Health Technology Assessment 2000; Vol. 4: No. 34

Methodology

Executive summary

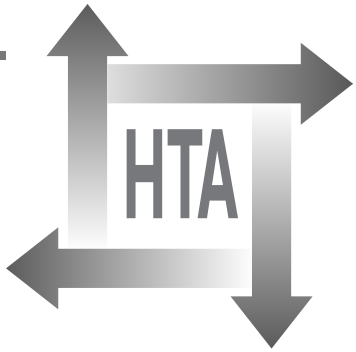
A systematic review of comparisons of effect sizes derived from randomised and non-randomised studies

RR MacLehose[1]
BC Reeves[2*]
IM Harvey[3]
TA Sheldon[4]
IT Russell[5]
AMS Black[6]

[1] South Essex Health Authority, UK
[2] Clinical Effectiveness Unit, Royal College of Surgeons and Health Services Research Unit, London School of Hygiene and Tropical Medicine, UK
[3] School of Health Policy and Practice, University of East Anglia, UK
[4] York Health Policy Group, University of York, UK
[5] Department of Health Sciences, University of York, UK
[6] Sir Humphry Davy Department of Anaesthesia, University of Bristol, Bristol Royal Infirmary, UK

[*] Corresponding author

Health Technology Assessment
NHS R&D HTA Programme



Executive summary

Background

There is controversy about the value of evidence about the effectiveness of healthcare interventions from non-randomised study designs. Advocates for quasi-experimental and observational (QEO) studies argue that evidence from randomised controlled trials (RCTs) is often difficult or impossible to obtain, or is inadequate to answer the question of interest. Advocates for RCTs point out that QEO studies are more susceptible to bias and refer to published comparisons that suggest QEO estimates tend to find a greater benefit than RCT estimates. However, comparisons from the literature are often cited selectively, may be unsystematic and may have failed to distinguish between different explanations for any discrepancies observed.

Objectives

The aim was to investigate the association between methodological quality and the magnitude of estimates of effectiveness by comparing systematically estimates of effectiveness derived from RCTs and QEO studies. Quantifying any such association should help healthcare decision-makers to judge the strength of evidence from non-randomised studies. Two strategies were used to minimise the influence of differences in external validity between RCTs and QEO studies:

- a comparison of the RCT and QEO study estimates of effectiveness of any intervention, where both estimates were reported in a single paper
- a comparison of the RCT and QEO study estimates of effectiveness for specified interventions, where the estimates were reported in different papers.

The authors also sought to identify study designs that have been proposed to address one or more of the problems often found with conventional RCTs.

Methods

Data sources

Relevant literature was identified from:

- the Cochrane Library, MEDLINE, EMBASE, DARE, and the Science Citation Index
- references of relevant papers already identified
- experts.

Electronic searches were very difficult to design and yielded few papers for the first strategy and when identifying study designs.

Choice of interventions to review for strategies 1 and 2

For strategy 1, any intervention was eligible. For strategy 2, interventions for which the population, intervention and outcome investigated were anticipated to be homogeneous across studies were selected for review:

- mammographic screening (MSBC) of women to reduce mortality from breast cancer
- folic acid supplementation (FAS) to prevent neural tube defects in women trying to conceive.

Data extraction and quality assessment

Data were extracted by the first author and checked by the second author. Disagreements were negotiated with reference to the paper concerned.

For strategy 1, study quality was scored using a checklist to assess whether the 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. For strategy 2, a more detailed instrument was used to assess study quality on four dimensions: the quality of reporting, the generalisability of the results, and the extent to which estimates of effectiveness may have been subject to bias or confounding. All quality assessments were carried out by three people.

Data synthesis and analysis

For strategy 1, pairs of comparisons between RCT and 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.

For strategy 2, three analyses were carried out:

- Attributes of the instrument were described by κ statistics, percentage agreement, and Cronbach's α values.
- Regression analyses were used to investigate variations in study quality.
- Meta-regression was used to investigate associations between study attributes and the size of estimates of effect for each intervention separately; the attributes considered included study design, study quality and sources of heterogeneity of the intervention and population between studies.

Results

Strategy 1

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.

Strategy 2

Thirty-four papers were identified, 17 evaluating MSBC and 17 FAS; eight and four papers, respectively, were individually or cluster assigned RCTs, five and six were non-randomised trials or cohort studies, and three and six were matched or unmatched case–control studies. Two studies, one of MSBC and one of FAS, used some other study design.

κ statistics for most items were < 0.4, although the percentage agreement usually exceeded 60%. Cronbach α values for different aspects of quality were < 0.5, suggesting that the instrument had limited ability to differentiate aspects of quality.

