Health Technology Assessment 1998; Vol. 2: No. 19 (Executive summary)

Review

Executive summary

Systematic reviews of trials and other studies

AJ Sutton[1]
KR Abrams[1]
DR Jones[1]
TA Sheldon[2]
F Song[3]

[1] Department of Epidemiology and Public Health, University of Leicester, Leicester, UK
[2] York Health Policy Group, University of York, York, UK
[3] NHS Centre for Reviews and Dissemination, University of York, York, UK

Health Technology Assessment
NHS R&D HTA Programme

Executive summary

Objectives

Systematic review and meta-analytical methods are already common approaches to the assessment of health technology and related areas, and increasing adoption of such approaches may be foreseen, in part in response to increasingly wide emphasis on evidence-based approaches to medicine and health care. This report is intended:

- to identify applications of systematic review and meta-analytical methods in Health Technology Assessment (HTA)
- to promote further, appropriate use of such approaches in these areas of application
- to begin to identify priorities for further methodological developments in this field.

How the review was conducted

Systematic literature searches using MEDLINE, EMBASE, and Institute of Scientific Information (ISI) Science/Social Science electronic databases and the Cochrane methods database were carried out to find relevant articles. Relevant reference collections of the study team were pooled. Grey literature and unpublished articles were obtained by writing to prominent researchers, and through the Internet; further papers were identified by inspecting the reference lists of all previously obtained articles.

Review findings

A large number of papers concerning methodology relevant to different aspects of systematic reviews were identified. While the ordering of the report follows the stages involved in carrying out a systematic review, it is highly structured in a way which enables readers with specific interests to locate particularly relevant sections easily. The main features of the report are now summarised briefly in turn.

A brief overview of the important issues to be considered prior to the appraisal and synthesis of studies, including a critical appraisal of search methods, is presented.

Methodology for critical appraisal of the research evidence, including ways of assessing the quality of the primary studies, and its incorporation into a review, is explored. No consensus has been developed as to which method is most appropriate for doing this.

An important consideration is the possibility of heterogeneity between study outcome estimates. Many assessments and formal tests for detecting heterogeneity are described. Methods for accounting/adjusting for heterogeneity are identified and assessed. No consensus has been reached concerning the best strategy for dealing with heterogeneity; currently a large degree of subjectivity is required on the part of the reviewer.

Both classical and Bayesian statistical approaches have been developed to combine study estimates. These encompass the relatively simple fixed effect approaches, through random effects models, to more sophisticated hierarchical modelling. The more complex methods were largely devised to deal with heterogeneous outcomes, systematic variation between studies, and the need to incorporate a fuller set of components of variability into the model. Several of these methods have come under criticism; it is concluded that neither fixed nor random effect analyses can be considered ideal.

In addition to these general methods, approaches specific to particular outcome scales/measures, and data types are identified. These include methods for combining ordinal, binary, and continuous outcomes; survival data; diagnostic test data; correlated outcomes; individual patient data; single arm studies; crossover trials; and finally, studies of differing designs. While some of these methods have become standard, others are less commonly used and so are at early stages of development.

Problems encountered by meta-analysts were identified. Two potentially serious ones are publication bias and missing data. Methods for detecting/adjusting for publication bias exist, and others are currently being developed. The validity of most is largely undetermined. Additionally, long-term policy measures such as registries for all trials have been suggested. Dealing with missing data within a meta-analysis has not been considered to the same extent. General methods do exist (in other literatures), but many of them are untested in a meta-analytical setting.

Further issues identified include methods used to report the results of systematic reviews; use of sensitivity analyses; prospective meta-analysis; and alternatives to traditional meta-analysis.

Several of the key methods are illustrated using a dataset comprising cholesterol lowering studies.

