When asked, patients say they want to know everything, but there is no evidence that more information is associated with better patient satisfaction. Information disclosure does not seem to have an impact on patient participation in trials (question 2). Empirical data are unclear for questions 2, 3 and 4.

#### **Authors' conclusions**

Informed consent is difficult to define and, when defined, difficult to achieve in practice and may not be feasible. The concept of clinical research is complex and may not be readily understood by patients. Informed consent in clinical care and research has been accepted without actual proof that the process works. Questions have been raised about the efficacy of informed consent and about whether it is possible to determine if trial participants have been fully informed, and if they have given true consent. The small amount of empirical data currently available are insufficient to answer these questions definitively. Further research is needed to study the process of informed consent. Future studies should use sound research methods and focus on determinants for patient participation in clinical trials.

---

## A Practihc/ESCORT Commentary

Shaun Treweek

A commentary addressing the methodology reviews:

- Edwards SJL, Lilford RJ, Braunholtz DA, Jackson JC, Hewison J, Thornton J. Ethical issues in the design and conduct of randomized controlled trials. *Health Technology Assessment* 1998;2(15).
- Verheggen FWSM, van Wijmen FCB. Informed consent in clinical trials. *Health Policy* 1998;36:131–153.

Only the informed consent section of Edwards et al was included in the methodology abstract since this was the only question reviewed systematically.

Randomized controlled trials are widely considered the best way to evaluate the effect of healthcare interventions. But trials need participants and participants need information if they are to make an informed decision as to whether to take part. These two reviews aimed to compile empirical data on informed consent.

The two reviews reached similar conclusions. Research participants (overwhelmingly patients in these reviews) often have a very poor understanding of what they are getting themselves into when they join a trial. The lengthy consent forms currently in vogue with ethics committees may actually obstruct the process of understanding. Patients want to know everything and there is some evidence to suggest that increased information does enhance understanding. However, it seems that there may also be a threshold beyond which provision of information is counterproductive in terms of understanding, anxiety and recruitment.

### **Are the results valid?**

Both reviews had broad inclusion criteria although Edwards et al. is the more comprehensive of the two. Analyses were largely restricted to a qualitative summary of the included studies with the main limitation of both reviews being the paucity of evidence that was found. It is interesting to note that of the 812 articles that were initially picked up by Edwards' search, only 203 came via a standard electronic search, which highlights the importance of using additional methods such as searching personal collections. Non-English articles were excluded from both reviews. Verheggen et al. were somewhat unclear regarding their selection criteria.

### **What are the implications?**

Trialists need to think carefully about whether trial participants really understand the trial and its aims. There are no easy solutions to this problem and, as one set of reviewers concluded, some degree of failure may be inevitable. Nevertheless, trialists have an ethical duty to do all they can to maximise participant understanding and should provide information that is both sufficient and understandable to ensure adequate knowledge.

Trialists and potential participants who are concerned about the risks of participating in a trial should be reassured by a recent review by Gunn Vist and colleagues,<sup>1</sup> which found that participation in randomized controlled trials, independent of the clinical interventions being compared, is unlikely to be harmful. Their concerns can be restricted to the specific risks and benefits of the interventions that are being compared and what is known about those, and any specific inconveniences or benefits of participating in the trial.

### **What are the implications for future methodology research?**

Future research should focus on randomized trials of the impact of alternative ways of obtaining informed consent on understanding, anxiety and recruitment rates.

### **Reference**

1. Vist GE, Hagen KB, Devereaux P, Bryant D, Kristoffersen DT, Oxman AD. Outcomes of patients who participate in randomized controlled trials compared to similar patients receiving similar interventions who do not participate. The Cochrane Database of Methodology Reviews 2004, Issue 4. Art. No.: MR000009. DOI: 10.1002/14651858.MR000009.pub2.

---

## PLACEBO EFFECTS

### Colours affect the perceived action of drugs and may influence effects

de Craen AJ, Roos PJ, Leonard de Vries A, Kleijnen J. Effect of colour of drugs: systematic review of perceived effect of drugs and of their effectiveness. *BMJ* 1996; 313:1624-1626.

#### STRUCTURED ABSTRACT

*From The Cochrane Collaboration Methods Groups Newsletter 1999 and peer reviewed by Joel Gagnier and Yoon Loke.*

In addition to a systematic review, an observational study was also conducted. This structured abstract addresses only the systematic review.

#### Background

A patient's response to a drug is not simply a reflection of its chemical compounds. Colours are important in daily life, so it is suspected that the colour of a drug may influence patients' expectations and have an influence on its therapeutic effect.

#### Question

What is the impact of the colour of a drug's formulation on its perceived effects and its effectiveness?

#### Search Strategy

MEDLINE (1966-1995), EMBASE (1974-1995) and ClinPSYC (1980-1994) were searched with the MESH terms colour and pill, tablet, medication, capsule, or drug. Reference lists of relevant articles were also used and pharmaceutical companies and investigators were contacted.

#### Selection criteria

Studies examining the perceived stimulant action and perceived depressant action of different coloured drugs and studies that considered the influence of the colour of a drug on its effectiveness within a clinical trial.

#### Data collection and analysis

Methodological quality of "effects of coloured drug formulation studies" was assessed using a 10-point scale. A meta-analysis was intended of the six published studies on effects of coloured drugs formulations, but all trials had different designs and outcome measurements so this was not possible.

#### Main results

Six studies examined the perceived action of different coloured drugs. They showed that red, yellow, and orange are associated with stimulant effect, while blue and green are related to a tranquillising effect. Six trials assessed the impact of the colour of drugs on their effectiveness within clinical trials. They showed inconsistent differences between colours. The quality of the methods of these trials was variable.

#### Authors' conclusions

Although the trends were not always consistent, colours appear to affect the perceived action of a drug and seem to influence the effectiveness of a drug.

### Non specific factors can affect the results in clinical trials

Kleijnen J, de Craen AJM, Everdingen JV, Krol L. Placebo effect in double-blind clinical trials: a review of interactions with medications. *Lancet* 1994; 344:1347-1349.

#### STRUCTURED ABSTRACT

*Prepared by Eduardo Bergel and peer reviewed by Joel Gagnier and Yoon Loke.*

#### Background

---



---

There are three causes for clinical improvement: 1) the natural course of disease; 2) the changes induced by non-specific effects of the therapist and the setting in which therapy takes place; and 3) the specific effects of physical or pharmacological interventions. Non-specific effects can interact with specific factors. Interactions can be measured by comparing two situations that differ for one non-specific factor and one specific factor in a 2 by 2 factorial design.

### **Question**

Can the specific effect of medications be influenced by non-specific factors (i.e. informed consent, pill colour, positive and negative attitudes or expectations during treatment)?

### **Search Strategy**

The reviewers searched their collection of 300 articles to identify terms for a MEDLINE, EMBASE and ClinPSYC search. Dates not specified. Search terms included: “balanced placebo design”, “placebo effect; placebo response; placebo pharmacotherapy” combined with “expectancy; expectations; suggestion; attitude; persuasion”. Reference lists of retrieved articles and textbooks on placebo effect were scanned. Authors of articles and books on placebo effect were contacted for additional references.

### **Inclusion criteria**

Placebo controlled trials investigating specific effects of medication in fixed dosages in two situations that differ only for one non-specific factor. Articles with alcohol or acupuncture as an intervention were excluded.

### **Data collection and analysis**

Based on the title and abstract the reviewers selected articles to retrieve the full text. To assess the quality of the studies the authors used their own scoring system that included the following ten items: 1) Well-described inclusion criteria (diagnostic criteria, duration and severity of disease, and previous treatment); 2) At least 50 patients per group; 3) Random allocation procedure described; 4) Presentation of relevant baseline characteristics; 5) Less than 10% dropouts and dropouts described; 6) Intervention well described (nature, number, and duration of treatment); 7) Double blinding; 8) Effect measurement relevant and well-described; 9) Intention-to-treat analysis; 10) Presentation of the results in such a manner that the analysis can be checked. Criteria were graded + or -, or ± if unclear. Criterion 2 was included to reduce publication bias.

### **Main results**

The search yielded 1100 articles. Ten met the inclusion criteria. Quality was assessed as good for two trials, acceptable for two and low for six. Most trials indicated that specific effects can be modified by non-specific factors. One good quality trial studied the impact of informed consent on the comparative efficacy of an analgesic against placebo. One group was randomized using informed consent, and the other was not. The difference in effect size between the analgesic and placebo was higher in the uninformed patients. The other good quality trial looked at the influence of different treatment settings in a comparison of postoperative naloxone versus placebo. The substances were given intravenously by one of three methods (open, hidden, or machine infusion). In this trial non-specific effects changed the direction of the effect from positive to negative. The reviewers also reported, to some extent, the results of trials that do not fulfill the inclusion criteria.

### **Authors' conclusions**

Specific and non-specific factors are sometimes synergistic, and in other circumstances antagonistic, suggesting that the implicit additive model of the randomized trial is too simple. Interactions can be of a significant size. In designing studies, non-specific factors which strongly interact with specific treatment effects must be accounted for. Positive or negative expectations of treatment effect by patients and physicians should be investigated. These include, among other things, the effect of informed consent, pill colour, impressiveness of the treatment, and suggestion.

## **Placebo effects vary greatly, are frequently large, and may be misattributed to specific treatment effects**

Turner JA, Deyo RA, Loeser JD, VonKorff M, Fordyce WE. The importance of placebo effects in pain treatment research. *JAMA* 1994; 271(20): 1609-1614.

