

Evaluating non-randomised intervention studies

JJ Deeks^{1*}
J Dinnes²
R D'Amico¹
AJ Sowden³
C Sakarovitch¹
F Song⁴
M Petticrew⁵
DG Altman¹

In collaboration with the International Stroke Trial and the European Carotid Surgery Trial Collaborative Groups

¹ Centre for Statistics in Medicine, Institute of Health Sciences, Oxford, UK

² Southampton Health Technology Assessments Centre, University of Southampton, UK

³ NHS Centre for Reviews and Dissemination, University of York, UK

⁴ Department of Public Health and Epidemiology, University of Birmingham, UK

⁵ MRC Social and Public Health Sciences Unit, University of Glasgow, UK

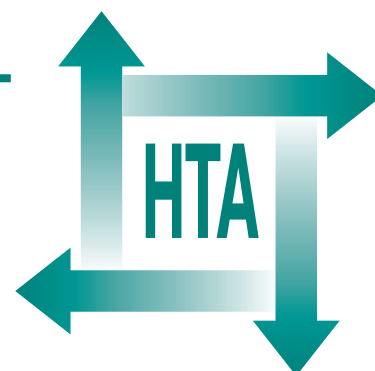
* Corresponding author



Executive summary

Health Technology Assessment 2003; Vol. 7: No. 27

Health Technology Assessment
NHS R&D HTA Programme





Executive summary

Background

In the absence of randomised controlled trials (RCTs), healthcare practitioners and policy-makers rely on non-randomised studies to provide evidence of the effectiveness of healthcare interventions. However, there is controversy over the validity of non-randomised evidence, related to the existence and magnitude of selection bias.

Objectives

To consider methods and related evidence for evaluating bias in non-randomised intervention studies.

Methods

1. Three reviews were conducted to consider:
 - empirical evidence of bias associated with non-randomised studies
 - the content of quality assessment tools for non-randomised studies
 - the use of quality assessment in systematic reviews of non-randomised studies.These reviews were conducted systematically, identifying relevant literature through comprehensive searches across electronic databases, handsearches and contact with experts.
2. New empirical investigations were conducted generating non-randomised studies from two large, multicentre RCTs by selectively resampling trial participants according to allocated treatment, centre and period. These were used to examine:
 - systematic bias introduced by the use of historical and non-randomised concurrent controls
 - whether results of non-randomised studies are more variable than results of RCTs
 - the ability of case-mix adjustment methods to correct for selection bias introduced by non-random allocation.

The resampling design overcame particular problems of meta-confounding and variability of direction and magnitude of bias that hinder the interpretation of previous reviews.

Results

Empirical comparisons of randomised and non-randomised evidence

Eight studies compared results of randomised and non-randomised studies across multiple interventions using meta-epidemiological techniques. The studies reached conflicting conclusions, explicable by differences in:

- whether data were sourced from primary studies or systematic reviews
- consideration of meta-confounding
- inclusion of studies of varying quality
- criterion for classifying discrepancies in results.

The only deducible conclusions were (a) results of randomised and non-randomised studies sometimes, but not always, differ and (b) both similarities and differences may often be explicable by other confounding factors.

Quality assessment tools for evaluating non-randomised studies

We identified 194 tools that could be or had been used to assess non-randomised studies. Around half were scales and half checklists, most were published within systematic reviews and most were poorly developed with scant attention paid to principles of scale development.

Sixty tools covered at least five of six pre-specified internal validity domains (creation of groups, blinding, soundness of information, follow-up, analysis of comparability, analysis of outcome), although the degree of coverage varied. Fourteen tools covered three of four core items of particular importance for non-randomised studies (How allocation occurred? Was the study designed to generate comparable groups? Were prognostic factors identified? Was case-mix adjustment used?). Six tools were thought suitable for use in systematic reviews.

Use of quality assessment in systematic reviews of non-randomised studies

Of 511 systematic reviews that included non-randomised studies, only 169 (33%) assessed study quality. Many used quality assessment tools designed for RCTs or developed by the authors

themselves, and did not include key quality criteria relevant to non-randomised studies. Sixty-nine reviews investigated the impact of quality on study results in a quantitative manner.

Empirical estimates of bias associated with non-random allocation

The bias introduced by non-random allocation was noted to have two components. First, the bias could lead to consistent over- or underestimations of treatment effects. This occurred for historical controls, the direction of bias depending on time trends in the case-mix of participants recruited to the study. Second, the bias increased variation in results for both historical and concurrent controls, owing to haphazard differences in case-mix between groups. The biases were large enough to lead studies falsely to conclude significant findings of benefit or harm.

