Health Technology Assessment 2007; Vol. 11: No. 43

Contamination in trials of educational interventions

MR Keogh-Brown, MO Bachmann,
L Shepstone, C Hewitt, A Howe, CR Ramsay,
F Song, JNV Miles, DJ Torgerson, S Miles,
D Elbourne, I Harvey and MJ Campbell

October 2007

Health Technology Assessment
NHS R&D HTA Programme
www.hta.ac.uk

How to obtain copies of this and other HTA Programme reports.
An electronic version of this publication, in Adobe Acrobat format, is available for downloading free of charge for personal use from the HTA website (http://www.hta.ac.uk). A fully searchable CD-ROM is also available (see below).

Printed copies of HTA monographs cost £20 each (post and packing free in the UK) to both public **and** private sector purchasers from our Despatch Agents.

Non-UK purchasers will have to pay a small fee for post and packing. For European countries the cost is £2 per monograph and for the rest of the world £3 per monograph.

You can order HTA monographs from our Despatch Agents:

– fax (with **credit card** or **official purchase order**)
– post (with **credit card** or **official purchase order** or **cheque**)
– phone during office hours (**credit card** only).

Additionally the HTA website allows you **either** to pay securely by credit card **or** to print out your order and then post or fax it.

Contact details are as follows:

HTA Despatch
c/o Direct Mail Works Ltd
4 Oakwood Business Centre
Downley, HAVANT PO9 2NP, UK

Email: orders@hta.ac.uk
Tel: 02392 492 000
Fax: 02392 478 555
Fax from outside the UK: +44 2392 478 555

NHS libraries can subscribe free of charge. Public libraries can subscribe at a very reduced cost of £100 for each volume (normally comprising 30–40 titles). The commercial subscription rate is £300 per volume. Please see our website for details. Subscriptions can only be purchased for the current or forthcoming volume.

Payment methods

Paying by cheque
If you pay by cheque, the cheque must be in **pounds sterling**, made payable to *Direct Mail Works Ltd* and drawn on a bank with a UK address.

Paying by credit card
The following cards are accepted by phone, fax, post or via the website ordering pages: Delta, Eurocard, Mastercard, Solo, Switch and Visa. We advise against sending credit card details in a plain email.

Paying by official purchase order
You can post or fax these, but they must be from public bodies (i.e. NHS or universities) within the UK. We cannot at present accept purchase orders from commercial companies or from outside the UK.

How do I get a copy of *HTA on CD*?

Please use the form on the HTA website (www.hta.ac.uk/htacd.htm). Or contact Direct Mail Works (see contact details above) by email, post, fax or phone. *HTA on CD* is currently free of charge worldwide.

The website also provides information about the HTA Programme and lists the membership of the various committees.

Contamination in trials of educational interventions

MR Keogh-Brown,[1] MO Bachmann,[1*] L Shepstone,[1] C Hewitt,[2] A Howe,[1] CR Ramsay,[3] F Song,[1] JNV Miles,[4] DJ Torgerson,[2] S Miles,[1] D Elbourne,[5] I Harvey[1] and MJ Campbell[6]

[1] School of Medicine, Health Policy and Practice, University of East Anglia, Norwich, UK
[2] Department of Health Sciences, University of York, UK
[3] Health Services Research Unit, University of Aberdeen, UK
[4] RAND Corporation, Santa Monica, California, USA
[5] London School of Hygiene and Tropical Medicine, UK
[6] Medical Statistics Group, School of Health and Related Research (ScHARR), The University of Sheffield, UK

* Corresponding author

Declared competing interests of authors: none

Published October 2007

This report should be referenced as follows:

Keogh-Brown MR, Bachmann MO, Shepstone L, Hewitt C, Howe A, Ramsay CR, *et al*. Contamination in trials of educational interventions. *Health Technol Assess* 2007;**11**(43).

Health Technology Assessment is indexed and abstracted in *Index Medicus*/MEDLINE, *Excerpta Medica*/EMBASE and *Science Citation Index Expanded (SciSearch®)* and *Current Contents®*/Clinical Medicine.

NIHR Health Technology Assessment Programme

The Health Technology Assessment (HTA) programme, now part of the National Institute for Health Research (NIHR), was set up in 1993. It produces high-quality research information on the costs, effectiveness and broader impact of health technologies for those who use, manage and provide care in the NHS. 'Health technologies' are broadly defined to include all interventions used to promote health, prevent and treat disease, and improve rehabilitation and long-term care, rather than settings of care.

The research findings from the HTA Programme directly influence decision-making bodies such as the National Institute for Health and Clinical Excellence (NICE) and the National Screening Committee (NSC). HTA findings also help to improve the quality of clinical practice in the NHS indirectly in that they form a key component of the 'National Knowledge Service'.

The HTA Programme is needs-led in that it fills gaps in the evidence needed by the NHS. There are three routes to the start of projects.

First is the commissioned route. Suggestions for research are actively sought from people working in the NHS, the public and consumer groups and professional bodies such as royal colleges and NHS trusts. These suggestions are carefully prioritised by panels of independent experts (including NHS service users). The HTA Programme then commissions the research by competitive tender.

Secondly, the HTA Programme provides grants for clinical trials for researchers who identify research questions. These are assessed for importance to patients and the NHS, and scientific rigour.

Thirdly, through its Technology Assessment Report (TAR) call-off contract, the HTA Programme commissions bespoke reports, principally for NICE, but also for other policy-makers. TARs bring together evidence on the value of specific technologies.

Some HTA research projects, including TARs, may take only months, others need several years. They can cost from as little as £40,000 to over £1 million, and may involve synthesising existing evidence, undertaking a trial, or other research collecting new data to answer a research problem.

The final reports from HTA projects are peer-reviewed by a number of independent expert referees before publication in the widely read monograph series *Health Technology Assessment*.

Criteria for inclusion in the HTA monograph series

Reports are published in the HTA monograph series if (1) they have resulted from work 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 research reported in this monograph was commissioned by the National Coordinating Centre for Research Methodology (NCCRM), and was formerly transferred to the HTA Programme in April 2007 under the newly established NIHR Methodology Panel. The HTA Programme project number is 06/90/20. The contractual start date was in May 2004. The draft report began editorial review in March 2007 and was accepted for publication in April 2007. The commissioning brief was devised by the NCCRM who specified the research question and study design. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' report and would like to thank the referees for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this report.

The views expressed in this publication are those of the authors and not necessarily those of the HTA Programme or the Department of Health.

ISSN 1366-5278

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.

MR

Abstract

Contamination in trials of educational interventions

MR Keogh-Brown,[1] MO Bachmann,[1*] L Shepstone,[1] C Hewitt,[2] A Howe,[1] CR Ramsay,[3] F Song,[1] JNV Miles,[4] DJ Torgerson,[2] S Miles,[1] D Elbourne,[5] I Harvey[1] and MJ Campbell[6]

[1] School of Medicine, Health Policy and Practice, University of East Anglia, Norwich, UK
[2] Department of Health Sciences, University of York, UK
[3] Health Services Research Unit, University of Aberdeen, UK
[4] RAND Corporation, Santa Monica, California, USA
[5] London School of Hygiene and Tropical Medicine, UK
[6] Medical Statistics Group, School of Health and Related Research (ScHARR), The University of Sheffield, UK
* Corresponding author

Objectives: To consider the effects of contamination on the magnitude and statistical significance (or precision) of the estimated effect of an educational intervention, to investigate the mechanisms of contamination, and to consider how contamination can be avoided.
Data sources: Major electronic databases were searched up to May 2005.
Methods: An exploratory literature search was conducted. The results of trials included in previous relevant systematic reviews were then analysed to see whether studies that avoided contamination resulted in larger effect estimates than those that did not. Experts' opinions were elicited about factors more or less likely to lead to contamination. We simulated contamination processes to compare contamination biases between cluster and individually randomised trials. Statistical adjustment was made for contamination using Complier Average Causal Effect analytic methods, using published and simulated data. The bias and power of cluster and individually randomised trials were compared, as were Complier Average Causal Effect, intention-to-treat and per protocol methods of analysis.
Results: Few relevant studies quantified contamination. Experts largely agreed on where contamination was more or less likely. Simulation of contamination processes showed that, with various combinations of timing, intensity and baseline dependence of contamination, cluster randomised trials might produce biases greater than or similar to those of individually randomised trials. Complier Average Causal Effect analyses produced results that were less biased than intention-to-treat or per protocol analyses. They also showed that individually randomised trials would in most situations be more powerful than cluster randomised trials despite contamination.
Conclusions: The probability, nature and process of contamination should be considered when designing and analysing controlled trials of educational interventions in health. Cluster randomisation may or may not be appropriate and should not be uncritically assumed always to be a solution. Complier Average Causal Effect models are an appropriate way to adjust for contamination if it can be measured. When conducting such trials in future, it is a priority to report the extent, nature and effects of contamination.

Contents

List of abbreviations

ASME	Association for the Study of Medical Education
CACE	Complier Average Causal Effect
CI	confidence interval
CRC	colorectal cancer
DA	data argumentation
EM	expectation maximisation
IBS	irritable bowel syndrome
ICC	intracluster correlation coefficient
ITT	intention-to-treat
IV	instrumental variable
LATE	Local Average Treatment Effect
ML	maximum likelihood
ML–EM	maximum likelihood methods using the expectation maximisation algorithm
RCM	Rubin's Causal Model
RCT	randomised controlled trial
SD	standard deviation
2SLS	two-stage least squares

All abbreviations that have been used in this report are listed here unless the abbreviation is well known (e.g. NHS), or it has been used only once, or it is a non-standard abbreviation used only in figures/tables/appendices in which case the abbreviation is defined in the figure legend or at the end of the table.

Executive summary

Objectives

The objectives of the study were to consider the effects of contamination on the magnitude and statistical significance (or precision) of the estimated effect of an educational intervention, to investigate the mechanisms of contamination, and to consider how contamination can be avoided.

Background

Educational interventions aimed at improving health, knowledge or health-related behaviour may be delivered to patients, health professionals or members of the general public. Contamination in controlled trials occurs when people who were not intended to receive an intervention inadvertently do so. Trials of educational interventions are especially prone to contamination because the active ingredients can be transportable and difficult to confine. Contamination tends to reduce the magnitude of effect estimates and therefore also to increase the chance that estimates will not be statistically significant. That is, contamination causes bias and reduces power.

Contamination can be avoided during the design, conduct or analysis of trials, but such strategies may be ineffective or may be in conflict with each other. With cluster randomised trials, groups of people are allocated to receive or not to receive an intervention, or to receive different interventions. This reduces contamination bias if it effectively separates people and so reduces the risk or extent of contamination. However, if, within each group, individuals are very similar to each other, compared with individuals in other groups, then the statistical significance and precision of effect estimates are reduced. Various methods of data analysis may adjust for this bias if the extent of contamination is known. However, these adjustments may reduce power and precision because they exclude part of the study sample, or they may cause bias by comparing dissimilar subgroups of the sample. Contamination is often assumed to be a problem when interpreting or designing trials of educational interventions, but whether it really is a problem is not well known. It has been argued that the problems of contamination have been exaggerated, and therefore cluster randomised trials are often inappropriate, given their statistical disadvantages. However, cluster randomisation may be appropriate if interventions are aimed at professionals or facilities that manage groups of people, regardless of contamination.

Methods

An exploratory literature search was conducted with major electronic databases being searched up to May 2005. The results of trials included in previous relevant systematic reviews were then analysed to see whether studies that avoided contamination resulted in larger effect estimates than those that did not. Experts' opinions were elicited about factors more or less likely to lead to contamination. We simulated contamination processes to compare contamination biases between cluster and individually randomised trials. Statistical adjustment was made for contamination using Complier Average Causal Effect analytic methods, using published and simulated data. The bias and power of cluster and individually randomised trials were compared, as were Complier Average Causal Effect, intention-to-treat and per protocol methods of analysis.

Results

Literature search

Although many studies have reported using cluster randomisation to avoid contamination, few have quantified contamination and its effects. We compared the results of cluster and individually randomised trials from previous systematic reviews of educational interventions aimed at health professionals and at patients. We examined whether individually randomised trials tended to show smaller effects, which could indirectly indicate contamination bias. This was true for the relatively few trials that evaluated very similar interventions, although there may be explanations other than contamination or its avoidance. It was

not true for the larger number of heterogeneous trials. One interesting trial randomised patients either to a cluster or an individually randomised sub-trial, both of which evaluated the same oral health intervention. Its results suggested that cluster randomisation reduced contamination bias, but some partial contamination also occurred in the cluster randomised trial. The results could also be explained by unblinded outcome measurement.

Consensus of expert opinion

Thirty-seven experts in trials of educational interventions took part in a Delphi study. They answered a questionnaire ranking the likelihood that contamination would occur in various situations, assuming for each situation that all other factors were constant; 27 completed the second-round questionnaire after feedback on all responses to the first round. In the experts' opinion, contamination was more likely in trials conducted in settings where subjects worked, lived or interacted closely together and where interventions were desirable, simple or easily transferable or were aimed at increasing knowledge. It was less likely when subjects were socially or physically separate, and where interventions were complex or aimed at changing behaviours. It was more likely with interventions aimed at health professionals than with interventions aimed at patients. It was more likely with interventions based on broadcast media, audiovisuals or written information and was least likely with computer-based reminders. Cluster randomisation was the design most likely to avoid contamination and individual randomisation was least likely to do so.

Simulating contamination

A computer model simulated the process of contamination to compare bias between cluster and individually randomised trials. When contamination is not cluster-wide but filters slowly amongst individuals, a cluster randomised design produces less bias. Individual randomisation produces less bias if entire clusters are contaminated at once, unless the risk of clusters being contaminated is low. Different combinations of the components of the contamination process favour either cluster or individual randomisation and should be considered when designing trials. Empirical evidence of the process of contamination during educational trials would be valuable.

Dealing with contamination using Complier Average Casual Effect analysis

Intention-to-treat analysis only answers the question of whether the offer of treatment to the intervention population is effective. Per protocol or on-treatment analyses, which attempt to account for contamination or non-adherence, are likely to be biased because of systematic differences between exposed and unexposed control subjects. Complier Average Causal Effect (CACE) analysis potentially overcomes these problems. It can be implemented in various ways. In this work, we used an instrumental variable technique entailing a two-stage regression. In the first stage, a dummy variable representing the treatment that the participants actually receive is regressed on a dummy variable representing the treatment to which the participants were randomised. Then in the second stage, the outcome is regressed on the treatment received variable and the residuals saved from the first stage. CACE tends to produce an unbiased estimate of the true treatment effect but also tends to reduce statistical power. To assess whether or not the power lost due to CACE analysis was better or worse than that due to cluster randomisation, we undertook a simulation exercise. With up to 30% contamination and using a CACE approach, individual randomisation was more powerful than cluster randomisation. This was true even when assuming small cluster sizes and intracluster correlation coefficients. Although analysis by intention-to-treat is generally the most valid primary analytic method for randomised trials, these methods may be appropriate for secondary analysis of randomised trials, or for analysis of non-randomised trials, if contamination has been measured.

Conclusions

The literature search found little evidence that contamination really is a problem in trials of educational interventions in health because very few studies reported whether contamination occurred. However, there is consensus about the types of situation in which contamination is more or less likely. If it is likely then cluster randomisation may reduce contamination unless entire clusters are contaminated. CACE analysis may reduce bias if contamination is measured. In future trials of educational interventions in health, it is a priority to report the extent, nature and effects of contamination.

Chapter 1
Introduction

The purpose of a controlled trial is to obtain a valid and precise estimate of the effect of one or more interventions. Controlled trials typically compare a group of individuals who have been exposed to an intervention with a group who have not been exposed, or have been exposed to another intervention. Ideally, randomisation is used to allocate individuals to the groups to be compared. This means that the groups can be considered to be identical, on average, at the start of the trial. Therefore, any difference in outcome between the groups can be attributed to the intervention. If randomisation is not possible, then baseline differences need to be adjusted for statistically but systematic differences cannot be entirely ruled out.

There are two problems that may arise in controlled trials:

1. The intervention is inadvertently received by members of the non-intervention or control group. This is called **contamination**.
2. The intervention is not received by members of the intervention group. This is called **non-adherence** (or non-adherence or non-concordance).

These two problems are similar to each other. Both tend to blur the distinction between the two groups. However, they are different processes. Whereas non-adherence is a potential problem in most kinds of trials, contamination is thought to be a particular problem in trials of educational interventions because of their transferability. That is, the active ingredients may be difficult to confine to the subjects for whom they are intended. For example, a patient who receives information about risks of obesity can pass that information on to others.

Effects of contamination on the magnitude and statistical significance (or precision) of the estimated effect

Contamination will only affect the results of a trial if the intervention is effective. Contamination of members of the non-intervention group will tend to shift the outcome in the control group in the same direction as the effect in the intervention group. Hence the effectiveness of the intervention, estimated as the ratio or difference between intervention and control groups, will tend to be underestimated. Thus contamination biases trial results towards the null – that is, towards a difference of zero or a ratio of one. For example, if half of the control group was contaminated, we might expect the control group overall to experience half of the effect experienced by the intervention group.