### **STRUCTURED ABSTRACT**

*Prepared by Joel Gagnier and peer reviewed by Eduardo Bergel and Signe Flottorp.*

---

**Background**

There are three general reasons for clinical improvement: natural history and regression to the mean; specific effects of the intervention; non-specific effects of treatment or, placebo effects. Clinicians and investigators require information on the contribution of placebo effects to patient improvement.

**Question**

What are the magnitude, duration, conditions that influence, and proposed explanations for the placebo effect?

**Search Strategy**

Databases searched included: MEDLINE (1980-1993) and PsycLIT (1967-1993). Searched personal files, bibliographies, and contacted experts in the field. Restricted to the English language. Search terms included: "placebo effect".

**Selection criteria**

Not reported.

**Data collection and analyses**

Not reported.

**Main results**

MEDLINE and PsycLIT searches yielded 195 articles. A total of three books and 75 articles were included in this review.

Placebo response rates varied across studies of sham treatments and of treatments initially (erroneously) believed efficacious for painful conditions. Two surgical trials indicate skin incision alone reduced angina; one trial found back pain relief in 43% and sciatica relief in 37% with negative surgical exploration. Non-specific influences plus natural history contribute to the pain relief effects of surgical interventions.

Studies indicate that placebos exhibit dose-response effects, carry-over effects, cumulative effects, time-effect curves, peak effects, and side effects. Placebo capsule size and colour influence perceived effects and injections may produce greater placebo responses than capsules. The duration of placebo effects is not clear.

Adverse effects (nocebo effects) of placebos have been estimated at 19% in drug trials. Placebos can also worsen existing conditions or produce pain in normal subjects.

Two studies indicate there is inconsistent placebo responding within individuals. Two studies show that patient expectations influence placebo responding. Also, a positive attitude toward provider and treatment can predict outcomes. Highly anxious subjects may have larger placebo responses and, as indicated in one trial, compliant individuals may have better outcomes, even when complying with a placebo. Two trials demonstrate that provider expectations of the treatment appear to influence placebo responses.

Though the role of anxiety in the placebo response is not clear, placebos seem to be most effective for highly anxious individuals and decrease anticipatory anxiety. Expectation of improvement may ameliorate symptoms by reducing anxiety, and may result in patients viewing pain as more controllable, noticing small improvements, disregarding negative events, interpreting ambiguous events positively, and increasing beneficial behaviours. Studies suggest that conditioned learning may be a component of the placebo effect and may be more powerful than expectancy. A variety of phenomena (drugs, people, places, procedures, etc.) may become conditioned stimuli for symptom alleviation. Past treatment experiences may influence responses to subsequent treatments. The role of endogenous opiate compounds as a mechanism for placebo responding remains unclear across four studies.

**Authors' conclusions**

Placebo effects influence patient outcomes after any treatment that the clinician and patient believe is effective. Placebo effects plus natural history and regression to the mean can result in high rates of good outcomes, which may be misattributed to specific treatment effects. The true causes of improvements in pain after treatment remain unknown in the absence of randomized controlled trials.

---

## A Practihc/ESCORT Commentary

Joel Gagnier

A commentary addressing the methodology reviews:

1. Kleijnen J, de Craen AJM, Everdingen JV, Krol L. Placebo effect in double-blind clinical trials: a review of interactions with medications. *Lancet* 1994;344:1347-1349.
2. Turner, JA, Deyo, RA, Loeser, JD, VonKorff, M, Fordyce, WE. The importance of placebo effects in pain treatment research. *JAMA* 1994; 271:1609-1614.
3. de Craen AJ, Roos PJ, Leonard de Vries A, Kleijnen J. Effect of colour of drugs: systematic review of perceived effect of drugs and of their effectiveness. *BMJ* 1996;313:1624-1626.

Three methodology reviews, conducted between 1994 and 1996, included a total of 91 original studies and three books. The three reviews addressed the following questions: Can the specific effect of medications be influenced by non-specific factors (i.e. informed consent, pill colour, positive and negative attitudes or expectations during treatment.)<sup>1</sup>, What are the magnitude, duration, conditions that influence, and proposed explanations for the placebo effect?<sup>2</sup>, What is the impact of the colour of a drug's formulation on its perceived effects and its effectiveness?<sup>3</sup>

### Were the results of the study valid?

All three reviews conducted comprehensive searches in relevant databases, retrieved article reference lists, personal files, and contacted experts in the field. Criteria for the inclusion of primary studies were listed and appropriate in one review<sup>1</sup> but not in the other two<sup>2-3</sup>. Although Kleijnen<sup>1</sup> did exclude a large number of alcohol and acupuncture trials (n=47) they were not related to the question of interest. Criteria to judge the quality of studies were reported and mostly appropriate in two reviews<sup>1,3</sup>. None of the three reviews reported methods to combine data from primary studies and did not report any findings of meta-analytic or synthesis methods. The conclusions were appropriate in two reviews which state that more research is required<sup>1,3</sup> whereas in the third review the conclusions seem to not extend beyond stating that the placebo effect is complex<sup>2</sup>. Overall two reviews had moderate flaws<sup>1,3</sup> and one review had major flaws<sup>2</sup>.

The three reviews covered in the commentary dated from 1994-1996, and this is a major limitation, as we are aware of many more new studies published since then. Additional methodology reviews are listed below (see additional reviews). The findings of these reviews may modify the findings presented in this commentary. Future commentaries will include the findings of these additional reviews.

### What are the implications?

Overall, given the small number of primary trials included in the reviews, the poor methods in these reviews, and the resultant lack of definitive conclusions in the reviews and original studies, we cannot offer any specific recommendations regarding placebos or placebo effects.

### What are the implications for future methodology research?

Future reviews must be careful to perform comprehensive searches, include only relevant trials, assess the methodological quality of the trials and attempt to statistically combine the primary data. These methods will improve the validity of methodology reviews. Future research should focus not only on the traditional aspects of placebo (i.e. use of pills, color, shape), but also on non specific effects like informed consent, and treatment expectation in patients and clinicians. In summary, more research is required to clarify the various attributes of the placebo effect.

### Additional reviews

Walach H, Jonas WB., Placebo research: the evidence base for harnessing self-healing capacities. *Journal of Alternative Complementary Medicine* 2004;10:S103-12

Hróbjartsson A, Gøtzsche PC. Placebo interventions for all clinical conditions. *The Cochrane Database of Systematic Reviews* 2004, Issue 2. Art. No.: CD003974. DOI: 10.1002/14651858.CD003974.pub2.

Hróbjartsson A, Gotzsche PC. Is the placebo powerless? Update of a systematic review with 52 new randomized trials comparing placebo with no treatment. *Journal of Internal Medicine* 2004;256:91-100.

- 
- Su C, Lichtenstein GR, Krok K, Brensinger CM, Lewis JD. A meta-analysis of the placebo rates of remission and response in clinical trials of active Crohn's disease. *Gastroenterology* 2004;126:1257-1269.
- Hróbjartsson A, Gøtzsche PC. Placebo interventions for all clinical conditions. *The Cochrane Database of Systematic Reviews* 2004, Issue 2. Art. No.: CD003974. DOI: 10.1002/14651858.CD003974.pub2.
- Zeller A, Estlinbaum T. [Placebo: an unappreciated factor in medicine] *Schweizerische Rundschau für Medizin Praxis* 2002; 91:1986-1991.
- Moyad MA. The placebo effect and randomized trials: analysis of alternative medicine. *Urologic Clinics of North America* 2002;29:135-155, x.
- Moyad MA. The placebo effect and randomized trials: analysis of conventional medicine. *Urologic Clinics of North America* 2002; 29:125-133, ix-x.
- Walsh BT, Seidman SN, Sysko R, Gould M. Placebo response in studies of major depression: variable, substantial, and growing. *JAMA* 2002; 287:1840-1847.
- Kirsch I, Scoboria A. Apples, oranges, and placebos: heterogeneity in a meta-analysis of placebo effects. *Mind Body and Medicine* 2001; 17:307-309.
- Greene PJ, Wayne PM, Kerr CE, Weiger WA, Jacobson E, Goldman P, Kaptchuk TJ. The powerful placebo: doubting the doubters. *Mind Body and Medicine* 2001; 17:298-307.
- Hróbjartsson A, Gotzsche PC. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. *New England Journal of Medicine* 2001; 24;344:1594-602.
- Thompson WG. Placebos: a review of the placebo response. *American Journal of Gastroenterology* 2000; 95:1637-1643.
- Weihrauch TR, Gauler TC. Placebo--efficacy and adverse effects in controlled clinical trials. *Rzneimittelforschung* 1999; 49:385-393.
- Weihrauch TR.[Placebo effect in clinical trials] *Medizinische Klinik (Munich)*. 1999; 94:173-181.
- Rochon PA, Binns MA, Litner JA, Litner GM, Fischbach MS, Eisenberg D, Kaptchuk TJ, Stason WB, Chalmers TC. Are randomized control trial outcomes influenced by the inclusion of a placebo group?: a systematic review of non-steroidal anti-inflammatory drug trials for arthritis treatment. *Journal of Clinical Epidemiology* 1999; 52:113-122.
- Kienle GS, Kiene H. Placebo effect and placebo concept: a critical methodological and conceptual analysis of reports on the magnitude of the placebo effect. *Alternative Therapies in Health and Medicine* 1996; 2:39-54.
- Piercy MA, Sramek JJ, Kurtz NM, Cutler NR. Placebo response in anxiety disorders. *Annals of Pharmacotherapy* 1996; 30:1013-1019.

