

'Sometimes similar, sometimes different': a systematic review of meta-analyses of randomised and non-randomised policy intervention studies

Jonathan Shepherd¹, Anne-Marie Bagnall³, Jill Colquitt¹, Jacqueline Dinnes¹, Stephen Duffy², Amanda Sowden²

¹Southampton Health Technology Assessments Centre (SHTAC), Wessex Institute for Health Research & Development, University of Southampton, UK;

²Centre for Reviews and Dissemination, University of York, UK;

³School of Health and Community Care, Leeds Metropolitan University, UK

Introduction

While the randomised controlled trial (RCT) is generally regarded as the design of choice in health care studies, within public policy there is considerable debate about its suitability. As the first stage of a study to explore effect sizes from comparable randomised and non-randomised policy evaluations, we conducted a systematic review of meta-analyses.

Policy interventions may be defined as a set of actions with a coherent objective to bring about change or produce identifiable outcomes. These include policy, regulatory initiatives, single strategy projects or multi-component programmes. Policy interventions are intended to serve communities or populations, and are distinguished from one-to-one services that are for the benefit of individuals. Interventions may fall within public policy for health, education, social care, welfare, housing, criminal justice, transport and urban renewal.

Objectives

To:
(a) search for, assess and synthesise systematic reviews reporting randomised and non-randomised studies (NRSs) of policy interventions
(b) describe the methods used by reviewers to identify factors other than the use of randomisation that may have influenced the results of the studies

Methods

Systematic reviews meeting the following criteria were eligible for inclusion:

- Completed or published between 1999 and 2004
- Evaluated a policy intervention
- Included both randomised and non-randomised studies and estimated intervention effects separately according to design
- Used quantitative synthesis (meta-analysis).

Each review potentially meeting the inclusion criteria was screened by one reviewer using a predefined electronic form, and checked by a second reviewer. Disagreements were resolved by consensus, with reference to a third reviewer if necessary.

A comprehensive search strategy was designed, tested and applied to a number of electronic bibliographic databases.

A web-based data extraction form was designed and piloted. Reviews meeting the inclusion criteria underwent data extraction by one reviewer and were checked by a second.

Data extraction focused on:

- The effects of interventions by RCTs and NRSs
- Whether authors attempted to examine similarities or differences across study designs (e.g. study design, study population, outcomes)
- Whether authors tried to assess heterogeneity either across or within study designs, using statistical methods or other approaches to identifying heterogeneity.
- Which criteria were used to establish equivalence (or otherwise) of the results of RCTs and NRSs and whether these criteria were sensible and objective

Reviews were classified into three groups according to the authors' judgment regarding the equivalence or otherwise of the results of RCTs and NRSs ('similar', 'not similar' or 'mixed')

Details of these issues were recorded so that any observed differences in the results of RCTs and NRSs could be considered and were not simply attributed to lack of randomisation.

Table - Summary of review methods

		TOTAL n (%)	Results judged 'similar' n (%)	Results judged 'not similar' n (%)	Results mixed n (%)
Statistical heterogeneity identified by design?	Yes	4 (25%)	2 (40%)	2 (25%)	0
	No	11 (69%)	3 (60%)	5 (62.5%)	3 (100%)
	Narrative only	1 (6.25%)	0	1 (12.5%)	0
Heterogeneity investigated?	Yes	4 (25%)	1 (20%)	3 (37.5%)	0
	No	12 (75%)	4 (80%)	5 (62.5%)	3 (100%)
	Narrative only	0	0	0	0
Obvious differences between RCTs/NRSs (author or reviewers' opinion)	<i>Population</i>				
	Yes	2 (12.5%)	1 (20%)	1 (12.5%)	0
	No	14 (87.5%)	4 (80%)	7 (87.5%)	3 (100%)
	<i>Intervention</i>				
	Yes	4 (25%)	2 (40%)	2 (25%)	0
	No	12 (75%)	3 (60%)	6 (75%)	3 (100%)
<i>Comparator</i>	Yes	2 (12.5%)	2 (40%)	0	0
	No	14 (87.5%)	3 (60%)	8 (100%)	3 (100%)
	<i>Outcomes</i>				
Yes	1 (6.25%)	1 (20%)	0	0	
No	15 (93.75%)	4 (80%)	8 (100%)	3 (100%)	
Rationale for pooling approach given?	Yes	6 (37.5%)	2 (40%)	3 (37.5%)	1 (33.3%)
	No	7 (44%)	2 (40%)	4 (50%)	1 (33.3%)
	Partially?	3 (18.5%)	1 (20%)	1 (12.5%)	1 (33.3%)
Criteria to judge equivalence of study results by design given?	Yes	3 (18.5%)	0	2 (25%)	1 (33.3%)
	No	13 (81.5%)	5 (100%)	6 (75%)	2 (66.7%)