Another adverse effect of contamination is the reduction in statistical power to detect a significant effect, because contamination reduces the estimated effect of an intervention. This is also seen as wider confidence intervals (CIs) around the effect estimate, that is, as decreased precision. This may lead to the conclusion that the intervention does not have a significant effect, when the true effect would have been significantly different from zero. This means failing to reject the hypothesis that there is no significant difference between groups, when that hypothesis is false (Type II error). Similarly, in an equivalence trial contamination reduces one's confidence that different interventions have equivalent effects. *Figure 1* shows how, with larger cluster sizes, contamination can severely reduce the power of trials to show significant effects. *Figure 2* shows how contamination leads to requirements for larger sample sizes, and more so with greater intracluster correlation. These figures were calculated with equations from Donner and Klar.[1]

Mechanisms of contamination with different types of educational intervention

The way in which contamination occurs may differ according to the educational mechanisms and the targets of the intervention. The target of the intervention might be a patient, a healthcare professional or members of the general public. Mechanisms, or processes, of contamination in such trials are discussed briefly here. A wider range of contamination processes are listed in Chapter 3.

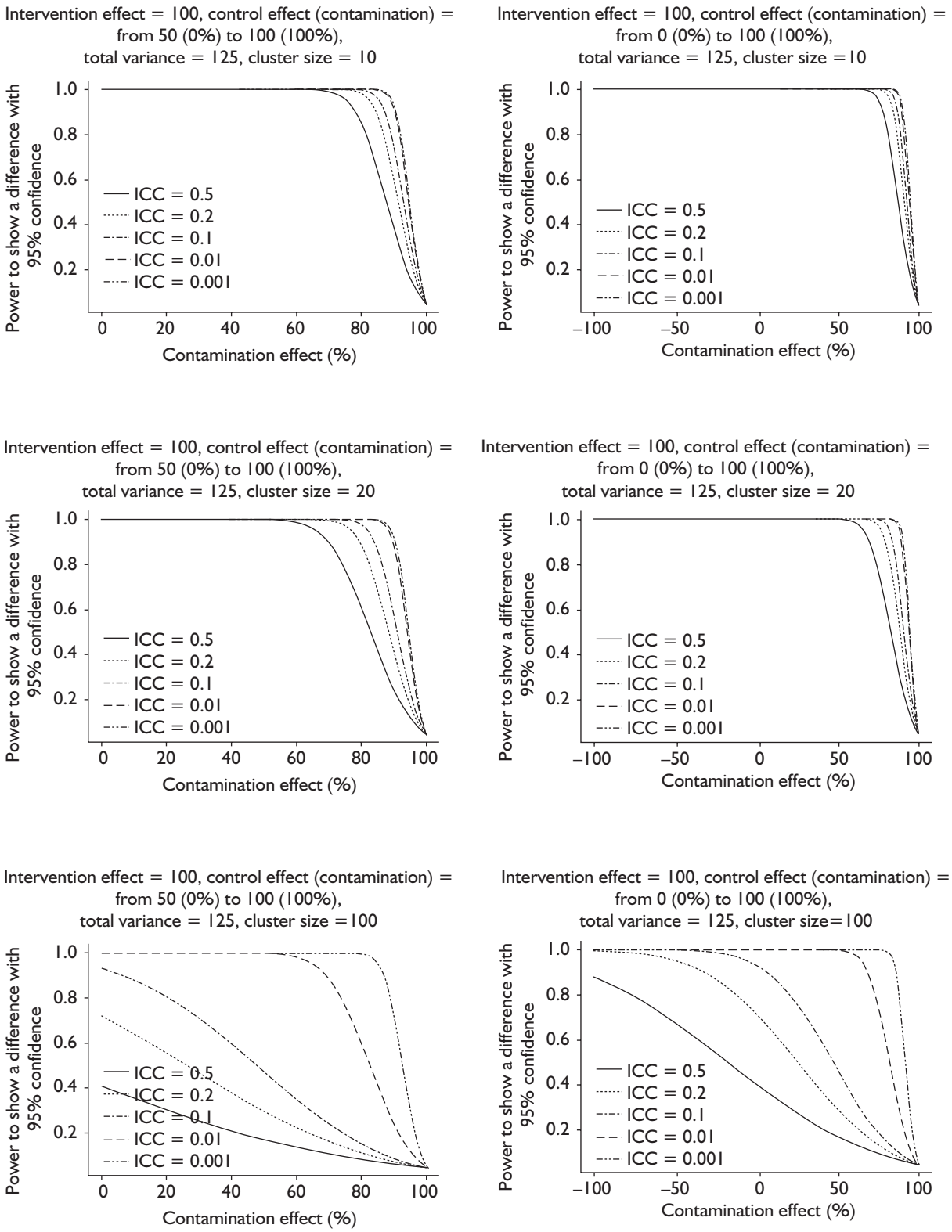

FIGURE 1 *Power to detect differences between intervention and control arms in cluster randomised trials, with different cluster sizes and degrees of contamination (5% significance level). ICC, intracluster correlation coefficient.*

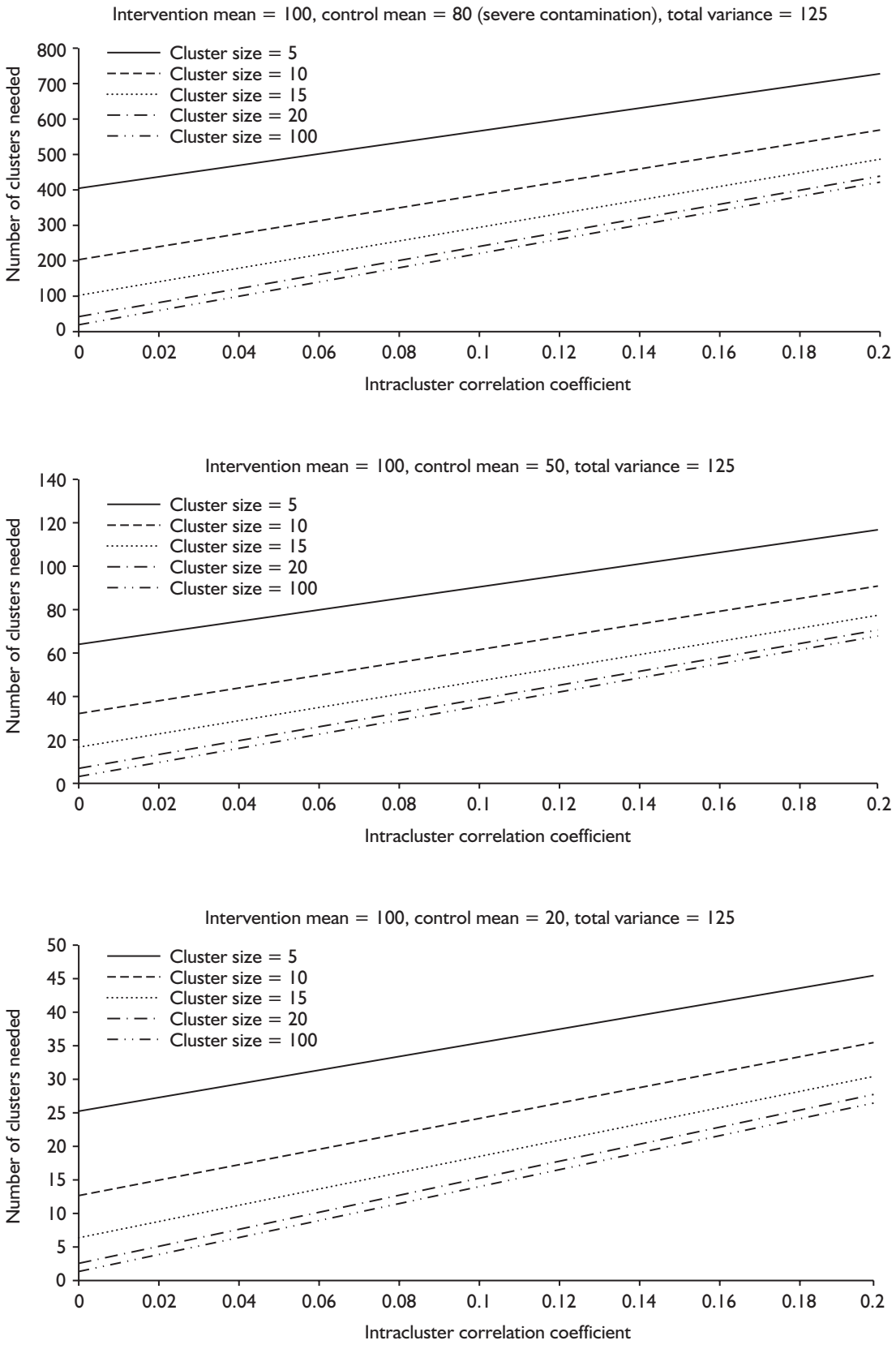

FIGURE 2 *Number of clusters needed with increasing degrees of contamination and intracluster correlation coefficients and different cluster sizes. Increasing degrees of contamination are represented by increasing control means. Total variance is between- plus within-cluster variance. Continued overleaf.*

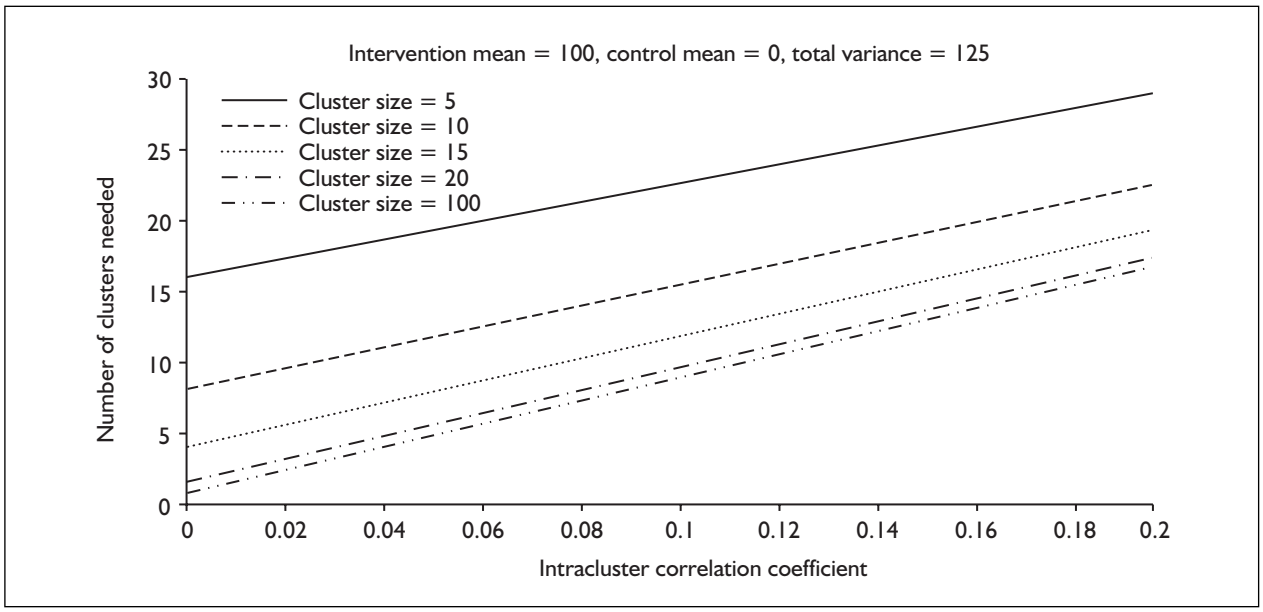

***FIGURE 2** (cont'd)*

Interventions aimed at patients

Interventions aimed at patients vary in their complexity and in their desirability. For example, a patient might receive a leaflet in the mail simply reminding them that a particular screening test is due. A more complex intervention might be a video informing them of how obesity affects health risks and how to lose weight. An even more complex intervention might involve the patient attending a series of motivational interviewing sessions aimed at helping the patient stop smoking. Contamination would occur if control patients were exposed to the intervention, by reading the leaflet, watching the video or attending the motivational interviewing sessions. However, contamination with complex interventions with many ingredients would tend to have a lesser effect in control patients than in intervention patients because the 'dose' of exposure would tend to be less in controls, since the whole intervention package would be difficult to transfer.

Interventions aimed at health professionals

Educational interventions aimed at health professionals can vary in complexity. Such interventions include written information, patient-specific reminders, clinical practice guidelines and in-depth training courses. Contamination would occur if some educational content were passed on to control subjects, either by disseminating the educational materials themselves or through mixing between professionals in intervention and control groups. Contamination would be especially likely if patients were allocated to intervention and control groups and the same health professionals treated patients from both groups. For example, the intervention could be a patient-specific reminder to doctors about elderly patients' need for influenza vaccination. When intervention patients visit their doctor, the doctor receives the reminder, but when control patients visit their doctor, the doctor does not. It would then be likely that the way in which the doctor treats control patients would be affected by the reminders about intervention patients.

Interventions aimed at the general public

Some education interventions are targeted at the general public. For example, health education broadcasts, such as a series of short smoking cessation programmes on local county radio, might be used as a general public intervention to particular counties. An alternative or no broadcast can be aimed at different counties for comparison. Contamination can occur since the broadcast reaches other counties.

Avoiding contamination in trials of educational interventions

Different methods and practices have been proposed to help avoid contamination. These methods and practices impact on the design of the trial and the analysis of the results.

A simple method to avoid contamination is to educate or instruct the trial participants to avoid

contamination. Patients receiving an intervention can be asked not to share information with other patients. Similarly, health professionals can be requested not to discuss the intervention with colleagues who might be involved in a different arm of the trial. Education with regard to general public interventions may be more difficult because personal contact with the trial subjects is unlikely.

Trials designs that avoid contamination

Another method to avoid contamination is to separate intervention and non-intervention groups geographically by randomising groups of subjects rather than individual subjects. For example, patients who attend the same GP surgery will tend to have similar social networks and to live near to one another. This is likely to lead to contamination if both intervention and non-intervention patients are selected from the same GP surgeries. Similarly, health professionals working in the same institution are likely to meet regularly. By administering the intervention arm of the trial at some sites and administering the non-intervention arm at other sites, it is less likely that the two groups will meet and discuss the intervention or share intervention materials, thus decreasing the likelihood of contamination. Similarly, with interventions aimed at the general public, it is wise to choose intervention and non-intervention groups that are distant from each other.

Trials that allocate groups of subjects who share a geographical location, place of treatment or health professional to intervention and control groups are known as cluster randomised trials.[1] Cluster randomised trials have the advantage of reducing the risk of contamination, but they have statistical disadvantages, compared with individually randomised trials, because randomisation information on subjects within clusters cannot be assumed to be independent of each other. Therefore, cluster randomised trials require that a statistical adjustment be made according to the correlation of outcomes between members of that cluster. Furthermore, many clusters are needed to ensure that the groups being compared are in fact comparable at baseline. For both reasons, a cluster randomised trial tends to require a larger sample size to detect a significant effect with a given power and without confounding, compared to an individually randomised trial. Therefore, it has been argued that in many circumstances individual randomisation should be used instead of cluster randomisation. The greater power of the individually randomised trial would outweigh the loss of power caused by contamination, especially where the risk of contamination is low and where outcomes are highly clustered.[2,3] However, this argument leaves aside the separate problem of biased effect estimation. Contamination leads to an underestimate of the magnitude of effect of the intervention, but cluster randomised trials may be confounded if the number of clusters is small. Such bias or confounding may mislead decisions about whether or not the intervention is worthwhile. The trade-offs between cluster and individually randomised trials therefore involve both bias and power.

The Zelen design entails randomising patients before asking for their consent to participate in a trial.[4] Consent can be elicited in one of two ways, namely the single consent design and the double consent design. The single consent design first randomises subjects to the control or intervention arm of the trial. Those randomised to the intervention group are then asked to consent to the intervention. Subjects who refuse consent to the intervention receive the control treatment. Those randomised into the control arm are not asked to consent to receiving the control treatment because this is what they would be receiving anyway. This method avoids disappointing those who may have liked to receive the intervention but are allocated not to, which could affect their behaviour or their reporting of outcomes. By reducing control subjects' awareness of the intervention, it also reduces the likelihood that control patients seek the intervention for themselves. All subjects are included in the analysis in the groups to which they were allocated [an intention-to-treat (ITT) analysis], regardless of their consent. This design may therefore help to avoid contamination of the control group, by not informing them of the alternative intervention, but it would cause non-adherence in the intervention group if many of those allocated to receive the intervention choose not to do so. Furthermore, it may be unethical to include control patients in a trial without their consent.

The double consent Zelen design is similar, except that the control group are asked for consent to receive the control treatment. If they refuse, they are offered an alternative treatment, which may be the experimental intervention. This design ensures that subjects are in agreement with the intervention that they receive. However, as with the single consent Zelen design, a dilution bias can be introduced if a large number of subjects cross over to an intervention contrary to the original randomisation.

Statistical analysis to adjust for contamination

It is possible to adjust for contamination when analysing the results of a trial. A simple method of adjustment where participants are known to be contaminated is to categorise contaminated control members as belonging to the intervention arm, and thus to analyse them as intervention subjects (per protocol analysis). The main problem with this is that one can no longer assume that randomisation has made the groups being compared the same on average. Furthermore, the intervention received via contamination cannot be assumed to be equivalent to the full intervention.

Alternative methods that retain the advantage of randomisation are discussed in Chapter 5. General statistical methods for adjusting for confounding in epidemiological studies can also be used to adjust for contamination. A confounding variable is one that is associated with the exposure (that is, is more or less likely in intervention than in control groups) and is also associated with the outcome. This is particularly a problem with non-randomised controlled trials (RCTs), in which statistical adjustment for confounding is usually necessary. For example, in a trial of nutritional advice, if control group subjects had less healthy diets than intervention group patients before the trial started, the effect of the intervention could be overestimated. Such a confounding factor would need to be adjusted for in the analysis to increase the validity of the effect estimate. Multiple regression methods are usually used to adjust for several confounding variables. Stratified analysis, for example using Mantel Haenzel tests, is an alternative. In controlled trials, this may be necessary if there is poor comparability between exposed and unexposed groups.