---

## PUBLICATION BIAS

### **Trials with positive results are published sooner and more often than those with negative results**

Hopewell S, Clarke M, Stewart L, Tierney J. Time to publication for results of clinical trials. *The Cochrane Database of Methodology Reviews* 2001, issue 3. art. no.: mr000011. DOI: 10.1002/14651858.mr000011. [Date new studies sought but none found 20 May 2005]

#### STRUCTURED ABSTRACT

*Prepared by the review authors and peer reviewed by Signe Flottorp and Yoon Loke.*

#### **Background**

It has been suggested that a time-lag bias exists whereby research studies with striking results are more likely to be stopped earlier than originally planned, published quicker or both. If time-lag bias exists, such that the results of studies with positive findings become available sooner than those with null or negative findings, new interventions might be mistakenly assumed to be effective in the absence of evidence to the contrary.

#### **Question**

To what extent is the publication of a clinical trial influenced by the significance of its result?

#### **Search strategy**

Studies were identified by searching the Cochrane Methodology Register, MEDLINE, EMBASE, Science Citation Index and by handsearching journals and conference abstracts. The date of the most recent search was May 2005.

#### **Selection criteria**

A study was considered eligible for this review if it contained analyses of any aspect of the time to publication of clinical trials and tracked the publication of a cohort of clinical trials.

#### **Data collection & analysis**

Data extraction was performed independently by two reviewers. Data were extracted on the median time from the date the trial started to the date of publication of its results. Data were also extracted on the methodological quality of the included research study, the source of the clinical trials under investigation, the source of funding, the area of health care, the means by which the publication status of these trials were sought, and the rate of publication.

#### **Main results**

Two studies met the inclusion criteria. In both studies just over half of all trials had been published in full. Trials with positive results (i.e. those with statistically significant results ( $p < 0.05$ ) in favour the experimental arm of the trial) tended to be published in approximately 4 to 5 years. This was quicker than for those with null or negative results (i.e. those with results that were not statistically significant ( $p = > 0.05$ ) or statistically significant ( $p < 0.05$ ) in favour of the control arm of the trial) with a time to publication of around 6 to 8 years. One of the studies suggests that this difference could, in part, be attributed to the length of time taken to publish the results of a trial once follow up has been completed. This study showed that trials with null or negative findings took, on average, just over a year longer to be published than those with positive results.

#### **Authors' conclusions**

The review confirms that trials with positive results are published sooner than those with null or negative results. This has important implications for the timing of the initiation and updating of a review, especially if there is an association between the inclusion of a trial in a review and its publication status. It is of particular concern when one considers reviews in which only a small number of studies have been published and are available for the review.

---

## **Published trials are generally larger than grey trials and may show an overall greater treatment effect**

Hopewell S, McDonald S, Clarke M, Egger M. Grey literature in meta-analyses of randomized trials of health care interventions. *The Cochrane Database of Methodology Reviews* 2002, issue 4. art. no.: mr000010. DOI: 10.1002/14651858.mr000010. [Date of most recent update 21 November 2005]

### STRUCTURED ABSTRACT

*Prepared by the review authors and peer reviewed by Elizabeth Paulsen and Signe Flottorp.*

#### **Background**

The inclusion of grey literature (i.e. literature that has not been formally published) in systematic reviews may help to overcome some of the problems of publication bias, which can arise due to the selective availability of data.

#### **Question**

What is the impact of grey literature in meta-analyses of randomized trials of health care interventions?

#### **Search Strategy**

The reviewers searched the Cochrane Methodology Register (The Cochrane Library Issue 1, 2002), MEDLINE (1966 to February 2002), the Science Citation Index (April 2002) and contacted researchers who may have carried out relevant studies.

#### **Selection Criteria**

A study was considered eligible for this review if it compared the effect of the inclusion and exclusion of grey literature on the results of meta-analyses of randomized trials.

#### **Data collection and analysis**

Data were extracted from each report independently by two reviewers. The main outcome measure was an estimate of the impact of trials from the grey literature on the pooled effect estimates of the meta-analyses. Information was also collected on the area of health care, the number of meta-analyses, the number of trials, the number of trial participants, the year of publication of the trials, the language and country of publication of the trials, the number and type of grey and published literature, and methodological quality.

#### **Main Results**

Eight studies meeting the inclusion criteria were identified. Four studies contained multiple meta-analyses and four contained single meta-analyses. Of the included studies, four multiple and three single meta-analyses, found that published trials showed an overall greater treatment effect than grey trials. This difference was statistically significant in two of the four multiple meta-analyses. The remaining single meta-analysis found that published trials showed no effect of treatment and that grey trials showed a negative treatment effect; this difference was not statistically significant. Overall there were more published trials included in the meta-analyses than grey trials [median 46 (IQR 4-300) versus 5.5 (IQR 4-88)]. Published trials had more participants on average. In the two studies that assessed methodological quality of the included trials, the published trials were of higher quality than the grey trials. The most common types of grey literature were abstracts (49%) and unpublished data (33%).

#### **Authors' conclusions**

This review suggests that published trials are generally larger and may show an overall greater treatment effect than grey trials. This has important implications for reviewers who need to ensure they identify grey trials, in order to minimise the risk of introducing bias into their review.

## **Studies reported as abstracts are published more frequently in full if they have significant results**

Scherer RW, Langenberg P. Full publication of results initially presented in abstracts. *The Cochrane Database of Methodology Reviews* 2001, Issue 2. Art. No.: MR000005. DOI: 10.1002/14651858.MR000005.pub2. [Date of most recent update 23 November 2004]

### STRUCTURED ABSTRACT

---

---

*Prepared by the review authors and peer reviewed by Elizabeth Paulsen and Yoon Loke.*

### **Background**

Abstracts of presentations at scientific meetings are usually available only in conference proceedings. If subsequent full publication of abstract results is based on the magnitude or direction of study results, publication bias may result. Publication bias, in turn, creates problems for those conducting systematic reviews or relying on the published literature for evidence.

### **Questions**

What is the rate at which abstract results are subsequently published in full, and the time between meeting presentation and full publication?

What is the association between study characteristics and full publication?

### **Search strategy**

Searches included MEDLINE, EMBASE, the Cochrane Library, Science Citation Index, reference lists, and author files.

### **Selection criteria**

All reports that examined the subsequent full publication rate of biomedical results initially presented as abstracts or in summary form. Follow-up of abstracts had to be at least two years.

### **Data collection & analysis**

Two reviewers extracted data. The weighted mean full publication rate and time to full publication were calculated. Dichotomous variables were analysed using relative risk and random effects models. Time to publication was assessed using Kaplan-Meier survival analyses.

### **Main results**

Combining data from 79 reports (29,729 abstracts) resulted in a weighted mean full publication rate of 44.5% (95% confidence interval (CI) 43.9 to 45.1). Survival analyses resulted in an estimated publication rate at 9 years of 52.6% for all studies, 63.1% for randomized or controlled clinical trials, and 49.3% for other types of study designs.

'Positive' results defined as any 'significant' result showed an association with full publication (RR = 1.30; CI 1.14 to 1.47), as did 'positive' results defined as a result favouring the experimental treatment (RR = 1.17; CI 1.02 to 1.35), and 'positive' results emanating from randomized or controlled clinical trials (RR = 1.18, CI 1.07 to 1.30).

Other factors associated with full publication include oral presentation (RR = 1.28; CI 1.09 to 1.49); acceptance for meeting presentation (RR = 1.78; CI 1.50 to 2.12); randomized trial study design (RR = 1.24; CI 1.14 to 1.36); and basic research (RR = 0.79; CI 0.70 to 0.89). Higher quality of abstracts describing randomized or controlled clinical trials was also associated with full publication (RR = 1.30, CI 1.00 to 1.71).

### **Authors' conclusions**

There is clear evidence of publication bias in the step between presentation of a study at a meeting and subsequent full publication. Studies reported primarily as abstracts are published more frequently in full if their results show a positive effect of the experimental treatment or have significant results. Researchers performing systematic reviews should make every effort to obtain unpublished study results in order to avoid making biased reviews. Health care providers should be aware that conclusions derived from systematic reviews may be biased due to lack of inclusion of not fully published or unpublished data. Researchers initiating randomized controlled trials should register trials prospectively to ensure availability of trial results and should endeavour to publish trial results regardless of magnitude and direction of the effect size.

## **Studies with significant results are more widely disseminated than those with non-significant results**

Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases. *Health Technology Assessment* 2000; 4(10).

### **STRUCTURED ABSTRACT**

*Prepared by Andy Oxman and peer reviewed by Elizabeth Paulsen and Yoon Loke.*

---

---

## **Background**

Systematic reviews of published studies can be misleading if the published studies comprise a biased sample of all the studies that have been conducted. The direction or strength of study findings may influence the decision to publish (publication bias), and may also affect the subsequent dissemination of the data.

## **Question**

How extensive is publication and related biases, what are the consequences of these biases, what causes these biases and what are the effects of methods for reducing or detecting publication bias?

## **Search strategy**

The Cochrane Methodology Register, MEDLINE, EMBASE, BIDS, Library and Information Science Abstracts, PsycINFO, Sociofile, ERIC, Dissertation Abstracts, MathSci, British Education Index, SIGLE and ASSIA were searched up to September 1998. The reference lists of identified articles were checked. Experts in the field were contacted on an informal basis to identify relevant studies.

