



METHODOLOGIC GUIDELINES FOR SYSTEMATIC REVIEWS OF RANDOMIZED CONTROL TRIALS IN HEALTH CARE FROM THE POTSDAM CONSULTATION ON META-ANALYSIS

D. J. COOK,^{1*} D. L. SACKETT^{2†} and W. O. SPITZER^{3‡}

¹Departments of Medicine, Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada L8N 4A6, ²Nuffield Department of Medicine, University of Oxford, Oxford, U.K. and ³Potsdam Institute of Pharmacoepidemiology & Technology Assessment, Potsdam, Germany

INTRODUCTION

Twenty scientists from 9 countries spent 4 days in Potsdam, Germany, this past March (appropriately enough at the site of the conference of the victors in WW II) reassessing the state of the science of meta-analysis and the systematic review of clinical trials. The meeting was convened by the Potsdam Institute of Pharmacoepidemiology and Technology Assessment, under the auspices of McGill University and the Einstein Forum. Present were some of the proponents, theoreticians, and practitioners of the method as well as critics of the strategy (a list of participants appears at the end of this report).

Discussions on how to improve meta-analysis under the constraints imposed by the real world of imperfect studies that require critical appraisal and synthesis soon led to agreement on the need to identify, describe, and disseminate "good overview practice" and to the production of draft guidelines for designing, conducting, and reporting high-quality overviews of randomized controlled trials.

In view of the relative infancy of attempts to pool the results of observational studies, the work on guidelines for the meta-analysis of such

studies could only be started at Potsdam. On the other hand, guidelines for the meta-analysis of randomized trials were judged to be sufficiently well-developed and agreed upon to be submitted for publication and wide dissemination. They are written for those who conduct systematic reviews and for students of this emerging science. The process of consensus building and the formulation of guidelines for observational epidemiologic studies has proved more difficult and more time consuming than expected. The deliberations will continue in subsequent consultations, to prepare a separate set of Guidelines for the Conduct and Interpretation of Meta-Analyses of Epidemiological Studies (which will be published as a companion to this document within six months).

The data sources we used for this report included manuscripts from conference participants, discussions during the Potsdam Conference, references provided by conference participants and other colleagues who conduct systematic reviews, as well as our personal files. For purposes of this document, we have defined terms in the following fashion:

Systematic Review = Overview = the application of scientific strategies that limit bias to the systematic assembly, critical appraisal, and synthesis of all relevant studies on a specific topic.

Meta-analysis = Quantitative Overview = a systematic review that employs statistical methods to combine and summarize the results of several studies.

*All correspondence should be addressed to: Dr D. J. Cook, Department of Medicine, Division of Critical Care, St Joseph's Hospital, 50 Charlton Avenue East, Hamilton, Ontario, Canada L8N 4A6.

†Dr Cook is a Career Scientist of the Ontario Ministry of Health.

‡Dr Sackett is a Trillium Scientist of the Ontario Ministry of Health.

GUIDING PRINCIPLES

1. A systematic review must address a specific health care question. The question will determine which studies and data are relevant, and how they should be synthesized.
2. Methodology must serve biology and the users and providers of health care. Therefore, a team with expertise in both the content area and methodology is ideally suited to conduct valid, useful systematic reviews.
3. A systematic review requires collaboration with the investigators who conducted the primary studies.
4. Systematic reviews are retrospective research, and are potentially subject to many of the same biases the affect other retrospective studies; therefore, a good systematic review has to rely on both good randomized controlled trial methodology and good review methodology.
5. For several reasons, review methods may vary (e.g. scarce resources may limit search strategies). Thus, the review methods actually employed must be described in detail.
6. The existence of unsatisfactory randomized trials, case-control studies and cohort studies does not mean that any of these study designs should be abandoned; it means that they should be critically appraised, empirically studied and improved. Overviews of observational studies require a great deal of methodological development.