Propensity score matching is also used to adjust for confounding or selection bias. Factors that are associated with patients being in one or another arm of the trial are identified, and used to predict the probability that each subject will be allocated to one arm or another. This probability should be 0.5 in a randomised trial with equal sized arms. Propensity scores can be used to adjust for these systematic differences in subjects' characteristics in various ways. There is evidence that adjusting for propensity scores is a better way to adjust for systematic differences in patients in each trial arm than adjustment for confounders.[5]

The main disadvantage of adjusting for baseline differences in randomised trials is that it undermines the fundamental assumption that the two arms are on average the same in all characteristics, including those that have not been measured, and that any baseline differences must have occurred by chance. Thus it may undermine the comparability of groups.

Definition of contamination

In this study, we define contamination to be **the process whereby an intervention intended for members of the trial (intervention or treatment) arm of a study is received by members of another (control) arm.**

This process could be simple and immediate or it could be complex and slow. It could occur both ways at once, so that some members of each group receive some of the intervention intended only for the other group. However, contamination is assumed to occur if at least some members of the control group receive at least some of the intervention.

Outline of the study

Our approach to this challenge has four distinct elements.

Review of the literature

To provide an empirical evidence base, we investigated the extent to which contamination has been reported in trials of educational interventions. Preliminary bibliographic searches identified little direct evidence of the degree of contamination and its effect in reports of educational trials. We therefore looked for indirect evidence of contamination examined in exemplar systematic reviews of trials of educational interventions by comparing effect estimates between trials in which contamination was more or less likely. We report in more detail on one trial which allowed comparison of the design effects of cluster and individually randomised trials.

Eliciting expert opinion

A two-stage Delphi questionnaire, together with a pilot study issued to members of the project team, was used to elicit a consensus of expert opinion on the extent to which different factors are likely to influence contamination and the most appropriate methods of avoiding it.

Simulating the mechanisms and effects of contamination over time

In order to investigate the effect of contamination

according to key parameters such as the efficacy, desirability or transferability of the intervention, a computer simulation was carried out. The simulation model was written for both cluster and individual randomisation and the two designs were compared.

Statistical adjustment for contamination with Complier Average Causal Effect and instrumental variable analysis

We used published examples of trials in which contamination or non-adherence was known to have occurred, to show how to adjust for contamination, with aggregated results, using the Complier Average Causal Effect (CACE) method. We then simulated individual level data to show how, if contamination is known to have occurred, it can be adjusted for using instrumental variable regression methods.

We conclude by bringing these various elements of the project together to provide an overview of when contamination is likely or unlikely and how best to avoid it.

Chapter 2

Literature search

The aim of the literature search was to locate and quantify evidence of, type and cause of contamination in existing research. The search comprised a review of journal articles identified by searching bibliographic databases and handsearching selected journals, and secondary analyses of systematic reviews of exemplar educational interventions. This was an exploratory search. A systematic review of all potentially relevant evidence was beyond the scope of this study because of the wide range of types of educational interventions that were covered. Finally, we report in more detail on a randomised trial within a randomised trial that compared effects estimated by cluster randomisation with effects estimated by individual randomisation.

Direct evidence of contamination reported in trials of educational interventions

In order to locate relevant studies, whether methodological studies or trials, in which contamination was reported, two database searches were performed and updated in May 2005. First, a MEDLINE search on the terms 'contamination' and 'trial' was performed. This identified 689 papers. The documents retrieved via this search method tended either to fall into the category of microbiological contamination or used the term 'contamination' as part of a justification for a cluster randomised trial rather than to provide any evidence about contamination. Five of the papers reported either the degree of contamination or methods to estimate the effect of contamination.

Second, the Web of Science database was searched using the same criteria; 274 papers were retrieved by this search, five of which reported the degree of or methods to estimate contamination. Four of these studies overlapped with the MEDLINE search.

Third, we performed a handsearch of the following journals:

- *Health Education*, 2000 and 2001
- *Medical Education*, 2004
- *American Journal of Health Promotion*, 2003 and 2004
- *Education for Health: Change in Training and Practice*, 1998–2001.

Trials in which extent of contamination was reported

Eight studies clearly stated that contamination had occurred and quantified the extent of contamination (*Table 1*).

Labarre and colleagues[6] conducted a controlled trial of physical education in schools in which intervention and control groups were selected from different age cohorts. They assessed contamination but found none. They speculated that this was because pupils of different ages tend not to play together.

TABLE 1 *Summary of contamination quantities in published literature*

Reference	Intervention	Contamination[a] (%)
Labarre *et al.*, 1994[6]	Physical education	0
Courneya *et al.*, 2003[7]	Moderate intensity exercise	22
Courneya *et al.*, 2004[8]	Home-based exercise	52
Courneya *et al.*, 2003[9]	Home-based exercise	52
Ross *et al.*, 2004[10]	Multiple media	65
Goel *et al.*, 1998[11]	Mammography screening	17
Tilgren *et al.*, 1998[12]	In-home health education	40/25/2.6[b]
Stewart-Brown *et al.*, 2004[13]	Parenting training	10

[a] Percentage of control patients who were exposed to the intervention.
[b] Three different types of contamination reported.

Courneya and colleagues[9] describe a randomised trial to evaluate enhanced quality of life in cancer survivors by comparing GP group psychotherapy with GP group psychotherapy plus moderate intensity exercise; 22% of control subjects received the intervention.

Courneya and and colleagues[7] measured adherence by the intervention group and contamination in the control group in a different randomised trial to determine effects of home-based exercise intervention on change of quality of life in recently resected colorectal cancer survivors; 52% of the control group and 76% of the intervention group were exposed to the intervention.

In another study, Courneya and and colleagues[8] conducted a trial to examine predictors of adherence and contamination in an RCT of a home-based exercise intervention for colorectal cancer survivors. The trial results suggest that contamination in the control group was 52% and adherence in the intervention group was 76%. Regression analysis showed that subjects' intentions ($\beta = 0.35$, $p = 0.001$) and stages of change ($\beta = 0.35$, $p = 0.095$) regarding exercise at baseline explained 30% of the variance in contamination.

Ross and colleagues[10] presented evidence of contamination in a community-level syphilis intervention trial using several media, namely brochures and posters, coasters, matchbooks, T-shirts, videos and billboards containing slogans and logos. Control members were exposed to an average of 2.04 intervention items whereas intervention members were exposed to an average of 2.99 items; 65% of the control group were exposed to at least one of the intervention media (contamination), compared with 71% of the intervention group (adherence).

Goel and and colleagues[11] estimated contamination in a trial of mammography screening by using insurance billing records to identify women who sought a non-screening mammogram outside the screening trial. The trial found that, for women aged 40–49 years, 2.2% in the intervention group and 14.1% in the control group had a claim for at least one bilateral mammogram, indicating adherence and contamination respectively. In the age group 50–59 years, 4.5% in the intervention group and 16.7% in the control group were exposed.

Tilgren and colleagues[12] conducted a study to assess the prevalence and intensity of contamination in a community-based trial of education about cervical cancer screening. The trial used individual randomisation and had two control groups. The intervention was conducted in the homes of subjects by lay health educators. Pre-test surveys were conducted in one of the control groups and post-test surveys in both control groups. Of the 185 control subjects who were interviewed, 40% were aware of the programme, 25% knew that the programme concerned cervical cancer and three had specific knowledge of the content or had seen the educational materials.

Stewart-Brown and colleagues[13] conducted a study to analyse the effectiveness of a parenting programme on a sample of parents of children whose behaviour was worse than average. A total of 116 parents were randomised into control and intervention groups; 23 intervention arm parents and 15 control arm parents were interviewed about their approach to, and difficulties with, parenting. Intervention parents were asked their opinion of the intervention. The intervention targeted behaviour and used videotape modelling and experiential learning including child play, praise and rewards. A significant change ($p < 0.05$) in a positive direction was observed for the intervention group at 12 months using the Eyberg inventory. A total of 31 of 60 parents in the intervention group attended 50% or more sessions. Control parents showed significant improvements on all scales of the parenting stress index and intervention patients showed significant improvements on all parenting stress index measurements except parent–child interaction. Control parents indicated that the questionnaires had encouraged them to reflect on their parenting and four of 41 control parents indicated that they had attended a parenting programme before 12-month follow-up (10% contamination); also one non-attender from the intervention group attended a parenting programme within 12 months.

In summary, these studies show how different types of contamination may occur in trials of different types of educational intervention. However, they do not provide evidence of how much contamination affected the statistical significance and validity of their results.

Indirect evidence of contamination among a diverse group of trials of educational interventions aimed at health professionals

In order to obtain a relevant source of data of trials of educational intervention in which to study

contamination, we examined one major systematic review, by Grimshaw and colleagues.[14] This reviewed 235 studies of educational interventions together with financial and strategic interventions. The investigators had categorised each study in terms of its potential for contamination as 'contamination avoidance done' (D), 'contamination avoidance not clear' (NC) and 'contamination avoidance not done' (ND). 'Contamination avoidance done' usually meant that a cluster randomised trial design was used. We analysed separately the following subsets of these studies:

1. All studies.
2. A subset of relatively homogeneous studies. Each study evaluated an intervention which used some type of reminder (to receive vaccination or screening) compared with a control group that received nothing.
3. A subset of item 2 that met at least one study quality criterion.

For each analysis we aimed to test the prior hypothesis that, for the wide range of educational interventions evaluated, effects would tend to be greater in trials that avoided contamination by cluster randomisation compared with trials that were individually randomised.

All studies analysis

Methods

The Grimshaw review data for 158 studies from 201 papers[15–215] were analysed, using the following variables reported for each study. The study designs included both cluster and non-cluster RCTs and controlled before–after studies:

1. Trial design (individual or cluster allocation to comparison groups).
2. Targeted behaviour (test ordering, prescribing, prevention or patient education).
3. Type of intervention (educational materials, educational meetings, reminders, audit and feedback or financial interventions).
4. Educational outcomes in control and intervention groups.
5. Difference in outcome between control and intervention groups.
6. Reviewers' classification of contamination avoidance:
   (a) Done: cluster allocation by community, institution or practice and unlikely that control group received the intervention.
   (b) Not clear: professionals allocated within a clinic or practice and possible that communication between experimental and control group professionals could have occurred.
   (c) Not done: likely that control group received the intervention, for example cross-over trials or if patients rather than professionals were randomised.

We compared outcomes in intervention and control groups using L'Abbé plots, as follows. These plots are explained in a paper by L'Abbé and colleagues[216] and used by Song.[217]

1. Each study from the Grimshaw review was plotted as a coordinate point on the plot with reference [intervention arm effect (%), control arm effect (%)].
2. A line representing control = intervention was plotted
3. The shortest distance from each coordinate point to the control = intervention line was calculated (the shortest distance will meet the control = intervention line at a right-angle). If the effect of the intervention is the same as the control, the mean L'Abbé distance will be zero. Otherwise, the distance represents the improvement or otherwise as a result of the intervention.
4. Also included was a line of best fit passing through the origin, calculated by linear regression, and summarising the tendency of the data.

To maximise the sample size, we initially examined all trials, regardless of their design, and compared them according to their contamination classification. The L'Abbé plots for the three types of trial are shown in *Figures 3–5*.

Results

The results of this preliminary analysis were the opposite of what we had hypothesised. We would expect contamination to decrease the L'Abbé distance since it would bias the control outcome in the direction of the intervention outcome. However, *Table 2* and *Figure 3* show that the trials classified as 'contamination avoided' have the smallest distances from the control = intervention line and therefore have the smallest difference between control and intervention effect. The trials with contamination avoidance not clear or not avoided have larger differences, with contamination not avoided trials having the largest difference between control and intervention effect.

Summary statistics for the distances from each point to the control = intervention line are given

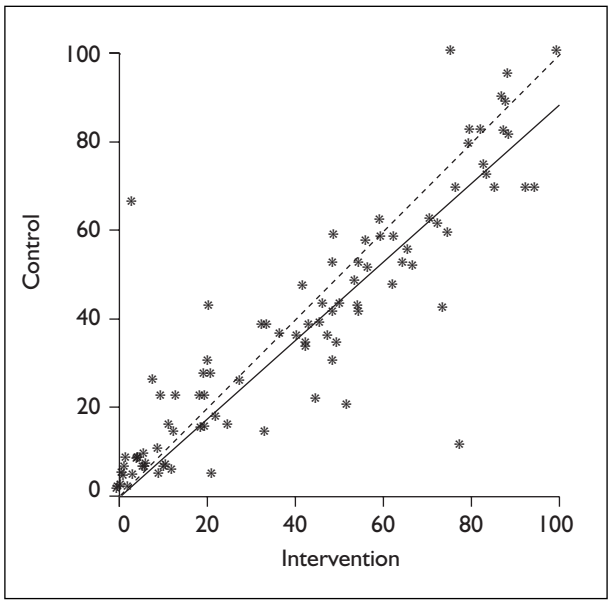

FIGURE 3 *Rate ratio fit: all trials, contamination avoided*

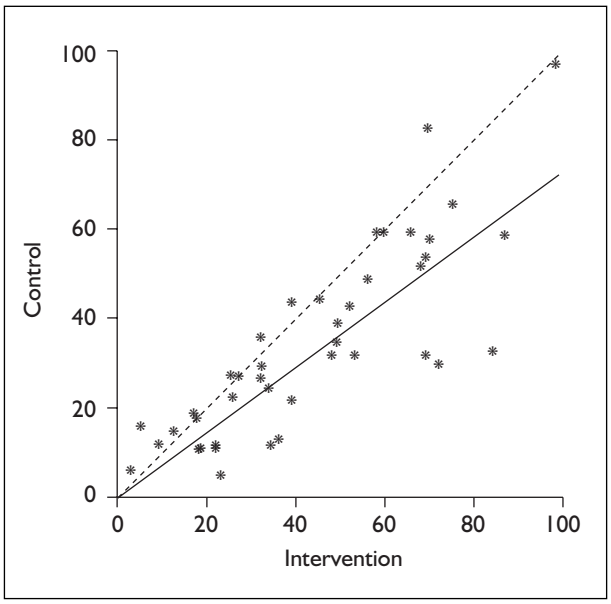

FIGURE 5 *Rate ratio fit: all trials, contamination not avoided*

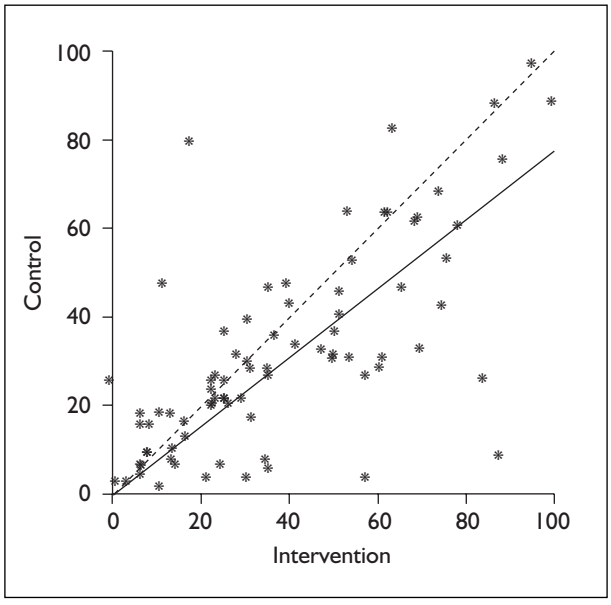

FIGURE 4 *Rate ratio fit: all trials, contamination avoidance not clear*

in *Table 2*. These figures are also summarised as box plots in *Figure 6*.

There are several possible explanations for this finding. First, the trials were very heterogeneous, so the above comparisons could be invalid. Second, in some trials the interventions may have had a negative, although beneficial, effect (a type of heterogeneity). Third, cluster randomisation may have been used more often in situations where contamination was most likely, but did not eliminate it. Fourth, researchers who were experienced enough to know where cluster randomisation was appropriate may also have been better at avoiding other sources of bias. Fifth, most trials were small, so publication bias could have resulted from those with larger effects being more likely to be published. In summary, therefore, the above results should be interpreted cautiously.