## **Selection criteria**

Any study if its main objectives involved any of the following issues: concepts, definition, causes, risk factors, existence and consequences of publication bias; and methods for preventing, reducing, detecting and correcting publication bias. Empirical evidence was defined as observations that could be used to reveal the existence, magnitude and consequences of publication and related biases.

## **Data collection and analysis**

The results of searches and the full text of potentially relevant studies were checked independently by two reviewers. Data from included studies were extracted by one reviewer and checked by another reviewer largely as free text without using explicit criteria to assess the quality of studies. The review did not focus specifically on randomized trials and results are for the most part not reported separately for trials.

## **Main results**

200 relevant articles were identified, including 64 containing empirical evidence. Empirical evaluation of four sets of registered trials found that those with statistically significant results were more likely to be published than those with non-significant results (overall odds ratio 2.54; 95% CI 1.44 to 4.47). Two studies found that pooled estimates of treatment effect based only on published results differed from estimates based on all registered trials.

Surveys of authors and members of professional organizations found that respondents were more likely to submit or publish statistically significant than non-significant results and that the most common reason for studies that were 'filed away' was non-significant results.

Empirical evaluation of three sets of trials found a longer time to publication for 'negative' versus 'positive' studies; another evaluation found that estimates of effect were larger in 20 of 26 meta-analyses of early trials compared with subsequent trials (average difference in relative odds 35%, 95% CI 15 to 55).

The definition of 'grey literature' varies, but five studies found similar biases to those of unpublished studies. The rate of full publication of studies presented in meeting abstracts ranged from 23 to 81% in 19 studies. Eight studies looked at the association between outcomes and subsequent publication but only one study found that abstracts with 'positive' results were significantly more likely to be published (OR 1.99, 95% CI 1.07 to 3.84).

There is limited evidence on dissemination bias, but researchers have found that significant results are more likely to be published in prominent journals, in English, and also in duplicate.

Three studies demonstrated the existence of selective reporting of significant outcomes.

Although there is no clear evidence for electronic database or indexing bias in retrieving trials, "negative" studies were less likely to be cited or covered in the media.

There appear to be many causes for publication bias ranging from investigators to peer reviewers, editors and funding bodies, but such bias is often due to the failure of researchers to submit results for publication. Although there are many statistical techniques for examining publication bias, these methods are unproven, and should only be used for sensitivity analyses.



---

### Authors' conclusions

Although the extent, direction and impact of publication and related biases are uncertain and may vary greatly, there is empirical evidence that studies with statistically significant or favourable results are more widely disseminated than those with non-significant or unfavourable results.

## A Practihc/ESCORT Commentary

Elizabeth Paulsen

A commentary addressing the methodology reviews:

1. Hopewell S, Clarke M, Stewart L, Tierney J. Time to publication for results of clinical trials. *The Cochrane Database of Methodology Reviews* 2001, Issue 3. Art. No.: MR000011. DOI: 10.1002/14651858.MR000011. [Date new studies sought but none found 20 May 2005]
2. Hopewell S, McDonald S, Clarke M, Egger M. Grey literature in meta-analyses of randomized trials of health care interventions. *The Cochrane Database of Methodology Reviews* 2002, Issue 4. Art. No.: MR000010. DOI: 10.1002/14651858.MR000010. [Date of most recent update 21 November 2005]
3. Scherer RW, Langenberg P. Full publication of results initially presented in abstracts. *The Cochrane Database of Methodology Reviews* 2005, Issue 2. Art. No.: MR000005. DOI: 10.1002/14651858.MR000005.pub2. [Date of most recent update 23 November 2004]
4. Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ. Publication and related biases. *Health Technology Assessment* 2000;4(10).

Two of the reviews above<sup>1,4</sup> both demonstrated that studies with 'positive' (statistically significant or favourable) results are published sooner and more often than those with 'negative' (non-significant) results.

One of the reviews<sup>4</sup> concluded that there is limited and often indirect evidence for full publication bias. The more recent review<sup>3</sup> combined data from 79 reports (29, 729 abstracts) and found evidence for full publication bias.

Two of the reviews<sup>2,4</sup> found that published trials may show an overall greater treatment effect than grey trials.

### Are the results valid?

All of the reviews stated the search methods used to find evidence, comprehensive searches were conducted in a variety of databases, and the results of searches were checked individually by two reviewers. The criteria used for selecting studies for the reviews and the criteria used for assessing the validity of the included studies were reported. Only one of the reviews<sup>3</sup> combined the data from the included studies in a meta-analysis.

### What are the implications?

- Systematic reviews should not be restricted to published studies
- All studies should be prospectively registered at their inception
- Researchers should publish trials results regardless of magnitude and direction of effect size
- The risk of publication bias should be assessed in all systematic reviews
- Reviewers should consider the possibility of time-lag bias when conducting a systematic review and update reviews when new studies become available.

Efforts are underway to support universal registration of all trials from onset and to identify and include unpublished as well as published data in systematic reviews.

### What are the implications for future methodology research?

Further research is needed about publication and related biases, specifically about the impact of publication bias on health decision making and outcomes of patient management, evaluation of existing methods for dealing with publication and related biases, and development of new methods to use for detecting publication bias in systematic reviews.

---

# QUALITY OF RANDOMIZED TRIALS

## Major weaknesses in scales and checklists for assessing the methodological quality of RCTs

Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Controlled Clinical Trials* 1995; 16:62-73.

### STRUCTURED ABSTRACT

*From The Cochrane Collaboration Methods Groups Newsletter 1998 and peer reviewed by Max Petzold and Merrick Zwarenstein.*

### Background

Assessing the quality of randomized controlled trials (RCTs) is important for those who might use the results and in the context of systematic reviews. Quality gives us an estimate of the likelihood that the results are valid.

### Question

What scales and checklists have been developed to assess the methodological quality of randomized controlled trials (RCTs)?

### Search strategy

MEDLINE search for 1966-1992. The search strategy included the key words: quality, clinical trials, scale, checklist and human. Wildcards were used for random\* and Medical Subject Headings (MeSH terms) were used for CLINICAL TRIALS and RANDOMIZED CONTROLLED TRIALS. Citations of retrieved articles were checked for additional scales and checklists. Some authors were contacted to ask whether they knew of any other scales or checklists.

### Selection criteria

Scales and checklists had to have been developed to measure the quality of RCTs. Scales had to have a numeric score attached to each item and to be able to produce an overall numeric score. Scales that were a minor modification of an existing one were not eligible.

### Data collection and analysis

The following information was recorded for each scale: name of scale and its principal author; whether it was developed to assess the quality of any trial or of specific types of trial; whether quality was defined; whether methodological quality or reporting quality was assessed; how items were selected; whether assignment, masking, follow-up and statistical analysis were included as items; number of items; whether the scale had undergone rigorous development; inter-rater reliability; the approximate time to complete the scale; range of possible scores; guidelines as to how a trial should be scored; and scores reported during scale development or in a meta-analysis. The following were recorded for each checklist: principal author; whether quality was defined; type of quality being assessed; how items were selected; whether assignment, masking, follow-up and statistical analysis were included as items; number of items; and approximate time to complete the checklist. The aim of this review was to produce an annotated bibliography; so combined analyses were not performed.

### Main results

Twenty-five scales were identified, 23 of which had been published. Nine checklists were identified, all of which had been published. Only six of the scales defined 'quality'. Eight scales and four checklists measured methodological quality alone (defined as the confidence that the trial design, conduct and analysis minimised or avoided biases in its treatment comparisons), three scales and three checklists measured the quality of reporting, and 14 scales and two checklists measured both. The number of items ranged from three to thirty-four for scales and from four to fifty-seven for checklists. The time required to score a trial ranged from less than 10 minutes to 45 minutes for the scales and from 10 to 30 minutes for the checklists. Inter-rater reliability was reported for 12 scales (intra-class correlation or Kappa range: 0.12-0.95). The overall average quality reported when the scales had been used in meta-analyses is approximately 50%. Only one scale satisfied the authors' criteria for rigorous development.

### Authors' conclusions

---

There are important shortcomings in scales and checklists that have been developed to assess the methodological quality of RCTs.

Other versions of this review:

Moher D, Jadad AR, Tugwell P. Assessing the quality of randomized controlled trials: current issues and future directions. *International Journal of Technology Assessment in Health Care*. 1996;12:195-208.

## **A Practihc/ESCORT Commentary**

Yoon Loke

A commentary addressing the methodology review:

Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Controlled Clinical Trials* 1995; 16:62-73.

This review evaluates the role of scales and checklists in measuring the quality of randomized controlled trials.

### **Are the results valid?**

This is a well-conducted review, which critically summarizes the key features (strengths and weaknesses) of the available tools for measuring quality. The authors used explicit criteria in assessing the validity of the quality tools.

While the search was comprehensive and well conducted, the selection process was not fully described. The inclusion and exclusion criteria are not explicitly stated, and other researchers may have difficulty replicating the methods. The paper also does not state whether articles were retrieved by two independent reviewers, or how the review team reached agreement on which articles to include.

However, this review only identified articles up to the final search date of December 1992, and there have been considerable developments in this field since then, particularly with the CONSORT statement.

### **What are the implications?**

There are major weaknesses in almost all the scales and checklists. Users of such quality tools may have difficulty with scoring the scales and completing checklists. It is also difficult to be certain what specific aspect of quality is being measured by these scales and checklists. Although the assessment of trial quality is of immense importance, the optimal scale or checklist has yet to be definitively established.

### **What are the implications for future methodology research?**

Rigorous methods should be applied to the development and validation of these quality assessment tools. Scales and checklists need to be consistent in their design, and easy to use.