Recommendations

Recommendations for good practice for the most part follow standard and widely agreed approaches. Greater latitude in the nature of studies potentially eligible for review, including non-randomised studies and the results of audit exercises, for example, may, however, be appropriate. The key stages are (with extensions and/or less widely agreed aspects in parentheses):

1. Specification in a protocol of the objectives, hypotheses (in both biological and health care terms), scope, and methods of the systematic review, before the study is undertaken.
2. Compilation of as comprehensive a set of reports as possible of relevant primary studies, having searched for all potentially relevant data, clearly documenting all search methods and sources.
3. Assessment of the methodological quality of the set of studies (the method being based on the extent to which susceptibility to bias is minimised, and the specific system used reported). Any selection of studies on quality or other criteria should be based on clearly stated *a priori* specifications. The reproducibility of the procedures in 2 and 3 should also be assessed.
4. Identification of a common set of definitions of outcome, explanatory and confounding variables, which are, as far as possible, compatible with those in each of the primary studies.
5. Extraction of estimates of outcome measures and of study and subject characteristics in a standardised way from primary study documentation, with due checks on extractor bias. Procedures should be explicit, unbiased and reproducible.
6. Perform, where warranted by the scope and characteristics of the data compiled, quantitative synthesis of primary study results (meta-analysis) using appropriate methods and models (clearly stated), in order to explore and allow for all important sources of variation (e.g. differences in study quality, participants, in the dose, duration, or nature of the intervention, or in the definitions and measurement of outcomes). This will often involve the use of mixed/hierarchical models, including fixed covariates to explain some elements of between-study variation, in combination with random effects terms.

7. Performance of a narrative or qualitative summary, where data are too sparse, or of too low quality, or too heterogeneous to proceed with a statistical aggregation (meta-analysis). In such cases the process of conduct and reporting should still be rigorous and explicit.
8. Exploration of the robustness of the results of the systematic review to the choices and assumptions made in all of the above stages. In particular, the following should be explained or explored:
   a) the impact of study quality/inclusion criteria
   b) the likelihood and possible impact of publication bias
   c) the implications of the effect of different model selection strategies, and exploration of a reasonable range of values for missing data from studies with uncertain results.
9. Clear presentation of key aspects of all of the above stages in the study report, in order to enable critical appraisal and replication of the systematic review. These should include a table of key elements of each primary study. Graphical displays can also assist interpretation, and should be included where appropriate. Confidence intervals around pooled point estimates should be reported.
10. Appraisal of methodological limitations of both the primary studies and the systematic review. Any clinical or policy recommendations should be practical and explicit, and make clear the research evidence on which they are based. Proposal of a future research agenda should include clinical and methodological requirements as appropriate.

Further areas of research related to the methods used for systematic reviews

Two priority areas are indicated below. Additionally, other areas needing further research are highlighted.

Priority topics

- Sensitivity analysis of the impact of many aspects of the design and analysis of the systematic review, and in particular of the meta-analysis, has been advocated. The result is a complex set of inter-related sensitivity analyses. Research into optimum, or at least efficient, strategies of multi-dimensional sensitivity analysis in these contexts would thus be useful.
- Evaluation of the role in HTA of meta-analysis of observational studies, and cross-design synthesis (which often features the inclusion of non-randomised evidence), possibly through systematic research and workshops of researchers active in the field.

Other areas needing further research

Study quality

- Investigation into the relevant dimensions of methodological quality and empirical research which establishes the relative importance of these dimensions in different contexts. This should eventually lead to the development of rigorous, validated, and parsimonious scales which can be adapted to a wide range of studies.
- Exploration of study quality as an explanation of heterogeneity.
- Empirical investigation into the basis for choice of cut-off values for exclusion of studies on grounds of quality.
- Systematic approaches to quality assessments of non-randomised controlled trials.

Heterogeneity

- Further investigation of its relationship with publication bias.
- Development of guidelines/recommendations for identifying and exploring heterogeneity.
- Investigation of degree of heterogeneity (both quantitative and qualitative) beyond which combining of all the studies should not be considered.
- Investigation into the effects of choice of measurement scale from both: a) a statistical perspective, and b) a clinical perspective.