To test the second of these explanations (whether these results had been biased by inclusion of trials with a negative direction of intervention effect), we repeated the analysis, first excluding studies with a negative direction of effect [leaving 82 trials with contamination avoidance done (the corresponding papers with contamination avoidance done and positive effect are Refs 15–19, 21, 25, 26, 28, 33, 38, 44, 52, 54, 55, 57–61, 65, 66, 68–74, 75, 81, 82, 85, 87–89, 93–99, 108–113, 118, 119, 122, 127, 128, 132–134, 137, 142–144, 148, 152, 153, 154, 160, 162, 175, 177, 181, 184, 185, 195–197, 201, 202, 211, 212), 65 not clear (the corresponding papers with contamination avoidance not clear and positive effect are Refs 20, 22, 24, 29, 30, 32, 43, 45, 46-50, 53, 56, 62–64, 67, 78, 80, 83, 106, 114, 115, 120, 121, 123, 129, 139, 140, 145, 147, 156, 161, 172–174, 178, 179, 183, 187, 192–194, 208, 210, 215, 216) and 39 not done (the corresponding papers with contamination avoidance not done and positive effect are Refs 27, 31, 35, 37, 39, 40, 42, 75, 84, 90, 91, 100, 101–104, 107, 125, 126, 136, 138, 155, 163, 167, 182, 188, 190, 191, 198, 203–207, 209)] and second excluding studies with a positive direction of effect (leaving 11[15,34,86,116,117,135,158,159,166,211,213]

TABLE 2 *Summary statistics for L'Abbé distances*

	All studies					
	Minimum	**1st quartile**	**Median**	**Mean**	**3rd quartile**	**Maximum**
Contamination avoided	0	1.414	4.384	6.631	9.192	48.08
Contamination avoidance not clear	0.2121	2.616	6.081	9.848	14.14	57.28
Contamination not avoided	0	2.121	7.354	9.259	13.26	38.18
	Studies with positive direction of effect					
Contamination avoidance done	0	1.573	4.95	7.284	9.458	48.08
Contamination avoidance not clear	0.2121	3.323	6.364	10.77	14.85	57.28
Contamination not avoided	0	3.359	8.485	9.908	13.44	38.18
	Studies with negative direction of effect					
Contamination avoidance done	0.354	0.746	1.414	1.763	2.475	3.96
Contamination avoidance not clear	0.354	0.530	0.707	0.825	1.061	1.414
Contamination not avoided	0.212	0.707	1.909	5.864	6.116	42.43

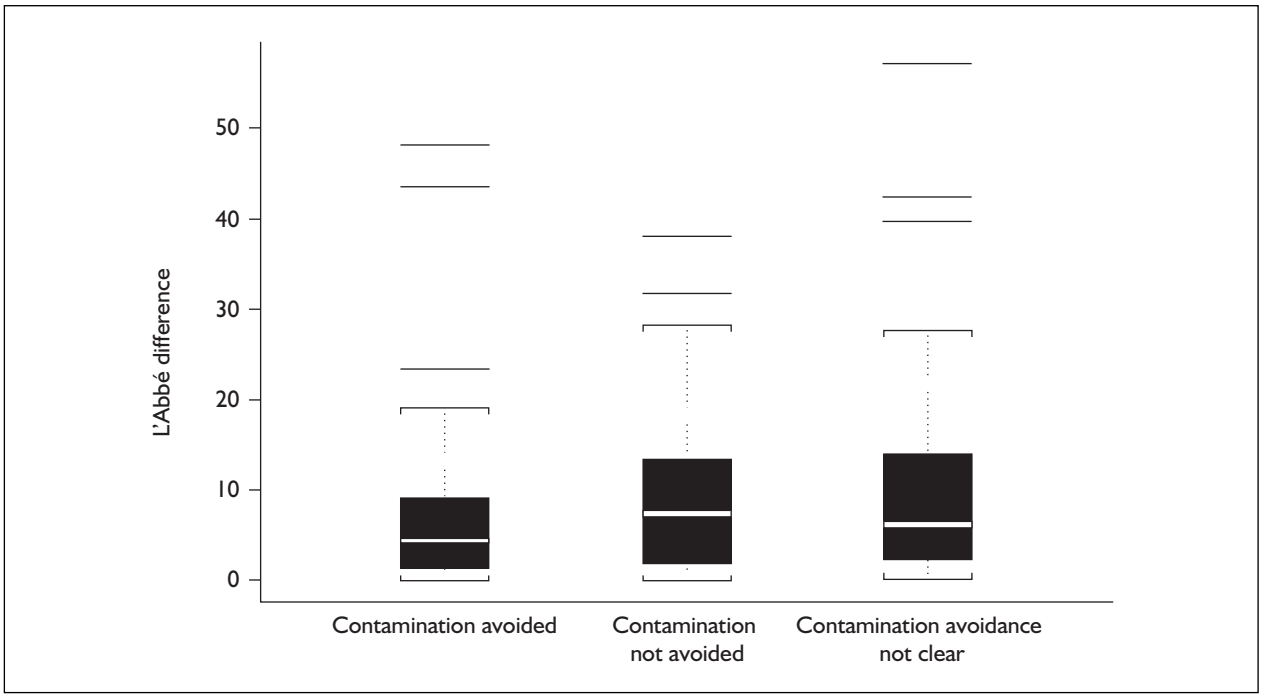

FIGURE 6 *Box plots of L'Abbé differences for all studies reviewed by Grimshaw and colleagues*[14]

trials with contamination avoidance done, 15[32,41,43,51,67,77,79,124,141,146,157,176,183] not clear and three[31,92,105] not done) (some trials in the Grimshaw review quoted more than one result; it is therefore possible for a study to feature in both positive and negative effect categories).

For trials where the effect of the intervention is expected to be positive, the distances measured by the L'Abbé plot are summarised in *Table 2* and displayed in the L'Abbé plots in *Figures 7–9* and as box plots in *Figure 10*.

These results show that, for positive effect interventions, 'contamination avoidance done' trials still had smaller effects than the other trials. However, the 'contamination avoidance not clear' trials show a slightly larger mean effect size than the 'contamination avoidance not done' trials.

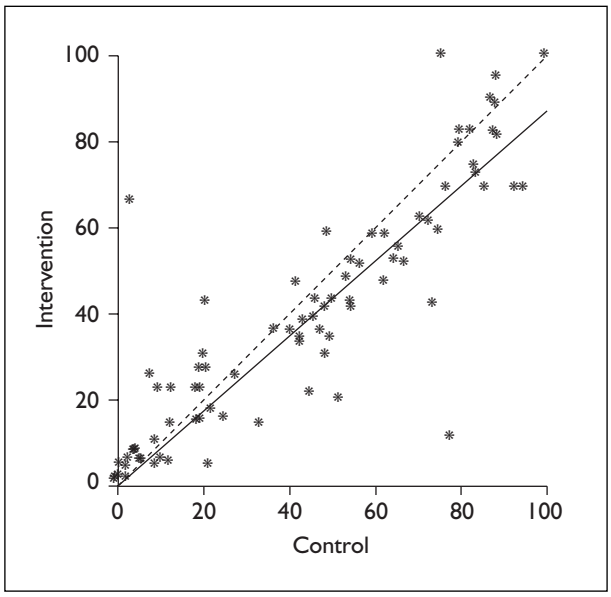

FIGURE 7 *Positive effects trials, rate ratio fit: contamination avoided*

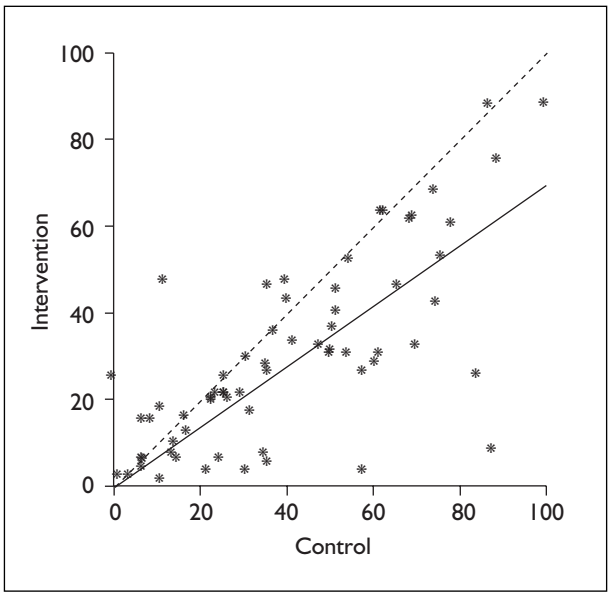

FIGURE 9 *Positive effect trials, rate ratio fit: contamination not avoided*

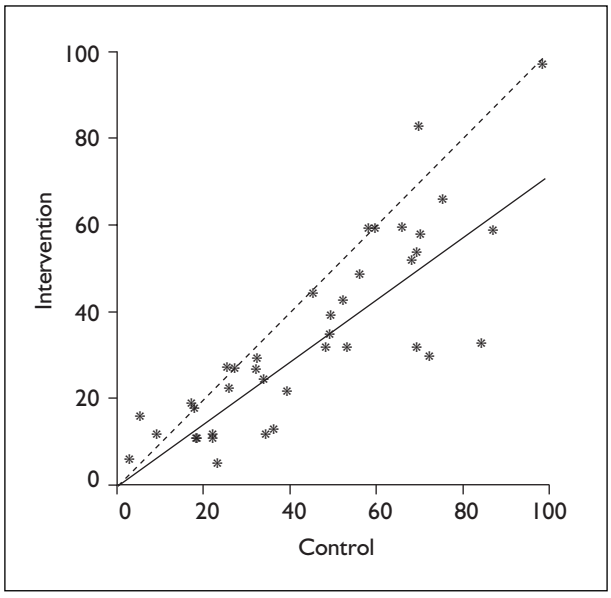

FIGURE 8 *Positive effect trials, rate ratio fit: contamination avoidance not clear*

We then excluded the interventions with positive effects and studied only those whose effectiveness decreased the measure of effect. The L'Abbé differences for studies with negative direction of effect are summarised in *Table 2*. The L'Abbé plots are shown in *Figures 11–13* and box plots of these differences are shown in *Figure 14*.

As the table and plots show, studies with negative effects agree with the hypothesis that avoiding contamination increases the effect size. However, the sample sizes are very small and so these findings could be due to chance.

Subset of higher quality studies

It is possible that the quality of the study will influence the effect of contamination since poor-quality trial results may be biased by factors other than contamination. We therefore examined only studies that met at least one quality criterion used in the Grimshaw review, namely randomisation concealed, outcome assessment blind, outcomes reliable and baseline measurements comparable. An analysis of the positive effect studies of this type follows.

L'Abbé plots are shown in *Figures 15–17* and the L'Abbé distances for these plots are summarised in *Table 3*.

A graphical representation of these L'Abbé distances is shown in *Figure 18*. The results suggest that the differences between control and intervention results are greater for trials where contamination was not avoided. This result is contrary to the hypothesis that contamination reduces the estimate of effect.

This preliminary analysis yielded no evidence to suggest that contamination biases the estimated efficacy of educational interventions. However, because of the variety of types of intervention and aim of the trials, it is possible that the above comparison is one of qualitatively different trials. We therefore performed a further, more detailed analysis of a more homogeneous set of interventions.

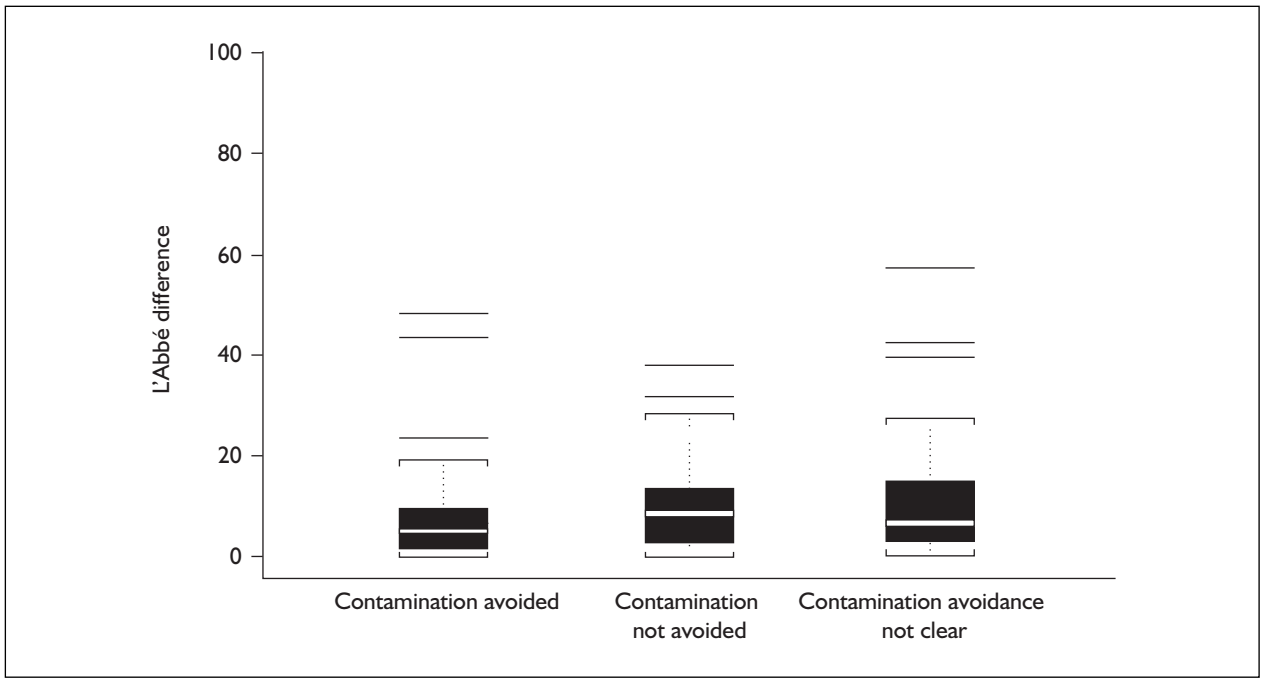

FIGURE 10 *Box plots of L'Abbé differences for positive effect studies in Grimshaw and colleagues[14]*

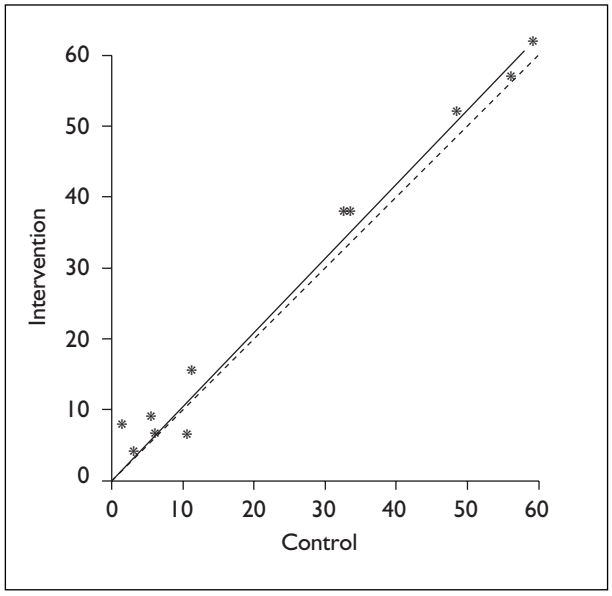

FIGURE 11 *Negative effect trials, rate ratio fit: contamination avoided*

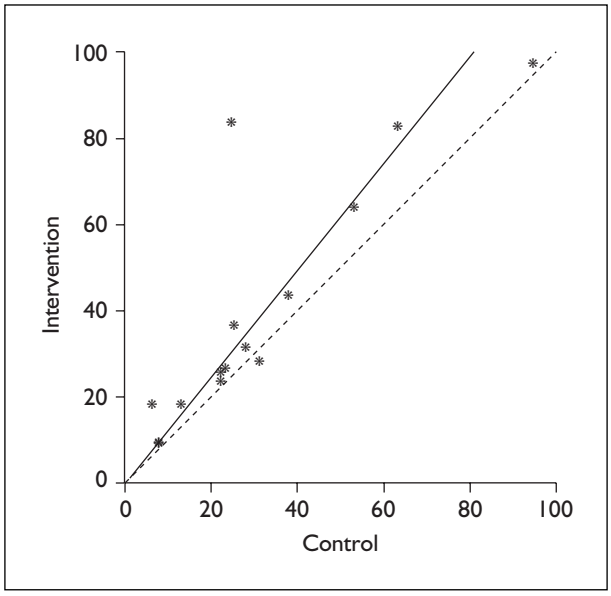

FIGURE 12 *Negative effect trials, rate ratio fit: contamination avoidance not clear*

Indirect evidence of contamination among a homogeneous group of trials of educational interventions aimed at health professionals

We then analysed a set of similar studies to test the hypothesis that the differences attributable to contamination between these studies would become clearer. This was a subset of the studies reviewed by Grimshaw and colleagues.[14] They all evaluated a reminder intervention, in which professionals were reminded about appropriate ways of managing particular patients. It includes 38 studies published between 46 research papers. The types of reminder vary from pop-up messages on a doctor's computer to the completion of patient encounter forms and letters to patients. The subjects of the intervention vary between

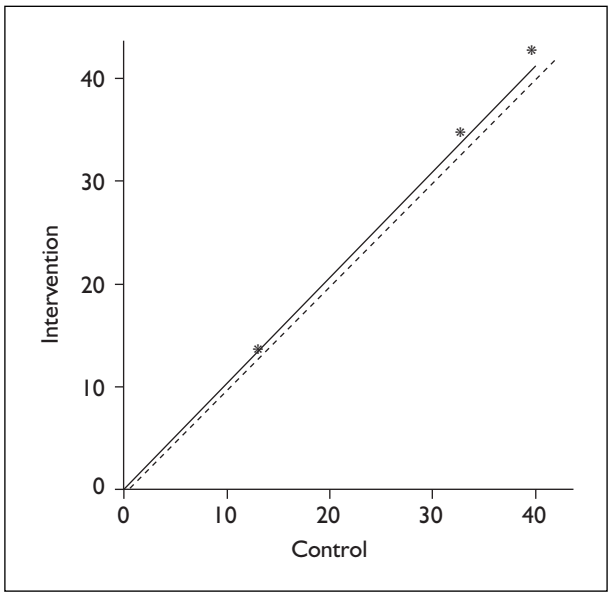

FIGURE 13 *Negative effect trials, rate ratio fit: contamination not avoided*

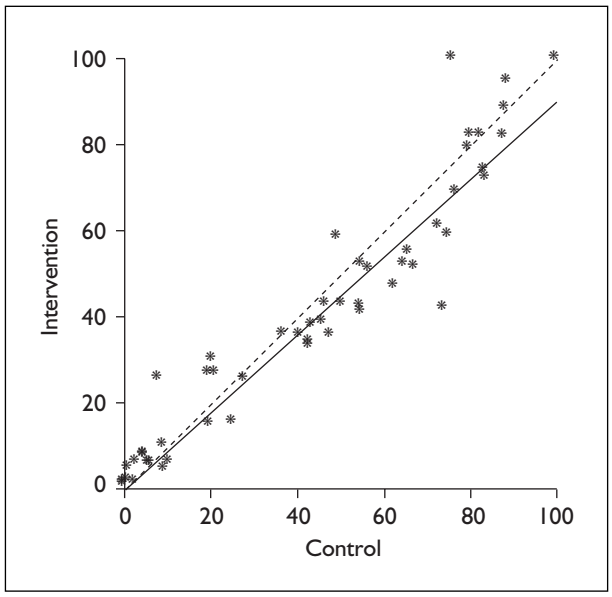

FIGURE 15 *L'Abbé plot for all high-quality studies with positive effect, rate ratio fit: contamination avoided*

physician, health professional and patient, and the duration of studies varies from weeks to years.