Empirical evaluation of case-mix adjustment methods

Four strategies for case-mix adjustment were evaluated: none adequately adjusted for bias in historically and concurrently controlled studies. Logistic regression on average increased bias. Propensity score methods performed better, but were not satisfactory in most situations. Detailed investigation revealed that adequate adjustment can only be achieved in the unrealistic situation when selection depends on a single factor. Omission of important confounding factors can explain underadjustment. Correlated misclassifications and measurement error in confounding variables may explain the observed increase in bias with logistic regression, as may differences between conditional and unconditional odds ratio estimates of treatment effects.

Conclusions

Results of non-randomised studies sometimes, but not always, differ from results of randomised studies of the same intervention. Non-randomised studies may still give seriously misleading results when treated and control groups appear similar in key prognostic factors. Standard methods of case-mix adjustment do not guarantee removal of bias. Residual confounding may be high even when good prognostic data are available, and in some situations adjusted results may appear more biased than unadjusted results.

Although many quality assessment tools exist and have been used for appraising non-randomised studies, most omit key quality domains. Six tools were considered potentially suitable for use in systematic reviews, but each requires revision to cover all relevant quality domains.

Healthcare policies based upon non-randomised studies or systematic reviews of non-randomised studies may need re-evaluation if the uncertainty in the true evidence base was not fully appreciated when policies were made.

The inability of case-mix adjustment methods to compensate for selection bias and our inability to identify non-randomised studies which are free of selection bias indicate that non-randomised studies should only be undertaken when RCTs are infeasible or unethical.

Recommendations for further research

- The resampling methodology utilised here should be applied in other clinical areas to ascertain whether the biases we describe are typical.
- Efforts should be focused on developing quality assessment tools for non-randomised studies, possibly by refining existing tools.
- Research should consider how quality assessment of non-randomised studies can be incorporated into reviews, and the implications of individual quality features for interpretation of a review's results.
- Reasons for the apparent failure of case-mix adjustment methods should be further investigated, including assessments of the generalisability of our results to risk assessments and epidemiological studies, and differences between conditional and unconditional estimates of treatment effects.
- The role of the propensity score should be further evaluated, and computer macros made available for its application.

Publication

By Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, *et al.* Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;7(27).

NHS R&D HTA Programme

The NHS R&D Health Technology Assessment (HTA) Programme was set up in 1993 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 provide care in the NHS.

Initially, six HTA panels (pharmaceuticals, acute sector, primary and community care, diagnostics and imaging, population screening, methodology) helped to set the research priorities for the HTA Programme. However, during the past few years there have been a number of changes in and around NHS R&D, such as the establishment of the National Institute for Clinical Excellence (NICE) and the creation of three new research programmes: Service Delivery and Organisation (SDO); New and Emerging Applications of Technology (NEAT); and the Methodology Programme.

The research reported in this monograph was identified as a priority by the HTA Programme's Methodology Panel and was funded as project number 96/26/99.

The views expressed in this publication are those of the authors and not necessarily those of the Methodology Programme, HTA Programme 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 any recommendations made by the authors.

Criteria for inclusion in the HTA monograph series

Reports are published in the HTA monograph series if (1) they have resulted from work commissioned for the HTA Programme, and (2) they are of a sufficiently high scientific quality as assessed by the referees and editors.

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.

Methodology Programme Director: Professor Richard Lilford

HTA Programme Director: Professor Kent Woods

Series Editors: Professor Andrew Stevens, Dr Ken Stein, Professor John Gabbay,
Dr Ruairidh Milne and Dr Rob Riemsma

Managing Editors: Sally Bailey and Sarah Llewellyn Lloyd

The editors and publisher have tried to ensure the accuracy of this report but do not accept liability for damages or losses arising from material published in this report. They would like to thank the referees for their constructive comments on the draft document.

ISSN 1366-5278

© Queen's Printer and Controller of HMSO 2003

This monograph may be freely reproduced for the purposes of private research and study and may be included in professional journals provided that suitable acknowledgement is made and the reproduction is not associated with any form of advertising.

Applications for commercial reproduction should be addressed to HMSO, The Copyright Unit, St Clements House, 2-16 Colegate, Norwich, NR3 1BQ.

Published by Gray Publishing, Tunbridge Wells, Kent, on behalf of NCCHTA.

Printed on acid-free paper in the UK by St Edmundsbury Press Ltd, Bury St Edmunds, Suffolk.