---

## RANDOMIZATION

### Differences in participants in randomized versus non-randomized studies may influence effect estimates

Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Choosing between randomized and non-randomized studies: a systematic review. *Health Technology Assessment* 1998; 2(13). Chapters 2, 4 and 5 are most relevant to this abstract.

#### STRUCTURED ABSTRACT

*Prepared by Shaun Treweek and peer reviewed by Dave Sackett and Andy Oxman.*

This HTA report contains four research questions, only one of which is addressed in this structured abstract. Two of the remaining questions are addressed in separate structured abstracts: 'Do non-randomized studies differ systematically from randomized studies in terms of the size of the estimated treatment effect?' and 'How important is patient preference in terms of outcome?' The other question was not addressed by a methodology review: 'To what extent is it possible to adjust for baseline differences between study groups?'

#### Background

Studies that compare healthcare interventions can be divided into those that involve random allocation of participants to comparison groups and those that do not.

#### Question

There are two questions:

- a) Do included and excluded individuals or centres differ in healthcare trials?
- b) If so, do these influence the estimated size of treatment effect?

#### Search strategy

MEDLINE was searched up to 1996 to identify studies looking at eligibility criteria. Two systematic reviews from 1996 were used to investigate the issue of centre participation. The search strategy for selective participation of patients was not designed to identify comparisons between randomized and non-randomized evaluations and it is not therefore considered here.

#### Selection criteria

Cohorts of trials, systematic reviews, meta-analyses or case studies of healthcare evaluations that considered the effect of exclusion criteria and the selective participation of centres on the type of centres participating and on the estimated size of treatment effect in randomized versus non-randomized evaluations.

#### Data collection and analysis

One reviewer went through all abstracts retrieved by the search. Full copies of papers that appeared to meet the inclusion criteria were obtained. In addition, cited references and studies identified through contact with other researchers were obtained. All selected studies were then assessed independently by two of the reviewers. The results of comparisons were put into tables.

#### Main results

##### *Question a)*

Two studies comparing the characteristics of participants and eligible non-participants in two randomized and three non-randomized evaluations were identified. Both studies found small baseline differences between randomized and non-randomized patients for characteristics such as age, sex and onset of diabetes. It was unclear whether a further two studies compared randomized and non-randomized evaluations, or simply compared participants who were included in, or excluded from, the same randomized evaluation.

Two systematic reviews of surgical procedures considered the representativeness of the centres and clinicians taking part in 26 randomized and 41 non-randomized evaluations. The randomized evaluations were more likely to be single-centre studies based in university hospitals.

##### *Question b)*

---

One study found that mortality and mortality reduction due to treatment were less in a randomized evaluation than in a patient cohort selected using the same eligibility criteria as the randomized evaluation. A second cohort to which the randomized evaluation's exclusion criteria had been applied showed no such difference. A second study found that the main outcome for the 'conventional treatment' arm of a randomized evaluation and that of a 'trial-eligible' cohort converged over the period of the randomized evaluation.

### **Authors' conclusions**

The review found very little evidence upon which to make judgements on these two questions. Randomized evaluations seem to be more likely than non-randomized evaluations to involve single university and teaching centres. The estimated size of treatment effects seen in randomized evaluations and in patient cohorts selected using the same inclusion criteria as the randomized evaluation seem to be similar. However, the small amount of empirical data currently available are indeterminate and are insufficient to answer these questions definitively.

Other versions of this methodology review:

McKee M, Britton A, Black N, McPherson K, Sanderson C, Bain C. Interpreting the evidence: choosing between randomized and non-randomized studies. *BMJ* 1999; 319:312-315.

Sanderson C, McKee M, Britton A, Black N, McPherson K, Bain C. Randomized and non-randomized studies: threats to internal and external validity. In: Stevens A, Abrams K, Brazier J, Fitzpatrick R, Lilford R, editors. *The Advanced Handbook of Methods in Evidence Based Healthcare*. London: Sage Publications Ltd, 2001:95-115.

## **The effect of allocating by patient preference rather than randomization is not known**

Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Choosing between randomized and non-randomized studies: a systematic review. *Health Technology Assessment* 1998; 2(13). Chapters 2 and 6 are most relevant to this abstract.

### **STRUCTURED ABSTRACT**

*Prepared by Shaun Treweek and peer reviewed by Dave Sackett and Andy Oxman.*

This HTA report contains four research questions, only one of which is addressed in this structured abstract. Two of the remaining questions are addressed in separate structured abstracts: 'Do non-randomized studies differ systematically from randomized studies in terms of the size of the estimated treatment effect?' and 'Are there systematic differences between included and excluded individuals and do these influence the size of the estimated treatment effect?' The other question was not addressed by a methodology review: 'To what extent is it possible to adjust for baseline differences between study groups?'

### **Background**

Studies that compare healthcare interventions can be divided into those that involve random allocation of participants to comparison groups and those that do not.

### **Question**

Do patients assigned to treatments by random allocation experience different outcomes from patients assigned by their treatment preferences?

### **Search strategy**

MEDLINE was searched up to 1996.

### **Selection criteria**

Cohorts of trials or case studies or systematic reviews or meta-analyses of healthcare interventions that considered the effect of patient preferences on the treatment outcomes obtained in randomized and non-randomized evaluations.

### **Data collection and analysis**

One reviewer went through all abstracts retrieved by the search. Full copies of papers that appeared to meet the inclusion criteria were obtained. In addition, cited references and studies identified through contact with other

---

researchers were obtained. All selected studies were then assessed independently by two of the reviewers. The results from each comparison were put into tables.

### **Main results**

Four studies considering the effect of patient preferences in three randomized and three non-randomized evaluations were identified.

Two studies found no significant differences in treatment outcomes between patients who were randomized to treatment and those who selected a treatment based on their personal preferences. The remaining two studies did not present primary outcome data because these data were not available at the time of publication.

### **Authors' conclusions**

Very few studies have looked at the effect of patient preferences on treatment outcomes and those that exist are either small or not yet able to report full results. The small amount of empirical data currently available are indeterminate and cannot answer the question of whether patients assigned to treatments by random allocation experience different outcomes from patients assigned by their treatment preferences.

Other versions of this methodology review:

McKee M, Britton A, Black N, McPherson K, Sanderson C, Bain C. Interpreting the evidence: choosing between randomized and non-randomized studies. *BMJ* 1999; 319:312-315.

Sanderson C, McKee M, Britton A, Black N, McPherson K, Bain C. Randomized and non-randomized studies: threats to internal and external validity. In: Stevens A, Abrams K, Brazier J, Fitzpatrick R, Lilford R, editors. *The Advanced Handbook of Methods in Evidence Based Healthcare*. London: Sage Publications Ltd, 2001:95-115.

## **Randomized versus non-randomized allocation often leads to a different estimate of effect**

Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Choosing between randomized and non-randomized studies: a systematic review. *Health Technology Assessment* 1998; 2(13). Chapters 2 and 3 are most relevant to this abstract.

### **STRUCTURED ABSTRACT**

*Prepared by Shaun Treweek and peer reviewed by Dave Sackett and Andy Oxman.*

This HTA report contains four research questions, only one of which is addressed in this structured abstract. Two of the remaining questions are addressed in separate structured abstracts: 'Are there systematic differences between included and excluded individuals and do these influence the size of treatment effect?' and 'How important is patient preference in terms of outcome?' The other question was not addressed by a methodology review: 'To what extent is it possible to adjust for baseline differences between study groups?'

### **Background**

Studies that compare healthcare interventions can be divided into those that involve random allocation of participants to comparison groups and those that do not.

### **Question**

What is the effect of randomized versus non-randomized allocation on the estimated size of treatment effects in healthcare evaluations?

### **Search strategy**

MEDLINE, EMBASE, Science Citation Index, Social Science Citation Index and the Cochrane Database of Abstracts of Reviews of Effectiveness were searched up to 1996. References from key articles were also obtained, key authors identified via Citation Index and researchers doing reviews for the National Health Service Research and Development Standing Group on Health Technology were contacted. Personal communication was also used.

### **Selection criteria**

---

Case studies, systematic reviews or meta-analyses of healthcare interventions that compared outcomes of randomized versus non-randomized evaluations.

### **Data collection and analysis**

One reviewer went through all abstracts retrieved by the search. Full copies of papers that appeared to meet the inclusion criteria were obtained. In addition, cited references and studies identified through contact with other researchers were obtained. All selected studies were then assessed independently by two of the reviewers. The results from each comparison were put into tables. Authors were contacted when additional information was required. The results across studies were assessed qualitatively to identify common trends or discrepancies. Estimated treatment effect sizes were recalculated by the reviewers to allow direct comparison and where possible a significance test was done on the difference in the estimated treatment effect sizes.

### **Main results**

Eighteen studies compared the treatment effect sizes for 47 randomized and 57 non-randomized evaluations. The methodological quality of evaluations was not explicitly considered. Two relevant papers published after the review was completed are also discussed.

In 11 of the 18 studies that compared random and non-random allocation, significant differences were found in the estimates of treatment effect. There was no obvious pattern to these differences, though slightly more surgical interventions were found to have a bigger effect in randomized trials. The remaining seven studies found no significant differences in the estimated size of effects between random and non-random allocation. Differences in the estimated treatment effect size reached statistical significance for at least one outcome measure in seven of the ten studies where it was possible for the reviewers to do a significance test. Three of these studies showed a greater estimated size of treatment effect in the randomized trials, three showed a smaller effect size and one study had mixed results depending on the outcome measure.

