Health Technology Assessment 1999; Vol. 3: No. 5 (Executive summary)

Review

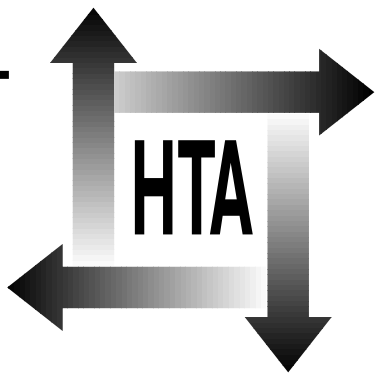
Executive summary

Methods for evaluating area-wide and organisation-based interventions in health and health care: a systematic review

OC Ukoumunne
MC Gulliford
S Chinn
JAC Sterne
PGJ Burney

Department of Public Health Sciences, Guy's, King's and St Thomas' School of Medicine, King's College London, UK

Health Technology Assessment
NHS R&D HTA Programme



Executive summary

Background

Health technology assessment often requires the evaluation of interventions which are implemented at the level of geographical area or health service organisational unit. Examples include health promotion interventions implemented in schools, workplaces or neighbourhoods, screening programmes in health authority populations, and healthcare interventions in general practices or hospitals. Interventions like these are implemented for clusters of individuals. Evaluation of cluster-based interventions presents a number of difficulties but some evidence suggests these are not always addressed in an optimal manner.

Aims and objectives

This report describes a systematic review of methods for evaluating cluster-based interventions. There were three objectives:

- to review the methodological literature and synthesise the findings into a checklist for practical use
- to evaluate existing practice in healthcare evaluation
- to present intraclass correlations for a range of outcome variables at different levels of organisational clustering in order to provide information for the design of future cluster-based studies.

Methods

- The review focused on methods for evaluating health and healthcare interventions that are implemented for clusters of patients or healthy individuals. References were obtained by handsearching journals, searching electronic databases, screening cited references, contacting expert informants, and searching the world wide web. Synthesis into a methodological checklist was by means of qualitative judgements concerning validity.
- A review of seven health science journals in 1996 yielded 56 papers reporting evaluations of cluster-based interventions. Evaluation against the checklist of methodological recommendations identified the main departures from good practice.
- A database of intraclass correlations was compiled by analysing data from a variety of sources.

Methodological recommendations

The main methodological findings of the review were synthesised into a 12-point checklist for investigators.

(1) **Recognise the cluster as the unit of intervention or allocation.** It is important to distinguish between cluster level and individual level intervention, as failure to do so can result in studies which are inappropriately designed or which give incorrect results.

(2) **Justify the use of the cluster as the unit of intervention or allocation.** For a fixed number of individuals, studies in which clusters are allocated are not as powerful as traditional clinical trials in which individuals are randomised. The decision to allocate at cluster level should be justified on theoretical, practical or economic grounds.

(3) **Include a sufficient number of clusters.** Evaluation of an intervention implemented in a single cluster will not usually give generalisable results. Valid designs should include a control group not receiving the intervention. Both intervention and control groups should include enough clusters to allow the effect of intervention to be distinguished from natural variability among clusters. Studies with fewer than four clusters per group are unlikely to yield statistically significant results, and more clusters will be required if relevant intervention effects are small.

(4) **Randomise clusters wherever possible.** The need for randomisation is generally accepted in the evaluation of individual level interventions but randomisation of clusters has not been practised as often as it should be in the evaluation of cluster-based interventions. Because of the risk of bias, use of quasi-experimental or observational designs should always be justified.

(5) **In non-randomised studies include a control group.** When randomisation is not feasible, a control group should be included. Each group should include a sufficient number of clusters (see point 3). The clusters allocated to groups should be stratified for important prognostic factors so far as possible (see point 8) and a wide range of confounders should be measured. Outcome variables should be measured before and after the intervention.

(6) In single group studies include repeated measurements over time. Sometimes it is not feasible to include a control group, as, for example, when a new policy is implemented at national level. In this case, repeated assessments should be made both before and after the intervention in order to control for secular changes in the outcome.

(7) Allow for clustering when estimating the required sample size. The total number of individuals required can be estimated by multiplying the result of a standard sample size calculation by the design effect. This will require an estimate of the intraclass correlation coefficient, which should be obtained from previous studies.

(8) Consider the use of pairing or stratification of clusters where appropriate. Cluster-based evaluations often include small numbers of clusters, and simple randomisation is unlikely to yield groups that are balanced with respect to cluster level baseline characteristics. Stratification or pairing of clusters according to characteristics that are associated with the outcome may reduce error in randomised studies and reduce bias in non-randomised studies. Limitations of the paired, or matched, design are underappreciated.

(9) Consider different approaches to repeated assessments in prospective evaluations. Either cohort or repeated cross-sectional designs may be used to sample individuals in studies with follow-up. The cohort design is more applicable to individual level outcomes, and may yield more precise results but is more susceptible to bias. The repeated cross-sectional design is more appropriate when outcomes will be aggregated to cluster level; it is usually less powerful but is less susceptible to bias.

(10) Allow for clustering at the time of analysis. Standard statistical methods applied to individual level outcomes should not be used because they will give confidence intervals that are too narrow and p values that are too small. There are three valid approaches to analysis: cluster level analysis, in which the cluster means or proportions are used as units of analysis; adjusted individual level analysis, in which standard univariate statistical methods are adjusted for the design effect; regression methods for clustered data, which allow for both individual and cluster level variation (hierarchical analysis). When the number of clusters is small, cluster level analysis will be most appropriate because between-cluster variation cannot be estimated with sufficient precision to implement analyses at the individual level. Regression methods for clustered data will usually be required for non-randomised designs.

(11) Allow for confounding at both individual and cluster level. Standard multiple regression methods are not appropriate. Use of regression methods for clustered data will allow the incorporation of both individual and cluster level confounders in the analysis. This approach will increase precision in randomised studies and reduce bias in non-randomised designs.

(12) Include estimates of intraclass correlation and components of variance in published reports. In order to provide information that may be used to estimate sample size requirements for future studies, estimates of the intraclass correlation coefficient should be included in published reports.

Case study: a review of seven health science journals

A review of 56 papers reporting evaluations of cluster-based interventions from seven health science journals showed that the present level of adherence to the methodological recommendations of the review was low. The main departures from recommendations were the evaluation of interventions in small numbers of clusters, and the incorrect use of standard methods for individual level analysis.

A database of intraclass correlation coefficients

In order to provide information which may be used in the design of future studies, the report presents intraclass correlation coefficients and components of variance for a range of outcomes in five areas: cardiovascular and lifestyle, cancer, respiratory, health service activity, and other. For community-based studies, data are presented for individuals clustered at the level of household, postcode sector and district and regional health authority. For healthcare-based studies, data are presented for clustering at the level of general practice, hospital, district health authority and family health services authority.

Publication

Ukoumunne OC, Gulliford MC, Chinn S, Sterne JAC, Burney PGJ. Methods for evaluating area-wide and organisation-based interventions in health and health care: a systematic review. *Health Technol Assess* 1999;**3**(5).

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 94/09/01.

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