METHODOLOGIC GUIDELINES FOR SYSTEMATIC REVIEWS OF RANDOMIZED CONTROL TRIALS

1. Protocol development for the systematic review

- (a) pose the question in both biologic and health care terms, specifying the population, intervention and outcomes (both beneficial and harmful) of interest
- (b) specify the methods used to search for all potentially relevant data, to select relevant data, to assess its methodologic quality, and to analyze it
- (c) specify hypothesis-testing subgroup analyses *a priori*
- (d) report changes in protocol, and the rationale for these changes

2. Search strategy

- (a) the requirements for comprehensiveness of any search depends on the field and question that the systematic review is designed to answer
- (b) possible sources include:
 - computerized bibliographic databases of published and unpublished research
 - review articles
 - abstracts
 - conference/symposia proceedings
 - dissertations
 - books
 - expert informants
 - granting agencies
 - trial registries
 - industry
 - journal handsearching
- (c) specify language constraints

3. Study selection

- (a) select the studies that address the question posed by the systematic review; the selection should be based on *a priori* specification of the population, intervention, outcomes, and study design
- (b) assess the reproducibility of study selection
- (c) present the reasons for rejecting studies, especially those at the margins of relevance and scientific quality

4. Methodologic quality assessment

- (a) can be used:
 - as a threshold for inclusion to describe primary studies
 - as a possible explanation for heterogeneity
 - or in sensitivity analyses
- (b) base quality assessments on the extent to which bias is minimized (e.g. the method of allocation and its concealment, the prevention of contamination and cointervention, blinding, objective criteria for important outcomes, completeness of follow-up and handling of departures from the original protocol)
- (c) make quality assessment scoring systems transparent and parsimonious; avoid confusing methodologic quality with the quality of reporting
- (d) evaluate the reproducibility of methodologic quality assessment
- (e) report the methodologic quality scoring system used in the publication

5. Data extraction

- (a) be explicit, unbiased and reproducible
- (b) include all relevant measures of benefit and harm of the intervention
- (c) contact the investigators of the original studies for clarification, especially when there are ambiguities in published study methods (e.g. patient characteristics, details of interventions, definitions of events, losses to follow-up)
- (d) extract individual patient data when published data do not answer questions about:
 - intention to treat analyses
 - time-to-event analyses
 - subgroups
 - dose–response relationships
- (e) quality, participants, in the dose or duration of the intervention, or in the definitions and measurement of outcomes)
- (d) if heterogeneity exists across studies, whether or not it can be explained, consider using the random effects model (which takes into account between-study differences in treatment effects)
- (e) if heterogeneity can be explained using hypotheses specified *a priori*, consider presenting results by these subgroups
- (f) if heterogeneity cannot be explained, acknowledge it, and proceed with caution in any further statistical aggregation and subgroup analysis

6. Analysis

- (a) include all relevant and clinically useful measures of treatment effect (including relative measures, such as relative risk reduction, odds ratios, and effect size; and absolute measures such as absolute risk reduction and number needed to treat)
- (b) when data are too sparse, of too low quality, or too heterogeneous to proceed with their statistical aggregation, perform a narrative, qualitative summary and avoid meta-analysis
- (c) specify whether the “assumption free” (fixed effects) or the random effects model is used
- (d) describe the proportion of randomized patients included in the final analysis
- (e) employ confidence intervals
- (f) include a power analysis
- (g) consider a cumulative meta-analysis to reveal the contribution of successive trials (ordered by publication date, baseline risk or study quality)
- (a) pre-specify hypothesis-testing subgroup analyses and keep them few in number
- (b) label all *a posteriori* subgroup analyses
- (c) when subgroup differences are detected, interpret them in light of whether they are established *a priori*, are few in number, are supported by plausible causal mechanisms, are important (qualitative vs quantitative), are consistent across studies, and are statistically significant (adjusted for multiple significance testing)

7. Evaluation of heterogeneity

- (a) define what is meant by heterogeneity in each systematic review
- (b) in testing large data sets, trivial heterogeneity may be statistically significant; therefore, specify *a priori* the clinically important degree of heterogeneity
- (c) if heterogeneity exists (i.e. the test is statistically significant and the magnitude clinically important), examine potential sources of heterogeneity (e.g. differences in study
- (a) test the robustness of the results relative to features of the primary studies and to key assumptions and decisions
- (b) include tests for bias due to the retrospective nature of systematic reviews (e.g. with/without trials for which there is ambiguity about whether they meet inclusion criteria or methodologic standards, with/without unpublished studies, with/without studies of lower methodologic quality etc.)
- (c) consider the fragility of results by determining the effect of small shifts in the number of events between intervention and control groups
- (d) consider using cumulative meta-analysis to explore the relationship between effect size and study quality, control event rates, and other relevant features (e.g. time to treatment)
- (e) test a reasonable range of values for missing data from studies with uncertain results (which cannot be resolved by contacting the investigators)