### **Authors' conclusions**

There was no obvious pattern to the differences in the estimated size of treatment effects found for random and non-random allocation. The estimated size of treatment effect was most similar between random and non-random allocation when both used the same inclusion and exclusion criteria, potential prognostic factors were well understood and differences between the arms of the non-randomized evaluation were adjusted for in the analysis.

Other versions of this methodology review:

McKee M, Britton A, Black N, McPherson K, Sanderson C, Bain C. Interpreting the evidence: choosing between randomized and non-randomized studies. *BMJ* 1999; 319:312-315.

Sanderson C, McKee M, Britton A, Black N, McPherson K, Bain C. Randomized and non-randomized studies: threats to internal and external validity. In: Stevens A, Abrams K, Brazier J, Fitzpatrick R, Lilford R, editors. *The Advanced Handbook of Methods in Evidence Based Healthcare*. London: Sage Publications Ltd, 2001:95-115.

## **Non-randomized and randomized trials with inadequate concealment differ from those with adequate concealment**

Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. *The Cochrane Database of Methodology Reviews* 2002, Issue 4. Art. No.: MR000012. DOI: 10.1002/14651858.MR000012.

### **STRUCTURED ABSTRACT**

*Prepared by the review authors and peer reviewed by Dave Sackett and Andy Oxman.*

### **Background**

Randomized trials use the play of chance to assign participants to comparison groups. The unpredictability of the process, if not subverted, should prevent systematic differences between comparison groups (selection bias), provided that a sufficient number of people are randomized.

### **Question**

---

What are the effects of randomized versus non-randomized allocation and adequately versus inadequately concealed allocation on the results of healthcare trials?

### **Search Strategy**

The Cochrane Methodology Register, MEDLINE, SciSearch and reference lists were searched up to August 2000. Personal communication was also used.

### **Selection Criteria**

Cohorts of trials, systematic reviews or meta-analyses of healthcare interventions that compared outcomes or prognostic factors for one of the following comparisons: randomized versus non-randomized trials, randomized trials with adequately versus inadequately concealed allocation, or high versus low quality trials where selection bias could not be separated from other sources of bias.

### **Data collection and analysis**

One reviewer went through all of the citations in the Cochrane Methodology Register and accumulated reference lists. Studies that appeared to meet the inclusion criteria were retrieved and assessed independently by two of the reviewers. The methodological quality of included studies was appraised and information extracted by one reviewer and checked by a second. Tabular summaries of the results were prepared for each comparison and the results across studies were assessed qualitatively to identify common trends or discrepancies.

### **Main Results**

Thirty-two studies including over 3000 trials were identified. Twenty-two studies compared randomized versus non-randomized trials, three compared adequately versus inadequately concealed allocation, and nine compared high versus low quality trials (some studies included more than one comparison). Five studies were of high methodological quality.

In 15 of the 22 studies that compared randomized and non-randomized trials of the same intervention, important differences were found in the estimates of effect. Some of these differences were due to a poorer prognosis in the control groups in the non-randomized trials. The results of the other seven studies that compared randomized and non-randomized trials across different interventions are less clear.

Comparisons of adequately and inadequately concealed allocation in randomized trials of the same intervention provided high quality evidence that concealment can be crucial in achieving similar treatment groups and, therefore, unbiased estimates of treatment effects. Studies with inadequate concealment tended to overestimate treatment effects.

Comparisons of high and low quality trials of the same intervention have found important differences in estimates of effect, but it is not possible to determine the extent to which these differences can be attributed to randomisation or concealment of allocation.

Omitting comparisons between randomized trials and non-randomized trials using historical controls did not substantially alter the results or conclusions of the review.

### **Authors' conclusions**

On average, non-randomized trials and randomized trials with inadequate concealment of allocation tend to result in larger estimates of effect than randomized trials with adequately concealed allocation. However, it is not generally possible to predict the magnitude, or even the direction, of possible selection biases and consequent distortions of treatment effects.

Other versions of this review:

Kunz R, and Oxman A. The unpredictability paradox: review of empirical comparisons of randomized and non-randomized clinical trials. *BMJ*. 1998; 317: 1185-1190.

## **Estimates of effect from non-randomized studies may be valid if important confounding factors are controlled for**

MacLehose RR, Reeves BC, Harvey IM, Sheldon TA, Russell IT, Black AMS. A systematic review of comparisons of effect sizes derived from randomized and non-randomized studies. *Health Technology Assessment* 2000; 4(34): Chapter 2, Strategy 1.



---

## STRUCTURED ABSTRACT

*Prepared by Andy Oxman and peer reviewed by Elizabeth Paulsen and Dave Sackett.*

### **Background**

Comparisons between randomized controlled trials (RCTs) and quasi-experimental and observational (QEO) studies are often cited selectively, may be unsystematic and may have failed to distinguish between different explanations for any discrepancies observed.

### **Question**

What is the association between the methodological quality of QEO studies and the magnitude of estimates of effectiveness relative to estimates from RCTs?

### **Search strategy**

MEDLINE and EMBASE were not found to be helpful. All of the abstracts in four databases were screened: a database from another HTA Methodology Review (Prescott 1999), the Cochrane Methodology Register, the Cochrane Database of Systematic Reviews, and DARE. Personal files were screened, other experts were contacted and the reference lists of relevant papers were screened.

### **Selection criteria**

A comparison of RCT and QEO study estimates of effectiveness for specified interventions, where the estimates were reported in a single paper.

### **Data collection and analysis**

Study quality was scored using a checklist to assess whether RCT and QEO study estimates were derived from the same populations, whether the assessment of outcomes was 'blinded', and the extent to which the QEO study estimate took account of possible confounding. QEO study estimates were classified as high or low quality. Seven indices of the size of discrepancies between estimates of effect size and outcome frequency were calculated, where possible, for each comparison. Distributions of the size and direction of discrepancies were compared for high- and low- quality comparisons.

### **Main results**

Fourteen papers were identified, yielding 38 comparisons between RCT and QEO study estimates. 25 were classified as low and 13 as high quality. Discrepancies between RCT and QEO study estimates of effect size and outcome frequency for intervention and control groups were smaller for high- than low-quality comparisons. For high-quality comparisons, no tendency was observed for QEO study estimates of effect size to be more extreme than RCT ones, but this tendency was seen with low-quality comparisons.

### **Authors' conclusions**

QEO study estimates may be valid if important confounding factors are controlled for. Treatment preferences and willingness to be randomized had a negligible effect on outcome. However, few papers were reviewed, the findings may depend on the specific interventions evaluated, and there is likely to be a publication bias related to a priori views of the authors of the included studies.

## **A Practihc/ESCORT Commentary**

Dave Sackett

The number of reviews and primary publications on this topic are burgeoning, and this commentary can only update this rapidly developing field to March 2005.

The reviews of methodology studies available at this time are:

1-3. Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Choosing between randomized and non-randomized studies: a systematic review. *Health Technology Assessment* 1998;2(13). Three reviews have been generated from this report.

4. Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. *The Cochrane Database of Methodology Reviews* 2002, Issue 4. Art. No.: MR000012. DOI: 10.1002/14651858.MR000012.

---

5. MacLehose RR, Reeves BC, Harvey IM, Sheldon TA, Russell IT, Black AMS. A systematic review of comparisons of effect sizes derived from randomized and non-randomized studies. *Health Technology Assessment* 2000; 4(34).

This resource contains abstracts summarizing the foregoing five reviews. However, two other pertinent reviews have just been published:

6. King M, Nazareth I, Lampe F, Bower P, Chandler M, Morou M, Sibbald B, Lai R. Impact of participant and physician intervention preferences on randomized trials; a systematic review. *JAMA* 2005;293:1089-1099.

They concluded: "Preferences influence whether people participate in randomized trials, but there is little evidence that they significantly affect (internal or external) validity."

7. Vist GE, Hagen KB, Devereaux PJ, Bryant D, Kristoffersen DT, Oxman AD. Systematic review to determine whether participation in a trial influences outcome. *BMJ* 2005; 330:1175.

They concluded: "Patients who participate in RCTs can expect similar outcomes as if they receive similar treatment outside of the trial. These results suggest that the results of RCTs are applicable to usual clinical practice."

This rapidly growing number of reviews (whose primary studies sometimes display little overlap) have documented and attempted to explain similarities and differences in the direction and size of treatment effects of the same interventions derived from groups of randomized and non-randomized studies. Taken as a whole, they support the conclusion that it is not possible to predict differences in the size, or even the direction, of estimates of treatment effects for the same intervention when it is generated in randomized and non-randomized studies.

However, especially in the more recent reports,<sup>5-7</sup> there is the suggestion that these disparities decrease when investigators have controlled for known confounders (between risk/responsiveness and treatment).

#### **Are the results valid?**

The absence of a common database for these reviews makes it impossible to reconcile their varying conclusions.

#### **What are the implications?**

Because randomized and non-randomized studies of the same intervention can disagree about both the direction and the size of treatment effects, and if one believes that the most valid estimates of treatment effects come from randomized trials, treatments not yet validated in randomized trials (unless they exhibit "all or none" effects should be tested in randomized trials before general use. By "all or none" effects I suggest that any treatment of a universally fatal condition ("all") that is followed by survival, or any treatment of an occasionally fatal condition that is invariably followed by survival ("none") has reduced uncertainty about its efficacy to near-zero, and need not undergo a randomized trial to confirm it).

#### **What are the implications for future methodology research?**

High priority should be given to generating and periodically updating a common data set of primary studies.