Results

16 reviews met the inclusion criteria

- 8 reviews focused on children with mainly school-based interventions
- 8 reviews targeted adults with mainly hospital- or workplace-based interventions.
- 5/16 reviews judged results of RCTs and NRSs 'similar'
- 8/16 reviews judged results of RCTs and NRSs 'different'
- In 5/8 reviews effect sizes were larger in NRS than RCTs
- In 2/8 reviews effect sizes were smaller in NRSs than RCTs
- In 1/8 reviews results of RCTs and NRSs were in opposite directions, although it is unclear which direction indicated a positive result.
- In many cases confidence intervals of the effect sizes overlapped

3/16 reviews judged results of RCTs and NRSs 'mixed' (i.e., similarity and differences between results of RCTs and NRSs varied across outcomes)

The table provides a summary of the meta-analyses

Statistical heterogeneity by study design was only identified in 4 reviews, and was investigated in only 4 reviews.

In most cases obvious differences between RCTs and NRS that might explain variations in effects were not identified. Where they were, differences in the intervention was the most cited factor, followed by the comparator and the study population.

Only 3 reviews reported the criteria used to judge the equivalence of the results of RCTs and NRSs. It is therefore not clear whether equivalence was judged in a systematic or pre-specified way, or, if so, how sensible and objective were the criteria used.

There were two approaches to meta-analysis:

- Keeping RCTs and NRSs separate throughout (8/16 reviews). The rationale for doing this was rarely explicit. 4 of these reviews stated *a priori* that they would investigate potential moderators of effect.
- Pooling all studies investigating potential moderators of effect, including randomisation, on the average estimate of effect (8/16 reviews). This approach usually involved a large number of studies that varied enormously in terms of intervention, population and outcomes measured

Results (cont)

Six of these reviews stated which potential moderator variables they would be investigating *a priori*. Methods used included sub-group analyses, sensitivity analyses, and meta-regression.

Conclusions

Considerable variation in the studies pooled within reviews, in terms of population, intervention, outcome and other methodological details, makes it difficult to separate the potential effect of random assignment from the potential effects of all the other variables. This reflects the broad nature of many systematic reviews of policy interventions compared to reviews of more tightly defined interventions.

The existing systematic reviews of policy interventions do not help us to determine whether RCTs and NRSs give similar results when evaluating policy interventions.

Systematic reviews should be carried out with the intention of investigating differences in effects of policy interventions between RCTs and NRSs. Sensible and objective criteria should be used to judge equivalence or otherwise of results of RCTs and NRSs

Acknowledgements

This analysis forms part of a larger project which was funded by the UK NHS Methodology programme which also involves meta-regression, and the re-analysis of primary data from trials. It is a collaborative project involving the Centre for Reviews and Dissemination at the University of York, the University of Southampton and the Institute of Education, University of London. See accompanying poster by Thomas *et al* at this conference for results of the meta-regression (Poster 291)

We would like to thank the following for their help:

- The project advisory group
- Sandy Oliver
- Zarnie Khadjesari

For further information, contact:
Jonathan Shepherd, Principal Research Fellow
Wessex Institute for Health Research and Development (WIHRD), Mail point 728, University of Southampton, UK
Tel: +44 (0)23 8059 7055, Fax: +44 (0)23 8059 5639
Email: jps@soton.ac.uk Web: www.wihrd.soton.ac.uk/