A summary of the L'Abbé differences for these homogeneous studies is given in *Table 4*. However, there were just two studies[81,127,128] classified as contamination avoided, 18[29,43,45,50,51,106,107,114,146,172,174,179] classified as not clear and 17[27,40,84,92,125,126,136,167,190,203,207] classified as not done.

These differences are further summarised in *Figures 19–21* and box plots are shown in *Figure 22*.

Although the sample size for contamination avoided studies is small, the above box plot is an illustration of the behaviour that might be expected if contamination reduces effect size.

For the majority of homogeneous studies, we have access to the study sample size. For contamination

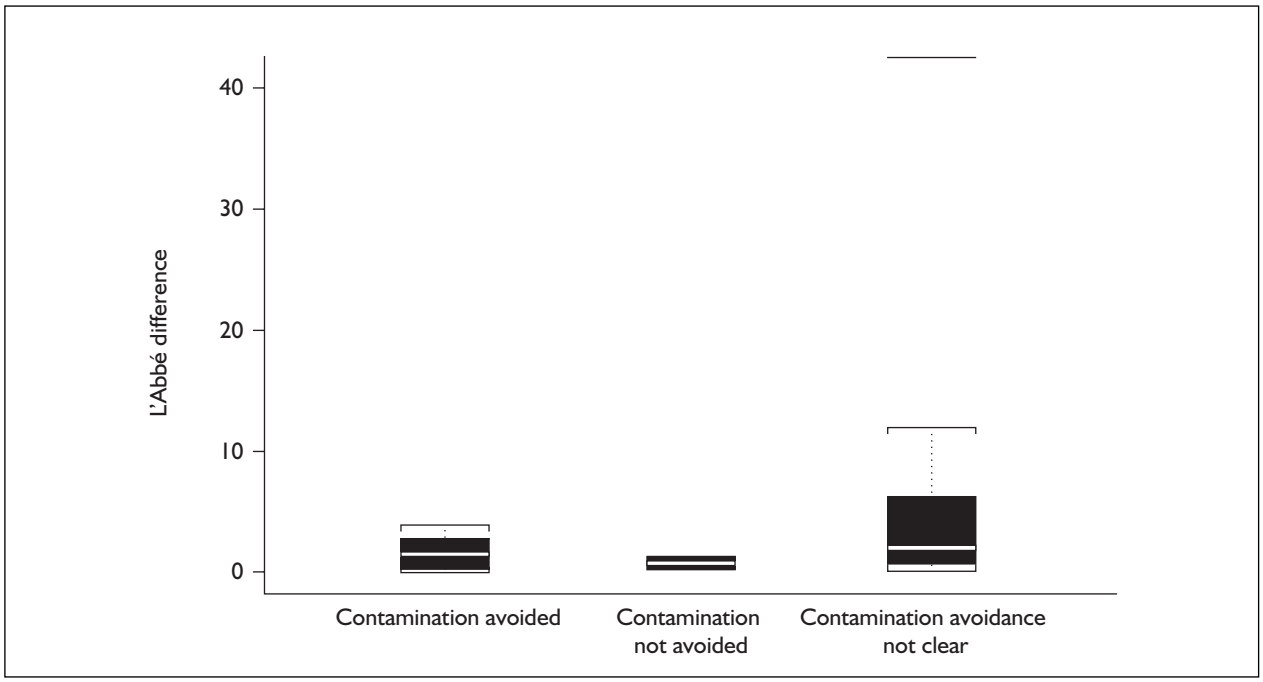

FIGURE 14 *Box plots of L'Abbé differences for negative effect studies in Grimshaw and colleagues*[14]

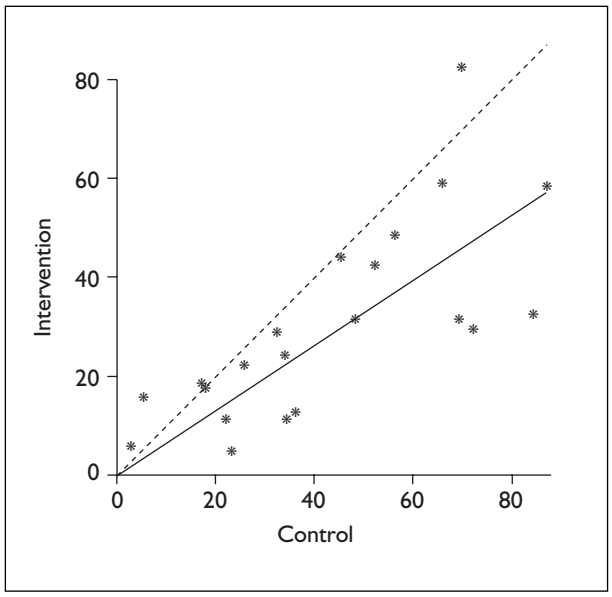

FIGURE 16 *L'Abbé plot for all high-quality studies with positive effect, rate ratio fit: contamination not avoided*

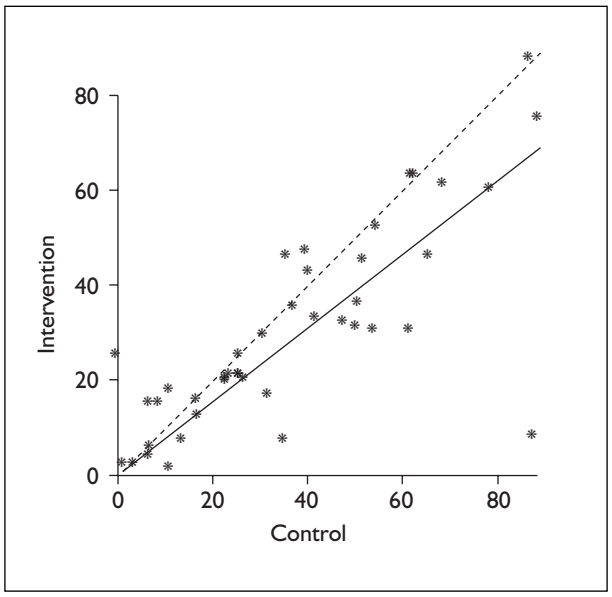

FIGURE 17 *L'Abbé plot for all high-quality studies with positive effect, rate ratio fit: contamination avoidance not clear*

TABLE 3 *Summary of L'Abbé distances for all high-quality positive effect studies*

	Minimum	1st quartile	Median	Mean	3rd quartile	Maximum
Contamination avoided[a]	0	1.414	4.419	5.406	7.973	23.330
Contamination avoidance not clear[b]	0.424	4.243	8.485	11.950	17.960	38.180
Contamination not avoided[c]	0.212	2.333	4.243	7.781	11.140	57.280

[a] This group consisted of 52 studies from Refs 15, 21, 25, 26, 28, 38, 44, 54, 55, 65, 68–72, 75, 81, 82, 85, 89, 93–96, 98, 108, 110, 111, 116, 117, 122, 132–134, 137, 142, 144, 148–151, 160, 162, 175, 177, 180, 181, 184–186, 195–197, 201, 202, 212.
[b] The contamination avoidance not clear group here consisted of 41 studies from the Refs 20, 22, 24, 29, 32, 43, 50, 62, 63, 80, 114, 115, 121, 123, 129, 139, 140, 145, 156, 161, 172–174, 178, 179, 190, 187, 193, 194, 210, 215.
[c] The contamination not avoided group here consisted of 21 studies from the Refs 31, 35, 39, 42, 76, 84, 100, 104, 125, 138, 163, 165, 167, 171, 191, 198, 209.

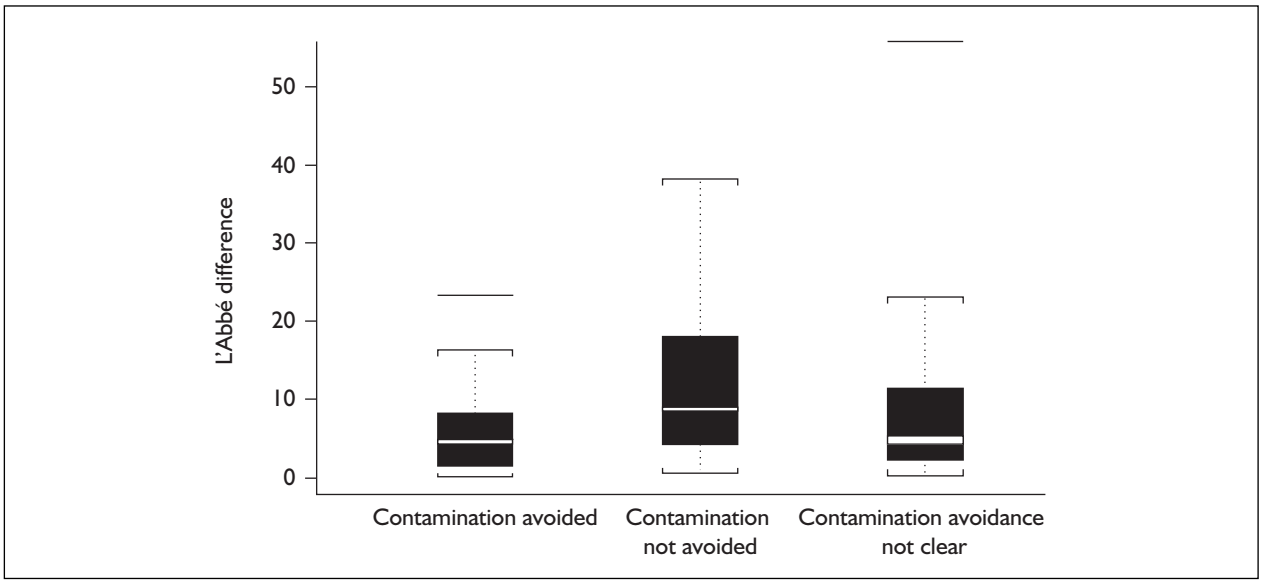

FIGURE 18 *Box plots of L'Abbé distances for all high-quality positive effect studies in Grimshaw and colleagues*[14]

TABLE 4 *Summary of L'Abbé distances for homogeneous studies in Grimshaw and colleagues*[14]

	Minimum	1st quartile	Median	Mean	3rd quartile	Maximum
Contamination avoided	14.14	16.44	18.74	18.74	21.04	23.33
Contamination avoidance not clear	0	0.808	2.687	4.674	7.707	14.71
Contamination not avoided	0.566	1.414	8.485	8.507	13.440	28.28

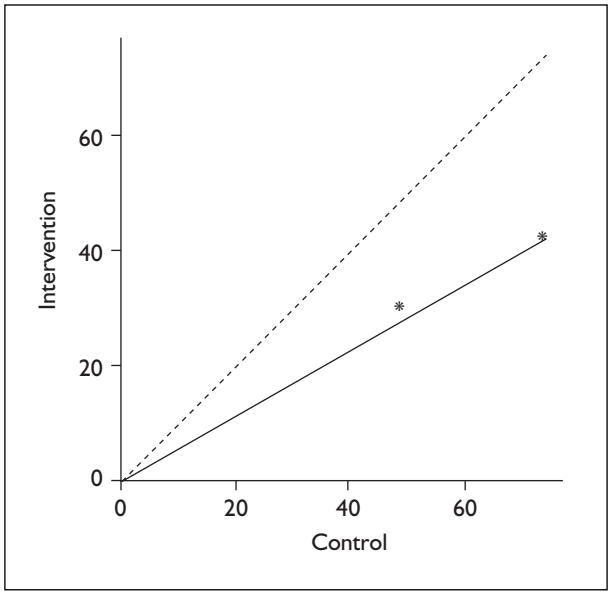

FIGURE 19 *Homogeneous studies, rate ratio fit: contamination avoided*

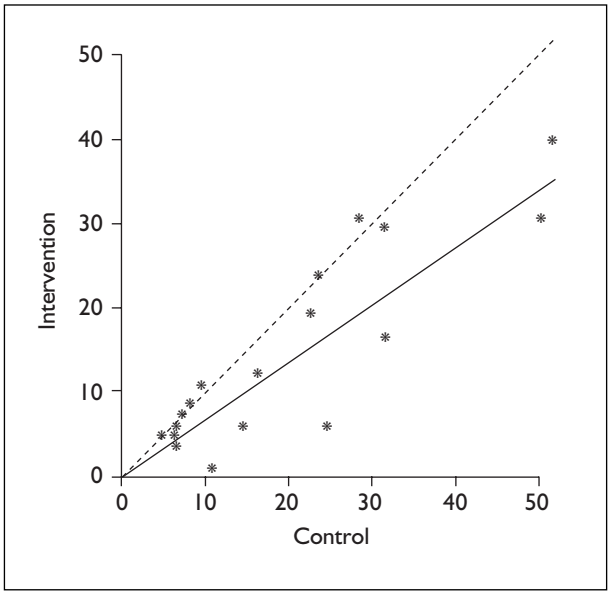

FIGURE 21 *Homogeneous studies, rate ratio fit: contamination not avoided*

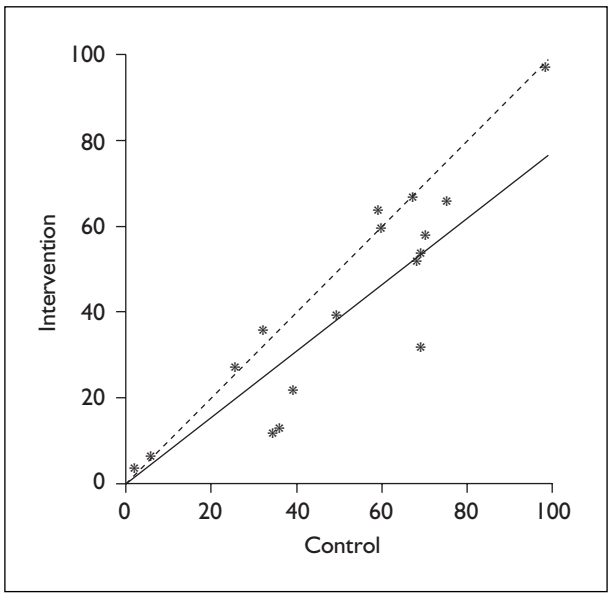

FIGURE 20 *Homogeneous studies, rate ratio fit: contamination avoidance not clear*

avoided studies, the mean sample size is 64.5 and for studies where contamination was not avoided, the mean sample size is 451.3; both datasets have very large variance and a simple test of mean difference between the samples shows no evidence to suggest equality of the means ($p = 0.21$).

Subset of higher quality studies

In addition to the analysis of data from the perspective of contamination likelihood, it was also decided to examine the data from the point of view of study quality. By removing low-quality studies from our analysis, we intended to obtain a set of studies that were both homogeneous in nature and with valid effect estimates.

Methods

We included studies that met at least one study quality criterion, that is, randomisation concealment, blinded assessment and baseline measurement comparability This further reduced the sample size. In order to make use of all studies, we omitted the classification 'contamination avoidance not clear' and divided the studies into individually randomised and cluster randomised trials (since individually randomised trials are more likely to be contaminated).

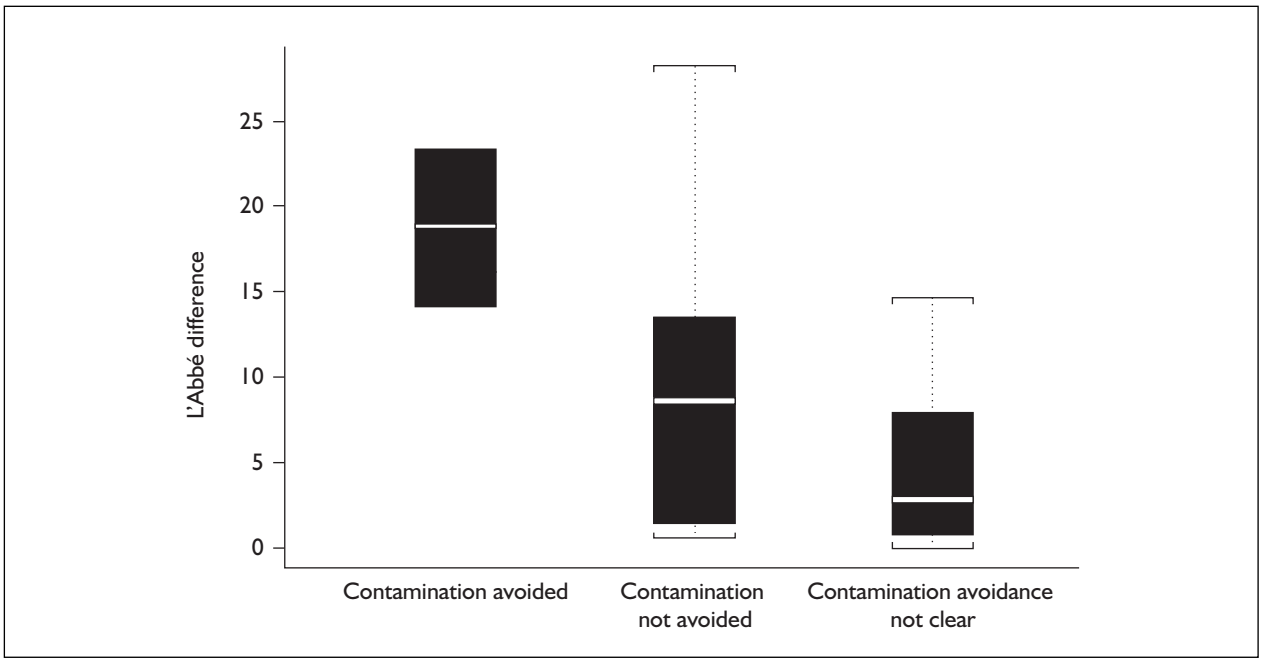

FIGURE 22 *Box plots for L'Abbé distances for homogeneous studies in Grimshaw and colleagues*[14]

Results

We performed L'Abbé plot analysis on the high-quality individually randomised and high-quality cluster randomised study results. Sample sizes were very small with just six[27,40,92,126,190,203] individual and five[43,114,167–171,179] cluster studies. L'Abbé plots for this analysis are shown in *Figures 23* and *24*.