8. Subgroup analyses

9. Sensitivity analyses

10. Presentation of results

- (a) include a structured abstract
- (b) include a table of key elements of each primary study (including design, sample size, characteristics of participants, treatment dose and duration, outcome measures, number of unreported/lost patients, collaboration with primary investigators, and any other information necessary for their interpretation)
- (c) include the summary data from which the efficacy measures are computed
- (d) report useful measures of efficacy such as the Number Needed to Treat to prevent one event (NNT), and the Number Needed to Treat to cause one harmful event due to treatment (NNH)
- (e) employ more informative graphic displays representing the confidence intervals, control group event rates, sample-sizes, etc.

11. Interpretation

- (a) interpret results within the context of current health care
- (b) state the methodologic limitations of both the primary studies and the systematic review
- (c) consider the size of the treatment effect in the primary studies as well as their aggregate, their consistency, and the presence of a dose-response relationship
- (d) consider interpreting results in the context of temporal cumulative meta-analyses
- (e) interpret the results in the context of other available evidence (human experimental or observational)
- (f) make any clinical recommendations practical and explicit
- (g) propose a future research agenda including clinical and methodologic requirements

12. Dissemination and updating

- (a) list known ongoing trials at the time of reporting
- (b) update the systematic review after important new trials are reported, and improve it after errors are identified
- (c) collaborate with journal editors in improving the quality of published systematic reviews

13. Conduct of trials in the future

- (a) ideally, investigators of potentially combinable trials will agree, before they begin the trials, to obtain the same "core" data on participants and interventions, and will agree on key outcomes (including quality of life) and their definitions
- (b) ideally, individual data from primary studies will be entered into a repository that protects confidentially while providing access by systematic reviewers

14. Conduct of systematic reviews in the future

- (a) more discussion and methodologic work is needed to determine in which situations the assumption free vs random effects models should be used
- (b) the effect of restricted language searches needs further study, preferably through international collaboration
- (c) systematic reviews should serve as background rationale for grant submissions
- (d) systematic reviews should help to set future research agendas
- (e) continuing development is necessary for more effective ways of applying the results of systematic reviews at the bedside
- (f) economists and meta-analysts should collaborate to determine how to incorporate economic evaluation into systematic reviews
- (g) the costs of conducting systematic reviews relative to their primary studies deserves further investigation
- (h) granting bodies should be encouraged to fund high quality systematic reviews

Potsdam Conference Participants

- C. Anello**, Food and Drug Administration, U.S.A.
- J. C. Bailar**, McGill University, Montreal, Canada
- V. Beral**, University of Oxford, U.K.
- I. Chalmers**, UK Cochrane Centre, Oxford, U.K.
- T. C. Chalmers**, Harvard University School of Public Health, Boston, U.S.A.
- D. J. Cook**, McMaster University, Hamilton, Canada
- A. R. Feinstein**, Yale University, New Haven, U.S.A.
- D. J. Finney**, Edinburgh, U.K.
- J. L. Fleiss**, Columbia University, New York, U.S.A.
- I. Guggenmooss-Holzmann**, Freie Universität, Berlin, Germany
- L. Heinemann**, Potsdam Institute, Potsdam, Germany
- A. Hildebrandt**, Institute for Drug Safety, Germany

- R. I. Horwitz**, Yale University, New Haven, U.S.A.
L. M. Irwig, University of Sydney, Sydney, Australia
H. Letzel, Staticon, Munich, Germany
A. Liberati, Mario Negri Institute, Milan, Italy
J. Le Lorier, University of Montreal, Montreal, Canada
I. Olkin, Stanford University, U.S.A.
R. Peto, University of Oxford, Oxford, U.K.
D. L. Sackett, University of Oxford, Oxford, U.K.
R. E. Slavin, Johns Hopkins University, Baltimore, U.S.A.
- W. O. Spitzer**, Potsdam Institute, Potsdam, Germany
C. Tingvall, Chalmers University, Sweden
J. Vandenbroucke, Leiden University, Leiden, Netherlands
N. Victor, Ruprecht Karls Universitat, Heidelberg, Germany.

Acknowledgements—The authors thank the Potsdam Consultation participants, and Drs Salim Yusuf and Andrew Oxman for their contributions to this document.