---

# RECRUITMENT FOR TRIALS

## A number of factors influence recruitment into clinical trials

Lovato LC, Hill K, Hertert S, Hunninghake DB, Probstfield JL. Recruitment for controlled clinical trials: literature summary and annotated bibliography. *Controlled Clinical Trials* 1997;18:328-352. (For most recent update see [www.fhcrc.org/phs/swog/recrcct/](http://www.fhcrc.org/phs/swog/recrcct/))

### STRUCTURED ABSTRACT

*Prepared by Yoon Loke and peer reviewed by Dave Sackett and Joel Gagnier.*

#### **Background**

The successful completion of a clinical trial depends upon recruitment of the appropriate number of participants within the designated time frame. Participants should be recruited at a steady rate to maintain power. Representative recruitment of gender and ethnic groups is also important.

#### **Question**

What are the factors that influence recruitment of participants into clinical trials?

#### **Search Strategy**

MEDLINE was searched from October 1986 to December 1995. Relevant articles prior to October 1986 were identified from a previous review published in 1987 and also through checking reference lists of retrieved papers. The publications list of the Office of Minority Health was also searched.

#### **Selection Criteria**

Included articles were those that explored recruitment issues in controlled clinical trials, particularly those that contain a data-based examination. Selected articles had to add to the knowledge-base or to confirm a previously published result. Articles were excluded if there were already two papers on the same topic.

#### **Data collection and analysis**

Reviewers checked the previously published review to identify articles that were considered relevant. Articles from the MEDLINE search were initially screened to exclude those that did not explore recruitment issues in detail. The remaining articles were sent out for further evaluation by reviewers.

#### **Main Results**

A total of 91 studies were identified. The studies encompassed the following broad areas:

**Diverse populations:** Lack of familiarity and trust in research, coupled with language barriers, need to be addressed when recruiting ethnic minorities. Greater enrolment of female participants may be achieved by having female investigators and providing more flexible time schedules.

**Recruitment strategies:** Potential participants may be readily identified through computerized disease registries, or via occupational screening schemes. Direct mailing is of limited value - choosing the most appropriate mailing list is a key component here. In such instances, recruitment may be improved through personalized letters, follow-up telephone calls, and a coordinated media campaign.

**Planning and management:** Initial projections of the recruitment rates may prove overly-optimistic, and even a pilot phase may yield falsely reassuring figures. Continuous monitoring and tracking of recruitment rates can identify problem areas early on, so that appropriate action can be taken. A skilled recruitment coordinator is essential.

**Participant and physician attitudes:** There are many barriers that hinder patient participation; e.g. lack of knowledge and trust, difficulties with time and transportation, unwillingness to be randomized, and difficulty in informed consent. Barriers to physician participation include time and financial constraints, disagreement with the trial protocol, and intrusiveness of the research process on relationships with patients.

---

Generalizability and adherence: Overly-strict inclusion criteria may hinder recruitment of diverse populations, thus limiting generalizability. Participant drop-outs should be minimized – pre-randomization adherence procedures may help, but these techniques remain controversial.

Cost of recruitment: Costs may be difficult to project and are often under-estimated. Participants' failure to attend follow-up may contribute substantially to escalating costs.

### **Authors' conclusions**

The following areas are considered important in the successful recruitment of participants – an overall plan, identification and elimination of barriers, logistical development of strategies, and awareness of special problems in recruiting to specific areas such as trials of prevention, as well as certain diseases like HIV.

## **Uncertainty about the effects of most strategies to improve recruitment into health-care studies**

Mapstone, J, Elbourne, D, Roberts, I. Strategies to improve recruitment to research studies. *The Cochrane Database of Methodology Reviews* 2002, Issue 3. Art. No.: MR000013. DOI: 10.1002/14651858.MR000013.pub2

### **STRUCTURED ABSTRACT**

*Prepared by Joel Gagnier and peer reviewed by Yoon Loke and Dave Sackett.*

### **Background**

Many studies fail to recruit the planned number of participants. Determining how to increase recruitment would improve research.

### **Question**

What are the contributions of strategies aimed at improving participation in research studies?

### **Search Strategy**

Databases searched: MEDLINE (1966-2001/11), Cochrane Library (2002, Issue 1), EMBASE (1980-Jan 2002), Spectr (1950-2002/02), PsycINFO (1887-Jan 2002), ERIC (1966- Jan 2002), CINAHL (1982-2001/11), National Research Register (2001, Issue 4), ISI Science Citation Index/Social Science Citation Index/Proceedings (1981-2002/02). Reference lists from retrieved trials were searched.

### **Selection Criteria**

Randomized or quasi-randomized controlled trials testing any intervention. Outcomes considered: proportion of participants, centres or researchers recruited, proportion of patients with full follow-up, number of people who agree to take part, difficulties/benefits. Excluded: Studies on improving questionnaire responses.<sup>1</sup>

### **Data Collection and analysis**

One reviewer screened titles and abstracts. Two individuals independently determined inclusion for full reports, assessed allocation concealment and extracted data. Disagreements were resolved by consensus. Trials were grouped by intervention. A test of statistical heterogeneity was performed. Heterogeneous groups of trials were described qualitatively. Risk ratios (RR) and 95% confidence intervals (CI) were calculated. Funnel plots were used to investigate publication bias.

### **Main Results**

Fifteen trials including 33,719 participants were included. 13 trials were randomized and 2 quasi-randomized. All studies tested recruitment strategies aimed at participants. Quality of allocation concealment was high in 4, unclear in 9 and poor in 2 trials. Owing to heterogeneity between trials and within strategies statistical pooling was not carried out.

Pre-warning (information prior to recruitment approaching) was tested in three trials (Total N=24,626). No influence was found for a letter sent prior to an enrolling phone call (RR 0.98, CI 0.91 to 1.06) or sending postcards prior to an eligibility questionnaire (RR 1.00, CI 0.83 to 1.20). A non-significant positive trend was found for leaving an answering machine message (RR 1.42, CI 0.89 to 2.27).

---

Extra information at recruitment was evaluated in 7 trials (N=3512). Non-significant positive trends were found for a one page letter compared to a standard leaflet (RR 9.83, CI 0.58 to 168.05), a video explaining study importance (RR 1.45, CI 0.92 to 2.30), extra cancer-specific information (RR 1.50, CI 0.29 to 3.62), or a booklet completed during follow-up. Recruitment was improved for providing a home safety questionnaire in a trial of safety risks (RR 1.38, CI 1.14 to 1.67) and for positive wording in consent forms (RR 1.62, CI 1.10 to 2.37). A telephone call explaining the consent process showed a trend towards lower recruitment (RR 0.87 CI 0.76 to 1.01).

The effect of study design (changes to study method) was tested in two trials (N=709). A non-significant negative trend was found for a lack of placebo arm (RR 0.95 CI 0.63 to 1.44), and for those agreeing to be randomized in a partially randomized patient preference design (PRPP; RR 0.95, CI 0.81 to 1.11) compared to a traditional RCT. Recruitment was better in the PRPP arm (RR 1.37, CI 1.22 to 1.53).

Consent change (changes to consent process) was tested in two trials (N=826). A non-significant positive trend was found for opting-out consent versus opting-in consent (RR 1.07, CI 0.81 to 1.41) or for five separate consent scenarios.

Incentives were evaluated in a smoking prevention trial (N=4046). Improved recruitment was found for monetary incentives with/after questionnaire completion (RR 1.43, CI 1.27 to 1.62: RR 1.53, CI 1.36 to 1.72, respectively) and being entered into a draw for prizes (RR 1.36, CI 1.20 to 1.54) compared to no incentive.

### **Authors' conclusions**

Though there are specific strategies that may improve recruitment, it is not possible to predict the effect of most of the reviewed interventions. The generalizability of these strategies to other contexts is questionable. These findings should be replicated in other contexts.

1. This is covered in another review: Edwards P, Roberts I, Clarke M, DiGiuseppi C, Pratap S, Wentz R, Kwan I, Cooper R. Methods to influence response to postal questionnaires. *The Cochrane Database of Methodology Reviews* 2001, Issue 3. Art. No.: MR000008. DOI: 10.1002/14651858.MR000008.pub2. [Date of most recent update 30 June 2003]

## **Many barriers to participation in randomized controlled trials by patients and clinicians**

Ross S, Grant A, Counsell C, Gillespie W, Russell I, Prescott R. Barriers to participation in randomized controlled trials: a systematic review. *Journal of Clinical Epidemiology* 1999;52:1143-1156.

### **STRUCTURED ABSTRACT**

*Prepared by Dave Sackett and peer reviewed by Yoon Loke and Joel Gagnier.*

### **Background**

Randomized controlled trials (RCTs) suffer when patients refuse invitations to participate, when investigators fail to recruit enough patients, and when clinical collaborators fail to adhere to trial protocols.

### **Question**

What are the barriers to participation in RCTs for patients and clinicians?

### **Search strategy**

MEDLINE, EMBASE, and CINAHL were searched from 1986 to 1996. 265 papers identified clinician or patient participation as an important issue.

### **Selection criteria**

From the foregoing they selected 78 primary research papers with empirical quantitative or qualitative data on "barriers" in recruiting clinicians or patients into RCTs.

### **Data collection and analysis**

Two readers reviewed each paper and abstracted the study design and relevant findings. "An iterative process of analysis was used to identify and describe factors thought to act as barriers."