Bearing in mind the small sample sizes, the fitted lines in the plots suggest that the cluster randomised trial results are further from the control = intervention line with two points close to control = intervention. The individually randomised trials have four points close to the control = intervention line and two studies with a larger effect. Summaries of these differences are given in *Table 5*.

These results show that the L'Abbé distances for the high-quality cluster randomised studies are larger for first quartile, median, mean and third

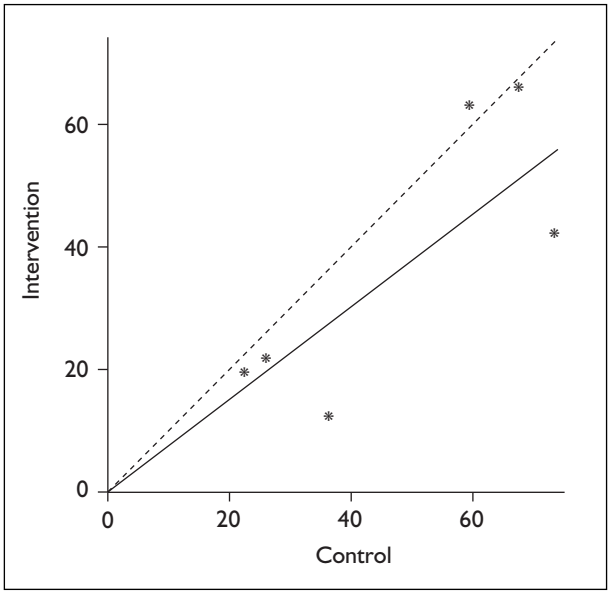

FIGURE 23 *Rate ratio fit: individually randomised high-quality studies*

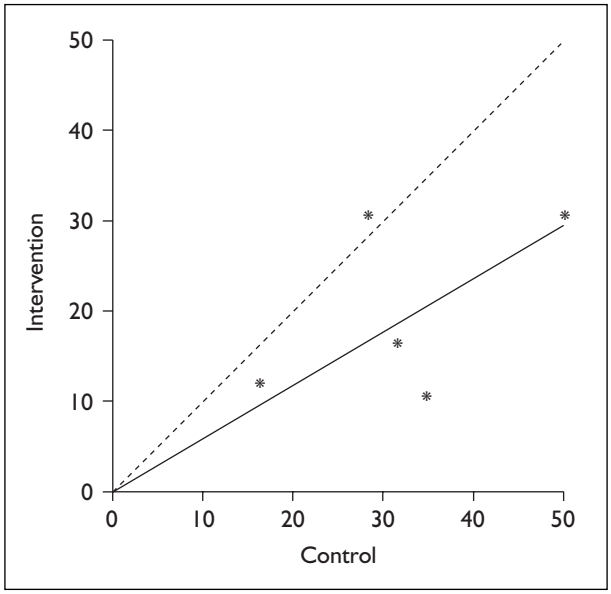

FIGURE 24 *Rate ratio fit: cluster randomised high-quality studies*

TABLE 5 *Summary of L'Abbé distances for high-quality homogeneous studies in Grimshaw and colleagues*[14]

	Minimum	1st quartile	Median	Mean	3rd quartile	Maximum
Individually randomised	1.414	2.422	3.783	8.768	14.690	23.330
Cluster randomised	0.7071	3.818	11.6	9.758	14.71	17.96

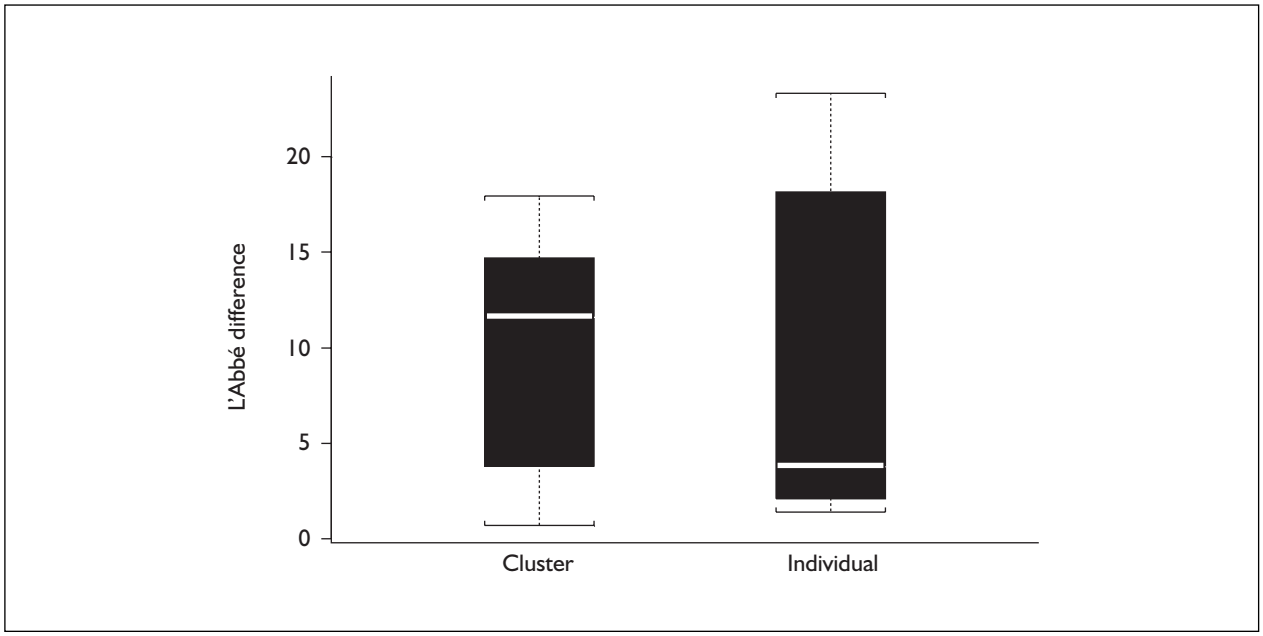

FIGURE 25 *Box plots for L'Abbé distances for cluster and individually randomised studies*

quartile than the individually randomised studies. These results are further illustrated in the box plots in *Figure 25*.

Although the sample size for this analysis is too small to enable us to make a confident inference, the results are in keeping with the suggestion that studies avoiding contamination tend to have larger effect size.

Indirect evidence of contamination among a homogeneous group of trials of educational interventions aimed at patients

We examined trials evaluating smoking cessation interventions based on the stages of change approach and reviewed in Riemsma and colleagues.[218] This exemplifies trials of a range of related interventions aiming to change patient attitudes or behaviour likely to affect their health.

Stages of change analysis

Our second group of homogeneous papers for analysis were those published in the stages of change review by Riemsma and colleagues.[218] We examined 13 homogeneous studies of smoking cessation interventions reviewed in that report.[219–231] Of these studies, 10[219–228] reported results in such a way that the results could be extracted and compared. The interventions used in these studies varied, but most considered the stages of smoking cessation pre-contemplation, contemplation and action. Other results included quit rates of various durations and over various intervals of time. The main problem with some of the results reported in these studies was that very few of the interventions reviewed resulted in a significant absolute quit rate, which makes estimation of the amount of contamination difficult (since if there is no effect, there can be no contamination effect).

Aim

As with our analysis of Grimshaw and colleagues' review,[14] it was our intention to compare the results of the smoking cessation

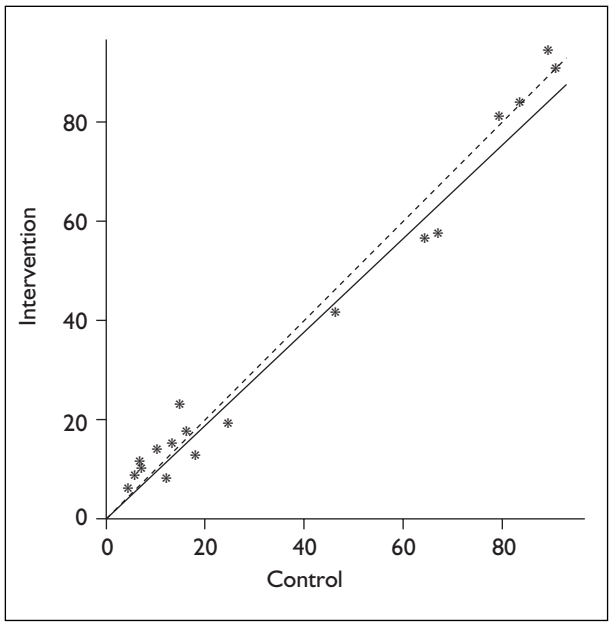

FIGURE 26 *Rate ratio fit: interventions with high likelihood of contamination*

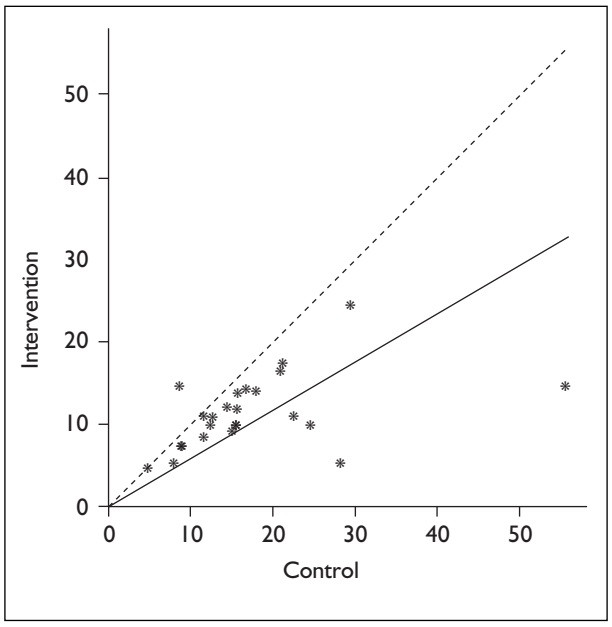

FIGURE 27 *Rate ratio fit: interventions with low likelihood of contamination*

studies and examine the effect, if any, that contamination has on the measured intervention effect.

Method

Having obtained the original papers for the 13 smoking cessation studies, we sought to identify results that could be compared between studies (such as quit rate percentages), and classified studies according to their potential for contamination. Trials where the intervention and control arms were clustered to geographically separate trial arms were classified as having low likelihood of contamination. Trials where the trial design was likely to lead to contamination (such as the same doctor administering different arms of the trial) were classified as highly likely to lead to contamination, even where the intervention was difficult to transfer. Results where the intervention was less effective than the control were ignored. Some studies asserted that contamination occurred during the trial and these were classified as having high likelihood of contamination.

Results

The L'Abbé plots in *Figures 26* and *27* show the results for high and low contamination likelihood studies.

The fitted line in *Figure 27* for studies with low likelihood of contamination suggests a larger general effect size than for the studies with high likelihood of contamination. In fact, the points representing the control and intervention effects are evenly scattered around the control = intervention line, suggesting a similar effect in both arms of the trial. This could be attributable to an ineffective intervention, but could equally be attributed to contamination, especially in view of the fact that some of the trial results in *Figure 26* are known to be contaminated.

Summaries of the L'Abbé distances are given in *Table 6*. These distances are also plotted in the box plots in *Figure 28*.

The results suggest that the contaminated trials have a smaller effect than the trials that are

TABLE 6 *L'Abbé distance summary for homogeneous smoking cessation studies*

	Minimum	1st quartile	Median	Mean	3rd quartile	Maximum
Contamination not likely	1.061	2.563	3.429	5.455	4.95	29.98
Contamination highly likely	0.141	0.583	1.732	2.726	4.685	8.344

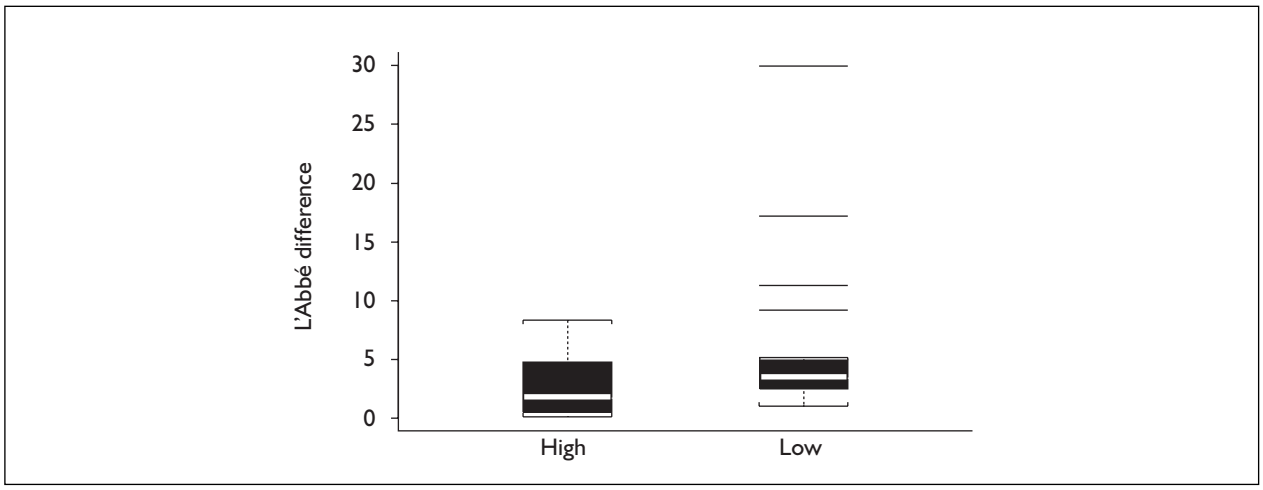

FIGURE 28 *Box plots of L'Abbé distances for smoking cessation studies*

unlikely to be contaminated. This conclusion is different from that observed for the Grimshaw data, but is in agreement with our theoretical expectations. However, the sample sizes used in this analysis are small.

Indirect evidence of contamination in a randomised trial of an educational intervention aimed at patients, comparing design effects of cluster and individual randomisation

A study entitled 'The effectiveness of enhanced oral health advice and instruction upon patient oral hygiene, knowledge and self-reported behaviour: two parallel trials using patient and cluster randomised controlled designs' is particularly relevant to the comparison of cluster or individually randomised trial designs. This study has been submitted for publication elsewhere and is included in this section by kind permission of the Vocational Dental Practitioners Trials Group (C Ramsay, University of Aberdeen: personal communication, 2005). We provide here an overview of the trial focusing on the evidence that it provides about contamination bias in cluster compared with individually controlled trials.

Introduction

Increasing emphasis is now being placed by patients, professionals and policy makers on the need for the provision of preventive care within the General Dental Service. A priority is the provision of effective oral health advice to dental patients. Studies show that current methods of delivering oral hygiene advice have varying success in influencing patient oral hygiene but the quality of the studies has been mixed. This study aimed to evaluate the effectiveness of providing oral health advice and instruction, in general dental practice, on the oral hygiene, cognitions and self-reported behaviour of adults. In addition, to provide information on whether a cluster or individual randomisation should be the preferred design of future practice-based oral health advice studies, the study investigated contamination effects by simultaneously conducting an individually and a cluster randomised trial.

Methods

Two trials of the same intervention were conducted in parallel. One trial used an individually randomised controlled design. The other trial used a cluster (dentist) randomised controlled design. Apart from the method of allocation, all procedures in each trial were identical.

Participants

Eligible dentists were those who were spending their vocational training year in Scotland. Eligible patients were dentate adults who had already made an appointment for a routine check-up.

Intervention

The intervention was an evidence-based package consisting of a powered toothbrush and oral hygiene advice. The content and delivery of the advice were framed to influence oral hygiene self-efficacy. Dentists demonstrated how to brush teeth using the toothbrush (including giving oral hygiene advice) and then asked the patient to

brush their teeth so that they could see if they were brushing correctly.

Primary outcome measures

- Clinical – the buccal and lingual aspects of all margins/surfaces in the upper left and lower right quadrants were investigated using the Silness and Loe index to record the amount of plaque and the presence or absence of bleeding at the gingival margin on gentle probing.
- Behavioural – self-reported brushing frequency, duration and rinsing behaviour.
- Cognitive – oral hygiene self-efficacy as measured by the mean score of five summed items examining the extent to which patients felt they could perform oral hygiene-related behaviours.

Outcomes for both trials were assessed in the primary care setting by self-completed questionnaires at baseline and at 8 weeks and by clinical examination at baseline on all patients, and at 8 weeks only for patients who required a further examination.