Regression analyses showed that both cohort and case–control studies had lower total quality scores than RCTs; cohort studies also had significantly lower scores than case–control studies. The latter, counter-intuitive finding may reflect a general tendency for quasi-experimental studies (which must use cohort designs) to have lower quality than observational studies.

Meta-regression of study attributes against relative risk estimates showed no association between effect size and study quality. Estimates from RCTs and cohort studies were not significantly different, but case–control studies gave significantly different estimates for both MSBC (greater benefit) and FAS (less benefit).

Identification of study designs

Ten study designs were identified; four, which include elements of both RCT and QEO study methods, were classified as hybrids and six, which adhere to the principle of randomisation but include some modification, were classified as RCT variants. Apart from the two-stage trial design, hybrid designs assume that non-randomised estimates are unbiased and that discrepancies between RCT and non-randomised estimates reflect the factors of interest (e.g. treatment preference). The majority of RCT variants have been designed to overcome the problems of non-compliance and patient drop-out; these designs therefore promote measures of efficacy as opposed to effectiveness. Three other types of variant were identified, namely response adaptive, randomised consent designs and change-to-open-label.

Conclusions

The findings of strategy 1 suggest that QEO study estimates of effectiveness may be valid if important confounding factors are controlled for. The small size of discrepancies for high-quality comparisons also implies that psychological factors (e.g. treatment preferences or willingness to be randomised) had a negligible effect on outcome. However, the authors caution against generalising their findings to other contexts, for three main reasons:

- Few papers were reviewed, and the findings may depend on the specific interventions evaluated.
- Most high-quality comparisons studied RCT and QEO study populations that met the same eligibility criteria, which may have reduced the importance of controlling for confounding.
- The literature reviewed is likely to have been subject to some form of publication bias. Authors of papers appeared to have strong *a priori* views about the usefulness of evidence from QEO studies, and the findings of papers appeared to support these views.

Strategy 2 found no association between study quality and effect size for either intervention, after taking account of study design. The lack of

association between quality and effect size could have arisen for a variety of reasons, the most likely being that study quality is not associated with relative risk in a predictable way or that the instrument failed to characterise methodological quality adequately.

There are several possible reasons for the finding that effect size estimates for case–control studies were significantly different from those for RCTs and cohort studies. The inconsistency of the direction of the discrepancy suggests that the direction is unpredictable and may be intervention specific. Case–control estimates of effectiveness should therefore be interpreted with extreme caution.

Several study designs were identified, which had been proposed to overcome a range of problems experienced with conventional RCTs, although the reported advantages were rarely substantiated. Discrepancies between RCT and QEO study estimates should not be attributed to factors such as patient preferences by default, since there may be residual confounding. Randomising patients prior to obtaining consent can cause as many problems as it solves, but may be useful when patients have a strong preference for an intervention. Other RCT variants may have a role when the aim is to measure efficacy.

The primary aim of quantifying any association between methodological quality and effect size was thwarted by several obstacles. For objective 1, the authors were unable to draw strong conclusions because of the paucity of evidence, and the potentially unrepresentative nature of the evidence they reviewed. For objective 2, the authors were unable adequately to distinguish, and measure, the variations in different aspects of quality between studies. The authors' recommendations relate directly to these obstacles.

Recommendations

- Most quasi-experiments reviewed were of poor quality. Quasi-experimental designs should not be rejected on the basis of this evidence.
- Standards for reporting of quasi-experimental and observational studies should be introduced. Enforcement of such standards, in the long term, might be expected to improve the standard of the research as well as reporting.
- More direct evidence about the comparability of findings from RCTs and QEO studies is needed. The comprehensive cohort study is probably the best study design for obtaining such evidence. Studies should be carried out in areas where RCTs are the preferred design, and in areas where RCTs are problematic, to assess the generalisability of evidence about the validity of QEO studies.
- There is a need to develop methods for identifying studies that provide a direct comparison of estimates from randomised and non-randomised data. A register should be established and studies entered into the register as they are identified. There is also a need for innovative search strategies to be developed.

Developing an instrument to characterise the quality of different studies is an urgent priority. The instrument must be able to assess all aspects of study design that may influence effect size. Separate instruments may be required for different study designs.

Publication

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

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.

Although the National Coordinating Centre for Health Technology Assessment (NCCHTA) commissions research on behalf of the Methodology Programme, it is the Methodology Group that now considers and advises the Methodology Programme Director on the best research projects to pursue.

The research reported in this monograph was funded as project number 93/45/02.

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.

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.

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) 23 8059 5639 Email: hta@soton.ac.uk
http://www.ncchta.org ISSN 1366-5278