Publication bias (HTA has commissioned a separate review in this area)

- Assessing the impact of the pipeline problem.
- Empirical study of degree and mechanisms of publication bias in meta-analysis of epidemiological and other non-randomised studies.
- Investigation into the extent to which the use of a prospective register for trials minimises publication bias.
- Further investigation into proposed statistical methods, including their power to detect publication bias, and their sensitivity towards its detection.

Approaches to modelling and analysis

- Investigation of the relative merits of the different approaches to combining studies in which some arms report no events (zeros in 2×2 tables)
- Comparison of new methods for random effects modelling which fully incorporate parameter uncertainty.
- Investigation of robustness of random effects models to departures from normality.
- Empirical investigation of model attributable weights with particular reference to over-weighting of large samples, in some models.
- Investigation of the impact of missing data at both the study level and patient level.
- Development of experience with practical applications of mixed models.
- Development of methodology for combining individual patient data with study level data.
- Investigation of the role of cumulative/sequential application of meta-analysis as a research methodology.
- Further development of methods for integration of qualitative assessments of studies with quantitative estimates of the results.
- Development of random/mixed effects models for meta-analysis of survival data.
- Use and implications of exact statistical methods for combining small studies.
- More extensive but critical use of Bayesian methods, including:
  a) encouragement of **expository papers** in the applied literature on the application of Bayesian methods
  b) more research on obtaining and using **elicited prior beliefs**.
- More research into the use of meta-analytic techniques in conjunction with decision analysis methods.
- General investigation of the impact of missing values, and extension of currently available methods to a wider range of circumstances with missing data, including the use of Bayesian methods.
- Development of the use of simulation of results of new studies before they are published or of hypothetical studies to allow their impact on meta-analysis to be assessed.

Miscellaneous

- More research into extrapolating the results of a meta-analysis to clinical practice.
- Further development of detailed publication guidelines to encourage uniform reporting of the results of studies, particularly of types other than randomised clinical trials.

Publication

Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. Systematic reviews of trials and other studies. *Health Technol Assess* 1998;**2**(19).

NHS R&D HTA Programme

The overall aim of the NHS R&D Health Technology Assessment (HTA) programme is to ensure that high-quality research information on the costs, effectiveness and broader impact of health technologies is produced in the most efficient way for those who use, manage and work in the NHS. Research is undertaken in those areas where the evidence will lead to the greatest benefits to patients, either through improved patient outcomes or the most efficient use of NHS resources.

The Standing Group on Health Technology advises on national priorities for health technology assessment. Six advisory panels assist the Standing Group in identifying and prioritising projects. These priorities are then considered by the HTA Commissioning Board supported by the National Coordinating Centre for HTA (NCCHTA).

This report is one of a series covering acute care, diagnostics and imaging, methodology, pharmaceuticals, population screening, and primary and community care. It was identified as a priority by the Methodology Panel and funded as project number 93/52/03.

The views expressed in this publication are those of the authors and not necessarily those of the Standing Group, the Commissioning Board, the Panel members or the Department of Health. The editors wish to emphasise that funding and publication of this research by the NHS should not be taken as implicit support for the recommendations for policy contained herein. In particular, policy options in the area of screening will be considered by the National Screening Committee. This Committee, chaired by the Chief Medical Officer, will take into account the views expressed here, further available evidence and other relevant considerations.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search, appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

The editors have tried to ensure the accuracy of this report but cannot accept responsibility for any errors or omissions. They would like to thank the referees for their constructive comments on the draft document.

Copies of this report can be obtained from:

The National Coordinating Centre for Health Technology Assessment,
Mailpoint 728, Boldrewood,
University of Southampton,
Southampton, SO16 7PX, UK.
Fax: +44 (0) 1703 595 639 Email: hta@soton.ac.uk
http://www.soton.ac.uk/~hta ISSN 1366-5278