Sample size

Both trials were powered to detect the same size of effect [80% power at the 5% significance level to detect an absolute reduction of 10% in the proportion of sites with bleeding – 40 to 30%; standard deviation (SD) of 35%]. The cluster randomised trial assumed an intracluster correlation coefficient of 0.05.

Randomisation

All participating dentists were first randomly selected to the individually or cluster randomised trial. For the individually randomised trial, patients were randomised to either the routine care group or the enhanced oral hygiene care group according to a practice level block randomisation sequence. For the cluster randomised trial, dentists were randomly allocated to either the routine care group or the enhanced care group and the patients received the intervention that the dentist was allocated.

Blinding

All patients were blinded to outcomes assessment at baseline. Given that the intervention was allocated by cluster in one of the trials, it was not possible to blind the dentist to allocation.

Results

Eighty-four dentists (34 in the patient randomised trial and 50 in the cluster randomised trial) took part in the study. In total, 799 patients consented to take part in the study; 310 were recruited to the patient randomised trial (155 control and 155 intervention) and 489 to the cluster randomised trial (248 control and 241 intervention). The mean (SD) number of patients recruited by each dentist was 8.6 (3.5) and 9.6 (3.1) for the patient and cluster randomised trials, respectively.

There were no substantive differences in patient characteristics at baseline between the control and intervention groups within or between each trial in demographic, clinical, behavioural or cognitive outcomes. The average age was approximately 37 years; just over 75% used a manual toothbrush; approximately one-third of gingival margins were bleeding; and about half of surfaces had plaque.

All primary cognitive and behavioural outcomes showed statistically significant positive changes in both trials (*Table 7*). For the clinical outcomes, only the cluster randomised trial demonstrated a statistically significant reduction in plaque, although all other clinical outcomes showed a similar direction of effect (plaque reduction) in both trials. For three of the primary outcomes, the magnitude of effect was substantially greater in the cluster randomised trial than in the individual randomised trial.

Measures of possible contamination are given in *Table 8*. The use of an electric toothbrush was an integral part of the intervention and we observed at follow-up that 8.2% of the patient randomised trial control group changed to an electric toothbrush compared with 4.7% in the cluster randomised trial. There were similar order of magnitude differences between the control groups for the behavioural measures, but the direction of effect was inconsistent, for example, the patient randomised trial control group sometimes showed less increase in knowledge compared with the cluster randomised trial control group.

Interpretation

This study, conducted throughout Scotland, was the first national study investigating the effectiveness on patient oral hygiene, knowledge and behaviour, of dentists giving oral health advice under normal day-to-day dental surgery conditions. The results of both trials indicated that self-reported patient oral hygiene behaviour and cognitions could be successfully influenced in general dental practice by targeting self-efficacy expectancies towards toothbrushing. However, the patient randomised trial and the cluster randomised trial obtained slightly different results

TABLE 7 *Primary outcomes*

	Patient randomised trial				Cluster randomised trial			
	Control	**Intervention**	**Adjusted[a] difference (95% CI)**	***p*-Value**	**Control**	**Intervention**	**Adjusted[b] difference (95% CI)**	***p*-Value**
Primary cognitive measures								
Self efficacy	*N* = 117	*N* = 115			*N* = 201	*N* = 165		
Baseline – mean (SD)	27.5 (5.3)	27.9 (5.9)			26.7 (5.6)	28.3 (5.1)		
	N = 17	*N* = 115			*N* = 201	*N* = 165		
Follow-up – mean (SD)	26.7 (5.2)	28.3 (5.8)			27.0 (5.3)	28.7 (4.4)		
		1.5	(0.2 to 2.8)	0.024		0.9	(0.01 to 1.8)	0.047
Primary clinical measures								
Percentage surfaces bleeding	*N* = 47	*N* = 36			*N* = 80	*N* = 105		
Baseline – mean (SD)	32.4 (25.6)	27.7 (27.7)			34.8 (28.8)	40.4 (27.6)		
	N = 47	*N* = 36			*N* = 80	*N* = 105		
Follow-up – mean (SD)	21.8 (25.4)	15.5 (16.7)			26.0 (26.3)	21.6 (20.6)		
		–3.5	(–11.8 to 4.8)	0.404		–7.4	(–15.0 to 0.2)	0.057
Percentage surfaces with plaque	*N* = 47	*N* = 37			*N* = 80	*N* = 105		
Baseline – mean (SD)	52.1 (30.4)	52.5 (27.7)			52.8 (32.4)	46.9 (34.2)		
	N = 47	*N* = 37			*N* = 80	*N* = 105		
Follow-up – mean (SD)	31.2 (23.5)	27.6 (19.8)			54.0 (31.1)	31.2 (26.4)		
		–4.5	(–12.7 to 3.7)	0.279		–16.7	(–25.7 to –7.7)	<0.001
			Odds ratio (95% CI)				Odds ratio (95% CI)	
Primary behavioural measures								
Brush teeth at least twice a day	*N* = 116	*N* = 117			*N* = 201	*N* = 165		
Baseline – *n* (%)	83 (71.6)	92 (78.6)			158 (78.6)	129 (78.2)		
	N = 116	*N* = 117			*N* = 201	*N* = 165		
Follow-up – *n* (%)	83 (71.6)	100 (85.5)			158 (78.6)	143 (86.7)		
		2.8	(1.2 to 6.9)	0.021		2.1	(1.2 to 3.6)	0.006
Brush teeth for at least 2 minutes	*N* = 116	*N* = 116			*N* = 201	*N* = 165		
Baseline – *n* (%)	44 (37.9)	32 (27.6)			81 (40.3)	69 (41.8)		
	N = 116	*N* = 116			*N* = 201	*N* = 165		
Follow-up – *n* (%)	51 (44.0)	68 (58.6)			91 (45.3)	117 (70.9)		
		3.3	(1.7 to 6.5)	<0.001		3.0	(1.9 to 4.8)	<0.001
Rinse but do not spit	*N* = 111	*N* = 113			*N* = 199	*N* = 161		
Baseline – *n* (%)	31 (27.9)	23 (20.4)			56 (28.1)	45 (28.0)		
	N = 111	*N* = 113			*N* = 199	*N* = 161		
Follow-up – *n* (%)	40 (36.0)	62 (54.9)			62 (31.2)	105 (65.2)		
		3.5	(1.8 to 6.6)	<0.001		5.3	(3.6 to 7.8)	<0.001

[a] Analysis adjusted for baseline measures.
[b] Analysis adjusted for baseline measures and clustering of patients within vocational dental practitioner.

when it came to the successful translation of these effects to the clinical outcomes. The patient randomised trial showed no statistically significant difference between the intervention and control groups, whereas the cluster randomised trial showed that the intervention group had more statistically significant and larger favourable clinical outcomes than the control group.

Assessment of contamination, however, suggests that this apparent difference was not necessarily a result of contamination. The results in *Tables 7*

TABLE 8 *Patient-reported measures of possible compliance*

Patient reported	Patient randomised trial				Cluster randomised trial			
	Control		Intervention		Control		Intervention	
	n	(%)	*n*	(%)	*n*	(%)	*n*	(%)
Changed from manual to electric toothbrush at follow-up	9	(8.2)	68	(60.2)	9	(4.7)	108	(69.2)
Changed to brushing teeth at least twice a day at follow-up	8	(6.9)	10	(8.5)	9	(4.5)	19	(11.5)
Changed to brushing teeth for at least 2 min at follow-up	15	(12.9)	37	(31.9)	33	(16.4)	59	(35.8)
Changed to rinsing but not spitting at follow-up	15	(13.5)	41	(36.3)	18	(9.0)	64	(39.8)

and *8* provide some evidence that partial contamination did occur, but that it was unlikely that the magnitude of contamination differed substantially between the patient and cluster randomised trials. We hypothesised *a priori* that the effect sizes for the patient randomised trial would be less than those in the cluster randomised trial because of contamination. In the clinical outcomes, such a trend was observed, and, if taken at face value, would indicate that contamination in the patient randomised trial reduced the effect size from –16.7 to –4.5%. However, the behavioural and cognitive measures did not show such a consistent trend. Given that we expected changes in behaviour and cognitions to correlate with changes in clinical outcome, it was suggestive that contamination was not the primary factor for the clinical outcomes to differ.

The measures of possible contamination in *Table 8* indicated that the level of contamination ranged from 6.9 to 13.5% in the patient randomised trial and from 4.7 to 16.4% in the cluster randomised trial. Within each measure, the difference in contamination between the patient and cluster randomised trials was approximately 3–4%, but did not consistently favour either of the trial designs.

In conclusion, this study demonstrated that some form of contamination occurred in both the patient and cluster randomised trials. Hence, although the cluster randomised trial design did not allow control patients to get the evidence-based oral hygiene advice intervention, nevertheless, some patients did acquire some of the knowledge (and electric toothbrush) during the follow-up phase. Further work is currently being undertaken to explore further the possible mechanisms and impact of this contamination on the results of the trials.

Conclusion

When considering all studies targeted at health professionals and reviewed by Grimshaw and colleagues,[14] there is no clear evidence that the effectiveness estimate in trials that do not avoid contamination is biased, compared with trials where contamination is avoided. However, when analysing more homogeneous and higher quality trials, those in which contamination was avoided by cluster randomisation tended to have larger effects. Our analysis of homogeneous trials of smoking cessation interventions targeted at patients found some evidence that trials that avoided contamination by cluster randomisation tended to show greater effects. The trial that directly compared effect estimates following cluster randomisation with effect estimates following individual randomisation also showed larger effects with cluster randomisation, but this could have been due to biased measurement rather than contamination. Overall, the published evidence of bias caused by contamination in trials of educational interventions is thus suggestive but weak. Several other possible reasons (discussed in the section ‘All studies analysis’, p. 11) could explain these findings.

Chapter 3

Expert opinion on situations most likely to lead to contamination and how to avoid it

Although there is little published evidence of the effects of contamination, many educational researchers and trialists have experience in conducting and analysing trials of educational interventions in which contamination was a potential problem. We used a Delphi method to summarise their expert judgements so as to identify situations in which contamination is most likely and how to avoid it. The Delphi method is an exercise in group communication among a panel of geographically dispersed experts.[232] The technique recognises human judgement as legitimate and useful input and allows experts to deal systematically with a complex problem or task and to generate forecasts. It comprises a series of questionnaires sent, either by mail or by computerised system, to a preselected group of experts and interspersed with controlled feedback of opinion.[232] The questionnaires are designed to elicit and develop individual responses to the questions posed and to enable the experts to refine their views as the group's work progresses in accordance with the assigned task. The principle of the Delphi method is to overcome the disadvantages of conventional committee action. According to Fowles,[233] anonymity, controlled feedback and statistical response characterise Delphi. The group interaction in Delphi is anonymous, in the sense that comments, forecasts and the like are not identified as to their originator but are presented to the group in such a way as to suppress any identification.

Our objective for the Delphi study was to elicit expert opinion with regard to the situations in which contamination and contamination avoidance were most likely.

Methods

The Delphi questionnaire was administered over three rounds:

1. Pilot round: the pilot study was distributed to a small group of experts. The aims were to ensure intelligibility of the questions, clear design of the questionnaire and adequate freedom to express opinions.
2. First round: the questionnaire was distributed individually to over 100 medical education experts, including those who completed the pilot round, and to members of the Association for the Study of Medical Education (ASME). This round was used to establish initial opinions of experts to inform the second round.
3. Second round: the questionnaire was distributed to all respondents to the first round. The second-round questions matched those in the first round, but allowed each respondent to view the popularity of each response in the previous round. Additionally, the second-round questionnaire incorporated a ranking exercise, based on first-round opinions, to identify design factors most likely to lead to contamination. Furthermore, in response to subjects' comments on the first round, we added two new sets of questions, as described below.

The questionnaires were disseminated via the Internet. The online questionnaires were constructed by Priority Research Ltd, Sheffield. Responses to the various questions were made via check boxes, for multiple-choice responses, and by free text entry. Respondents were invited to participate via email. Email reminders to potential respondents were also used. By this method, postal delays and data entry from paper copies were avoided, and reminders did not require additional paperwork or postage.

The pilot round of the Delphi was used primarily to assess the suitability and clarity of the questionnaire. The respondents invited to participate in this round were restricted to researchers involved in the project. This round led to re-wording of questions, an extension of the definitions and an increase in the Likert scale from three options ('highly likely', 'neither likely nor unlikely', 'highly unlikely') to include the two intermediate points, 'moderately likely' and 'moderately unlikely'. The option 'no comment/not sure' was added for each item.

The study sample was identified as follows. We compiled a list of experts known to be active in educational research in health. Our personal knowledge was supplemented by searches of researchers active in relevant Cochrane Collaboration and Campbell Collaboration groups. This list consisted of over 100 people to whom the first round of the Delphi questionnaire was emailed. It was also emailed to all members of the ASME. Each email contained an overview of the study, an invitation to participate and an attached covering letter. The letter and email contained hypertext links both to the online Delphi questionnaire and the research project's website. A total of 37 individuals completed the first-round questionnaire and were included in the second round and 27 completed the second round.

Questionnaire design

The first section of the questionnaire elicited respondents' views on the likelihood of contamination in various circumstances. They were asked to consider each factor in isolation, assuming that other factors were the same. To maximise the final response rate by limiting the number of rounds required, the questionnaire was based on our knowledge and on expert opinions expressed in the literature, instead of starting with elicitation of relevant factors from respondents. Respondents were able, however, to suggest other factors that should be considered.

The questions covered the following categories:

1. aim of interventions: transferring knowledge, skills, altering behaviour and changing attitudes
2. complexity of interventions
3. intervention targets: patients, professionals or the general public
4. movement of individuals between treatment arms
5. geographical separation of trial subjects
6. social separation of trial subjects
7. desirability of the interventions
8. transferability of the interventions
9. educating participants against contamination
10. different trial designs
11. situations where cluster randomisation is needed
12. the medium of the intervention, i.e. booklet, CD, etc.
13. social, occupational and geographical proximity of patients
14. social, occupational and geographical proximity of professionals.

For each factor within these categories, respondents indicated how likely they thought contamination was in that situation. In order to analyse these results, a scoring system was used. The possible responses are given below with their scores in brackets. Usage of these scores will be explained later.

- highly likely (5)
- moderately likely (4)
- neither likely nor unlikely (3)
- moderately unlikely (2)
- highly unlikely (1)
- don't know/not applicable (0).

Most of the questions in this first section of the questionnaire required multiple-choice responses. There were two opportunities to make free text responses. One could be used to suggest additional mediums of intervention that were likely to lead to contamination. The other could be used to state specific situations in which cluster randomisation should be used to avoid contamination. Additionally, we asked respondents to make free text responses to the following specific questions:

1. How do you think researchers can protect against contamination in controlled trials of educational interventions?
2. Under which circumstances would you employ a particular study design specifically to avoid contamination?
3. How would you know if contamination were occurring in a study?

The first-round questionnaire also allowed free text responses to be made in order to comment generally on the questionnaire, improve the clarity of the questionnaire, modify the definitions provided or list any issues regarding contamination that were not covered by the questionnaire.

In order to compare different types of factors with regard to their likelihood of contamination, a scoring system was used. The number of subjects who gave each response to each question was multiplied by that response's score to produce a mean score for each question. The resulting values were summed and the results divided by the number of responses. For example, if two respondents give the answer 'highly likely' (score 5) and one respondent gives the answer 'neither likely nor unlikely' (score 3), the mean score would be $[(2 \times 5) + (1 \times 3)]/3 = 4.33$. Using this method, a high score indicates a high likelihood of

contamination whereas a low score may indicate either a low likelihood of contamination or a high level of uncertainty [i.e. 'don't know' responses (score 0)].

The aim of the second round was to elicit a consensus of the first-round responses in terms of their likelihood to lead to contamination. The second-round questionnaire was largely identical to the first-round questionnaire, but provided additional information about first-round responses. Each respondent was provided with the same questions (minus the free text questions), and was asked to consider their response in the light of the other experts' opinions. For each question they were presented with a percentage breakdown of the response by all experts and reminded of their previous response.

The second-round questionnaire included a new section. Respondents were presented with two examples of an intervention which, in accordance with first-round responses, was highly likely to lead to contamination. Each intervention was assumed to be:

- highly desirable
- made up from simple elements that stand alone
- aimed at transferring new knowledge and
- easy to transfer.

One example, of an intervention aimed at patients, was a workout video to aid weight loss. The other example, of an intervention aimed at patients, was a CD-based computer program for managing diabetes in primary care using National Institute for Health and Clinical Excellence guidelines. For each example, respondents were presented with a ranking of the eight factors deemed generally most likely to lead to contamination, according to first-round responses. Respondents were asked either to agree with this ranking or to specify a new ranking.

Examples of the first- and second-round questionnaires are provided in Appendices 1 and 2, respectively. The definitions provided to respondents are given in Appendix 3.

Results

Scores for the second round were very similar to those for the first round ($R^2 = 0.98$). *Table 9* shows the ranking within each question, in the order in which it was asked. *Table 10* re-ranks the items under three general headings so as to facilitate comparisons. Our scoring method entails that items in the top 27 ranks were, on average, considered more likely than unlikely to lead to contamination. A detailed breakdown of responses to the first and second rounds is available from the first author.

The results show that, in the opinion of respondents, contamination was most likely to occur in trials conducted in settings where subjects worked, lived or interacted closely together and where interventions were desirable, simple or easily transferable or were aimed at increasing knowledge. Contamination was least likely where subjects were socially or physically separate, where interventions were complex or aimed at changing behaviours and when cluster randomisation was used. Contamination was more likely with interventions aimed at health professionals than with interventions aimed at patients. It was more likely with interventions based on broadcast media, audiovisuals or written information, was least likely with computer-based reminders, and was intermediate with training interventions.