---

## Main results

Barriers to patient participation were: concern that the (additional) trial procedures might be uncomfortable, inconvenient, or costly; travel time, effort, and costs; the lack of payment for participation; strong preference for one of the treatment arms (didn't want to change their current treatment, didn't want to take any medication, didn't want to take a placebo, or didn't want to take an "experimental" drug); aversion to treatment assignment through randomization; uncertainty about unproven treatments and fear of the unknown; disapproval of the trial by spouse, family member, or friend; low confidence in, or dislike of, their clinician; distrust of hospitals or medicines; and preference that treatment decision be made by their clinician.

Although both participating and non-participating patients wanted more information on a trial, providing information was thought to reduce participation and alter treatment responses. Clinicians reported concerns that information might cause fear and increased morbidity and mortality. The most common motivation for patients' participation was altruism. The authors concluded that there was no consistent effect of illness severity, age, or level of education on patient participation.

Barriers to clinician participation were: lack of a sufficiently interesting trial question; preference for one of the treatments; unwillingness to recruit patients to a "no treatment" arm; insufficient time to recruit, consent, and follow trial patients; excessive and poorly designed data collection; lack of staff (e.g., trial nurses) or a stable research team; lack of training and experience in RCT participation; perceived lack of benefit to their or their institution's reputations; lack of financial reward; lack of personal encouragement and support; the consent process itself; loss of clinical autonomy (concerning decision-making, loss of independence, inability to individualize patient care, and being accountable to a third party); perceived lack of skill in applying both trial treatments; damage to rapport with patients from having to admit uncertainty and from conflicts between their clinical and research roles; concern that their patients might suffer toxicity and side-effects; reluctance to recruit severely ill patients; reluctance to withdraw patients from maintenance therapy before randomisation; concern about their patients' time and travel costs; and a fear of feeling responsible if their patient failed to receive the treatment that was later found to be best.

## Authors' conclusions

Patient participation in RCTs might be improved by minimizing the demands on their time and effort, by clearly explaining the purpose of the trial and the necessary investigations, by providing support from dedicated research staff as they make their participation decision, and by not pressuring them to consent.

Clinician participation in RCTs might be improved by posing an interesting and pragmatic question, and by providing them a clear and simple protocol that requires straightforward data collection and makes minimum demands on their time (by providing research support staff if needed).

## A Practihc/ESCORT Commentary

Yoon Loke

A commentary addressing the methodology reviews:

- Lovato LC, Hill K, Hertert S, Hunninghake DB, Probstfield JL. Recruitment for controlled clinical trials: literature summary and annotated bibliography. *Controlled Clinical Trials* 1997;18:328-352. (For most recent update see [www.fhcrc.org/phs/swog/recrcct/](http://www.fhcrc.org/phs/swog/recrcct/))
- Mapstone, J, Elbourne, D, Roberts, I. Strategies to improve recruitment to research studies. *The Cochrane Database of Methodology Reviews* 2002, Issue 3. Art. No.: MR000013. DOI: 10.1002/14651858.MR000013.pub2.
- Ross S, Grant A, Counsell C, Gillespie W, Russell I, Prescott R. Barriers to participation in randomized controlled trials: a systematic review. *Journal of Clinical Epidemiology* 1999;52:1143-1156.

Three methodology reviews looked at factors that influenced the recruitment of participants into clinical trials. None of the reviews attempted to statistically pool the results; they generally provided a descriptive summary of individual studies. The reviews did not reach any consistent conclusions as to which aspects of recruitment were most important in ensuring adequate participant numbers.

---

**Are the results valid?**

All three methodology reviews conducted electronic searches in relevant databases. The three reviews had different criteria for selecting studies for inclusion, although they broadly aimed to select those that presented data on recruitment strategies. Lovato's inclusion criteria may not have led to a complete dataset as "selected articles had to add to the knowledge-base or to confirm a previously published result, and articles were excluded if there were already two papers on the same topic." The primary aim of Lovato's review was to produce an "annotated bibliography...to make documented information on recruitment in clinical trials easily accessible", and the selection criteria were not systematic given that inclusion of articles from the previously published bibliography appeared to be based on a subjective level of interest.

Both Lovato and Ross did not specify particular study designs, whereas Mapstone limited the review to randomized or quasi-randomized controlled trials testing any intervention. The three reviews did not attempt to systematically assess the methodological quality of the primary studies.

**What are the implications?**

None of the reviews were able to reach a consistent conclusion on the optimal recruitment strategy. All three reviews identified a host of factors that had been shown to influence recruitment. However, it is not known which factors are relevant to specific situations. A wide range of options to minimize barriers to participation need to be considered when designing a recruitment strategy. Logical, but largely unevaluated strategies that should be considered include: identification and elimination of specific barriers and minimizing the demands on patients' and clinicians' time and effort.

**What are the implications for future methodology research?**

Randomized trials of promising recruitment strategies across a range of clinical trials are needed.

---

# STRATIFIED RANDOMIZATION

## When, why and how to stratify in clinical trials

Kernan WN, Viscoli CM, Makuch RW, Brass LM, Horwitz RI. Stratified randomization for clinical trials. *Journal of Clinical Epidemiology* 1999; 52:19-26.

### STRUCTURED ABSTRACT

*Prepared by Max Petzold and peer reviewed by Merrick Zwarenstein and Shaun Treweek.*

#### **Background**

Stratified randomization is a two-stage procedure in which patients who enter a clinical trial are grouped into strata/blocks according to clinical features that may influence the outcome. Stratified randomization prevents imbalance between treatment groups for known predictors and may prevent misleading inference.

#### **Question**

What are the purposes, indications, accomplishments, risks and alternatives of stratified randomization?

#### **Search strategy**

MEDLINE and reference lists were searched for the period 1966-1997 for English-language references. Selected articles were searched for additional references and additional sources of information included in textbooks on RCTs and their bibliographies.

#### **Selection criteria**

Works that include original research on stratification or include stratification as the major focus.

#### **Data collection and analysis**

Titles and abstracts were scanned to identify works that met the selection criteria.

#### **Main results**

The search identified 33 studies that included original research plus an unspecified number of secondary sources. A number of potential advantages of stratification are discussed. First, stratification provides greater assurance that compared groups are similar with respect to known prognostic factors. This is especially important for small trials ( $n < 100$ ). Further, stratification will protect against misleading inference, i.e. to larger extent achieve the chosen significance level and increase the power for a given sample size. Thus the efficiency, referred to as the number of patients that are required to detect a difference in two treatments at a prespecified level of significance and power, might be increased. Stratification is especially efficient in equivalence trials where large sample size reductions have been observed (12-42%). As potential disadvantages of stratification the administrative problems to run studies with large numbers of strata and the increased complexity for trials of therapies that require emergency administration are pointed out. Minimization and post-stratification are discussed as alternatives to stratified randomization.

The review ends with a set of 10 guidelines where the first three concern when to stratify. The main conclusion is to consider stratification for trials with less than 200 patients per study arm. The fourth concerns the calculation of the required sample size and guidelines 5-7 concern the choice of stratification factors. When factors with a large prognostic effect are not known or readily available, stratification becomes less important. Guidelines 8-10 concern alternatives and reporting.

#### **Authors' conclusions**

Stratification is a simple method of restricted randomization that is harmless always, useful frequently, and important rarely. It is important only for small trials in which treatment outcomes may be affected by known clinical factors that have large effect on prognosis, large trials when interim analyses are planned with small number of patients, and equivalence trials.



---

## A Practihc/ESCORT Commentary

Dave Sackett

A commentary addressing the methodology review:

Walter N. Kernan, Catherine M. Viscoli, Robert W. Makuch, Lawrence M. Brass and Ralph I. Horwitz. Stratified Randomization for Clinical Trials. *Journal of Clinical Epidemiology* 1999;52:19-26.

We rely on randomization to balance treatment groups for unknown factors that affect study participants' risks and responsiveness. But what should we do when these factors are already known? This review considered the strategy of guaranteeing balance for known prognostic factors by separating (stratifying) study participants into groups who do and don't have them, and then randomizing separately from each stratum. The review considered the effects, pros, and cons of stratification prior to randomisation and proposed guidelines for their application by trialists.

### **Are the results valid?**

The literature search was restricted to publications in English. There is no second, independent search against which to compare its thoroughness and appropriateness. The empiric data reviewed are certainly consistent with current logic and opinion.

### **What are the implications?**

Stratification prior to randomization generates experimental and control groups that are closely similar with respect to known factors affecting their risk and responsiveness. The benefits are twofold. First, by preventing the confounding of these factors with treatment, the validity of all-inclusive comparisons of outcomes is enhanced. Second, by creating roughly equal groups of experimental and control participants within each stratum, the trial usually becomes more "efficient" (that is, the confidence interval around the treatment effect is reduced).

On the other hand, stratification prior to randomization creates more work for trialists, and the number of strata for dichotomized known factors increases to the power of the number of factors selected for stratification (e.g., stratifying for sex, and the presence/absence of diabetes, heart failure, and hypertension generates 16 strata).

### **What are the implications for future methodology research?**

An unresolved implication is whether stratification prior to randomization ought to be replaced by minimization.<sup>1</sup>

How much the increasing development and application of methods for conducting trials on the internet will remove the "workload" of manual stratification and minimization is already a topic of considerable research interest.<sup>2</sup>

### **References**

1. Treasure T, MacRae KD. Minimization: the platinum standard for trials? *BMJ* 1998;317:362-363.
2. Litchfield J, Freeman J, Schou H, Elsley M, Fuller R, and Chubb B. Is the future for clinical trials internet-based? A cluster randomized clinical trial. *Clinical Trials* 2005;2:72-79.