For each of the two scenarios in which contamination was stated in advance to be likely, additional factors more or less likely to cause contamination were ranked as follows. For the patient education scenario – comprising an exercise workout video – the location and movement of patients were ranked as more likely to cause contamination than professional-related factors (*Table 11*). Contamination was more likely if intervention and control patients lived in the same residence than if they shared social networks or geographical location.

For the professional education scenario – CD-based diabetes guidelines – the location and movement of health professionals were ranked as more likely to cause contamination than patient-related factors (*Table 12*). Working in the same workplace was more likely to cause contamination than moving between intervention and control arms or sharing a place of employment.

Free text responses

When asked, 'Are there any situations which usually require cluster randomisation to avoid contamination?', their responses were as follows:

- When the intervention is delivered by a health professional (nine responses).
- Individually randomised trial where the participants are likely to meet (four responses).
- General public or mass media interventions (four responses).

TABLE 9 *Likelihood[a] of contamination for different intervention characteristics, contexts or trial designs, ranked within each question*

Intervention characteristic, context or trial design	Score	Intervention characteristic, context or trial design	Score
Transfer of new knowledge	3.89	Individually randomised parallel group trial	3.59
Changing attitudes	2.63	Before/after comparisons	2.37
Transfer of new skills	2.59	Repeated time series	2.37
Altering behaviours	2.52	Cluster randomised trial	2.08
Simple – elements stand alone	3.78	Broadcast media (TV, radio)	4.44
Modest complexity – some elements stand alone	3.33	Audiovisual, e.g. video, CD-ROM	3.93
Complex – multiple interdependent parts	2.44	Written information, e.g. booklet	3.89
Targeted at health professionals	4.07	Training programme (multiple attendances)	2.41
Targeted at patients	3.26	Training event (single attendance)	2.38
Targeted at the general population	2.89	Use of specialist resources (models, simulations)	2.04
Staff move from intervention arm to control arm	4.59	Computer reminders (popups on computer)	1.93
Patients move intervention arm to control arm	3.96	Patients living in the same residence	4.93
Subjects from the same geographical site	4.22	Patients with shared social networks	4.19
Subjects from nearby sites in same community	3.52	Patients' healthcare comes from same practice	3.89
Subjects geographically separated	1.74	Patients living in same geographical locality	3.74
Subjects from same social networks	4.78	Health professionals in the same workplace team	4.78
Subjects from overlapping social networks	3.96	Health professionals sharing a place of employment	4.37
Subjects unlikely to have social networks in common	1.52	Health professionals in the same clinical directorate or equivalent	3.85
High desirability	4.07	Health professionals sharing an employer	3.33
Medium desirability	3.04	Geographical/social separation of subjects	4.00
Little desirability	1.85	Education of subjects to avoid transfer of intervention	3.37
Easy to transfer	3.78	Restriction on medium of intervention	3.19
Difficult to transfer	1.56	Avoid allocation of subjects to less desirable arm	2.48
No efforts to educate participants against contamination	3.67		
Participants educated to avoid contamination	2.26		

[a] Higher score indicates greater likelihood. Scores from Delphi second round.

- 'If the intervention is easily transferred and desirable, and the physical/social/trial design conditions make that possible/likely' (one response).
- 'Cluster randomisation is usually required for logistical reasons not because of fears of contamination' (one response).

When asked how respondents thought researchers can protect against contamination in controlled educational intervention studies, their responses included the following:

- Cluster randomisation (or equivalent, i.e. geographical/social separation) (18 responses).
- Educate against contamination (six responses).
- Don't educate against contamination (two responses).
- Provide clear information on nature and purpose of study and on the intervention and controls (four responses).
- Attempt to measure contamination (three responses).
- You cannot protect against contamination/can't do much (three responses).
- Increase sample size to compensate (two responses).
- Use less transferable interventions (two responses).
- '(a) Expect it, (b) design it out as far as possible – using the usual methods – as much blindness as possible' (one response).

Asked under which circumstances they would employ a particular study design specifically to avoid contamination, several suggested a cluster randomised design for the following situations:

- When educating health professionals (five responses).
- When the intervention cannot be delivered at individual level (two responses).
- When the intervention occurs at population levels (two responses).
- When individuals are likely to discuss intervention (two responses).

TABLE 10 *Likelihood[a] of contamination ranked within intervention characteristics, contexts or trial designs, and each question*

Characteristic	Score	Characteristic	Score
Content, medium or aim of intervention		Patients with shared social networks	4.19
Media output (TV, radio)	4.44	Targeted at health professionals	4.07
High desirability	4.07	Patients move from intervention arm to control arm	3.96
Audiovisual, e.g. video, CD-ROM	3.93	Subjects from overlapping social networks	3.96
Transfer of new knowledge	3.89	Patients' healthcare comes from same practice	3.89
Written information, e.g. booklet	3.89	Health professionals in same clinical directorate or equivalent	3.85
Simple elements stand alone	3.78		
Elements easy to transfer	3.78	Patients living in same geographical locality	3.74
Modest complexity – some elements stand alone	3.33	Subjects from nearby sites in the same community	3.52
Medium desirability	3.04	Health professionals sharing an employer	3.33
Aimed at changing attitudes	2.63	Targeted at patients	3.26
Aimed at transfer of skills	2.59	Targeted at the general population	2.89
Aimed at altering behaviours	2.52	Geographically separated	1.74
Complex – multiple interdependent parts	2.44	Subjects unlikely to have social networks in common	1.52
Training programme (multiple attendances)	2.41		
Training event (single attendance)	2.38	**Design or conduct of trial**	
Use of specialist resources (models, simulations)	2.04	Geographical/social separation of subjects	4.00
Computer reminders (popups on computer)	1.93	No efforts to educate participants against contamination	3.67
Little desirability	1.85		
Elements difficult to transfer	1.56	Individually randomised parallel group trial	3.59
		Restriction on medium of intervention	3.19
Context or setting of intervention		Avoid allocation of subjects to less desirable arm	2.48
Patients living in the same residence	4.93	Before/after comparisons	2.37
Subjects from same social networks	4.78	Repeated time series	2.37
Health professionals in the same workplace team	4.78	Education of participants to avoid transfer of intervention	3.37
Staff move from intervention arm to control arm	4.59		
Health professionals sharing a place of employment	4.37	Participants educated to avoid contamination	2.26
Subjects from the same geographical site	4.22	Cluster randomised trial	2.08

[a] Higher score indicates greater likelihood. Scores from Delphi second round.

TABLE 11 *Likelihood of contamination with a patient-targeted intervention*

Patient target	Revised rank	SD	Original rank	Change
Patients receiving the intervention live in the same residence as patients receiving the control	1.6	0.69	2	0.4
Patients receiving the intervention are from the same social network as patients receiving the control	3.1	1.15	3	–0.1
Patients move from intervention arm to control arm	3.9	2.57	7	3.1
Health professionals administering the intervention are in the same workplace team as health professionals administering the control	4.3	2.16	1	–3.3
Staff move from intervention arm to control arm	5.5	1.68	4	–1.5
Patients receiving the intervention are from the same geographical site as patients receiving the control	5.5	1.76	6	0.5
No effort is made to educate participants against contamination	6.1	2.21	8	1.9
Health professionals in the intervention arm share a place of employment with health professionals in the control arm	6.2	1.19	5	–1.2

- When there is a strong likelihood of contamination (two responses).

Several wrote that the main reason for doing cluster randomised trials was that the unit of inference was a professional or workplace and so it was logical to randomise at this level. Other designs suggested included the following:

- Use an interrupted time series design to collect control data before the intervention (two responses).

TABLE 12 *Likelihood of contamination with a professional-targeted intervention*

Patient target	Revised rank	SD	Original rank	Change
Health professionals receiving the intervention are in the same workplace team as health professionals receiving the control	1.8	1.57	1	–0.8
Staff move from intervention arm to control arm	3.2	1.70	4	0.8
Patients in the intervention group live in the same residence as patients in the control group	3.6	1.80	2	–1.6
Health professionals receiving the intervention share a place of employment with health professionals receiving the control	4.4	1.41	5	0.6
Patients in the intervention group are from the same social network as patients in the control group	4.8	1.78	3	–1.8
Patients move from intervention arm to control arm	5.4	1.93	7	1.6
No effort is made to educate participants against contamination	6.2	2.40	8	1.8
Patients in the intervention group are from the same geographical site as patients in the control group	6.6	1.08	6	–0.6

- Find a balance between contamination sample size and cost (two responses).

Asked how they would know if contamination were occurring in a study, their responses included the following:

- Measure it by questioning control group (13 responses).
- Look for delayed parallel improvement in control arm (two responses).
- 'Comparison with non-study samples' (one response)
- 'Very difficult without totally independent control groups' (one response).

Asked how to improve the questionnaire, several respondents wrote that they would have preferred more specific examples to have been provided or would have preferred a more specific context to be given for some of the questions. Several commented that holding the assumption of 'other things being equal' while considering each factor in isolation did not make sense because the risk of contamination depended on combinations of factors.

Discussion

These results reflect the opinions of experts in educational research on study designs to avoid contamination, and give some guidance to those designing trials of educational studies. There was a high degree of consensus among these experts. Many of the findings have face validity and may even seem obvious. However, they are worth considering when contamination is possible. They suggest that researchers should consider using a cluster randomised design, or take stringent efforts to avoid and measure contamination, in the following settings where contamination is most likely:

- any groups working in the same healthcare setting, or living in the same household
- any group with significant social network linkage (e.g. GPs in the same community)
- media-based interventions.

Interventions with less risk of contamination include multifaceted interventions and those with a significant behavioural component that cannot easily be transferred from one participant to another. Experts were sceptical about the value of specific preventive instruction in reducing contamination, particularly in patient-based interventions. The consensus was that, for cluster randomised designs, contamination was more likely where both the intervention and control health professionals were part of the same workplace team. In these situations, early contamination of all, or a large proportion of, the cluster may occur. Rapid contamination of a large proportion of a cluster might also occur when a health professional moves from the intervention to the control group.

In conclusion, there may be practical or logical reasons, unrelated to contamination, to use cluster or individually randomised trials designs to evaluate educational interventions. However, the findings of this study give some guidance on the types of studies for which cluster randomised designs may be best. As this study shows, there are many other ways to avoid contamination if it is a risk.

Chapter 4

Contamination simulation

In order to examine the influence of contamination upon a trial, simulation programs were produced to reflect the effect of the intervention and contamination upon the experimental and control groups.

Aim

The aim of our simulation program was to compare the bias of cluster randomised trials relative to the bias of individually randomised trials. Contamination was modelled in both individual and cluster randomised trials under different scenarios.

Methods

Bias of the effectiveness estimate is a function of three components of the contamination process:

1. The proportion of the control group exposed to the intervention.
   (a) The proportion exposed may or may not be a function of subjects' baseline 'need' for the intervention, that is, according to the baseline value of the outcome measure.
2. The timing of the exposure.
   (a) In cluster randomised trials, individual control subjects may be exposed either at the time of the cluster's first exposure, or later.
3. The intensity of exposure.

These variables all vary randomly between control subjects. Any of these components could also be a function of the desirability or transportability of the intervention, which do not need to be modelled separately.

For example, consider a trial of training GPs to use a clinical practice guideline. All or some of the control group GPs could inadvertently receive the guideline. Those who do could be those who are less or more informed. They could receive the guideline but not the training. In a cluster randomised trial, control practices could start receiving the guideline after intervention practices, and individual GPs in these practices could receive the guideline together then, or later.

The contamination simulation was written in SAS version 8. Whereas the basis of the contamination model remained consistent, two main versions of the simulation were produced: one to implement an individually randomised design and the other to implement a cluster randomised design. Both designs of simulation were based on a two-armed trial consisting of an intervention which is assumed to be effective and a no-intervention control. The outcomes in the arms of the trial are modelled as the 'education' level of trial subjects so that intervention leads to increased levels of education.

Model for simulation

The individually randomised trial simulation implemented the following mathematical models for control and intervention:

$$y_{i0} = x + \epsilon_{i0}$$
$$y_{iT} = x + \kappa\tau_{i0} + \epsilon_{it}$$

where $x \sim N(40,500)$, y_{iT} is the education level for participant i at time T, the end of the trial, y_{i0} is the baseline education level of participant i, τ is the intervention effect and κ is a proportion between 0 and 1 and $\epsilon_{it} \sim N\,(0,125)$ are random errors.

This model induces a positive correlation between the baseline and follow-up education levels.

For those individuals in the intervention group κ is fixed at one, for those in the control, with no contamination, κ is fixed at zero, and for those contaminated, κ takes on a value above zero, possibly equal to one.

Cluster randomised trials

The cluster randomised model implemented a similar mathematical model to that of the individual model, but includes some additional terms. The control and intervention models are of the following form:

$$y_{i0} = x + \varepsilon^{i}_{ij0} + \varepsilon^{j}_{ij0}$$
$$y_{iT} = x = \kappa\tau + \beta_c\tau + \varepsilon^{i}_{ijT} + \varepsilon^{j}_{ijT}$$

where $x \sim N(40,500)$, y_{iT} is the education level for participant i at time T, the end of the trial, y_{i0} is the baseline education level of participant i, τ is the intervention effect, κ is a proportion between 0 and 1, $\varepsilon_{ij}^{i} \sim N(0, \sigma_W{}^2)$ are individual random errors, $\varepsilon_{ij}^{j} \sim N(0, \sigma_B{}^2)$ are cluster random errors, $\sigma_B{}^2$ and $\sigma_W{}^2$ are the between-group and within-group variances respectively, $\sigma_B{}^2 + \sigma_W{}^2 = 125$ and $\beta_C \sim N(0, 0.1)$ is the cluster effect variation.

For those individuals in the intervention group κ is fixed at one, for those in the control, with no contamination, κ is fixed at zero and for those contaminated, κ takes on a value above zero, possibly equal to one; β_C is zero in the control model.

The intervention effect, τ, was arbitrarily set at 100. For those individuals in the intervention group κ is fixed at one, for those in the control, with no contamination, κ is fixed at zero and for those contaminated, κ takes on a value above zero, possibly equal to one.

The above model is used for all of our simulations. However, by varying the contamination parameter κ and its dependence on other parameters, it is possible to model contamination in different ways.

Modelling time to contamination

Time to contamination in individually randomised trials

The time of contamination was modelled using a Weibull distribution. It is widely used for modelling time to event data and is an obvious choice of model for time to contamination.

The random variable Y follows a Weibull probability model with parameters $\alpha > 0$ and $\beta > 0$ denoted $Y \sim \text{Weibull}(\alpha, \beta)$, when Y has a power hazard of the form $h(y) = (\alpha/\beta^{\alpha})y^{\alpha-1}$.

By modelling time to contamination in this way, individuals can be labelled as contaminated or not. For a trial of length time T, an individual i is only contaminated if $t_i < T$.

Since generating random variates from a Weibull distribution is not a standard function in SAS, variates from an exponential distribution were generated and then transformed into Weibull variates.

The transformation of a standard exponential random variate x into a Weibull random variate y with shape parameter α and scale parameter β can be accomplished via the expression

$$y = \beta x^{1/\alpha}$$

where α and β are as defined above.

The hazard function, $h(t)$, of the Weibull distribution provides the instantaneous risk of contamination occurring at time point t given that contamination has not occurred by that stage. Mathematically it is defined by

$h(t) = \alpha t^{\alpha-1}/\beta^{\alpha}$, where α is the shape parameter and β the scale parameter of the Weibull distribution, given above. By altering α, the nature of the risk of contamination with time can be varied. The parameter β controls the overall, or average, level of risk.

For $\alpha = 1$, $\lambda = h(t) = 1/\beta$; this provides a constant hazard, that is, a constant risk of contamination over time. To ensure a hazard function between zero and one, β needs to be >1.

For $\alpha = 2$, $\lambda = h(t) = 2t/\beta^2$; this provides a hazard which increases linearly over time. To ensure a hazard function between zero and one, β needs to be $>\sqrt{2}$.

For $\alpha = 3$, $\lambda = h(t) = 3t^2/\beta^3$; this provides a hazard which increases quadratically over time. To ensure a hazard function between zero and one, β needs to be $>\sqrt[3]{3}$.

In order to relate the risk of contamination to an individual's baseline education level y_{i0}, the parameter β, and hence the hazard function, varied across individuals according to y_{i0} via the function

$$\beta_i = f(y_{i0}) = \sqrt[\alpha]{\alpha}\exp[k - b(y_{i0} - \mu)]$$

where $\sqrt[\alpha]{\alpha}$ is the lower limit for β, μ is the mean education level at baseline, b determines the variation in β and the influence of the baseline education level y_{i0}, k determines the average level of β and the exponent prevents negative values of β.

Therefore, the risk of contamination will depend on the value of k_j (a lower value of producing higher levels of risk) and the dependence of contamination on ignorance will be determined by b (a higher value providing a greater dependence and a value of zero producing no relationship between the risk of contamination and baseline education level).

Time to contamination in cluster randomised trials

Contamination of clusters uses the same model to simulate contamination time; the Weibull random variates are used to provide times to contamination within the trial period. A Weibull variate is generated for each cluster but, in addition, a Weibull variate is generated for each subject within clusters.

In this way, two independent contamination time values, t_c and t_i, for the cluster c and individual i are generated according to the Weibull distributions detailed earlier. The cluster contamination time depends only on the parameter k_c (since no baseline education value is generated for a cluster), via the parameter β, that is, $\beta = \sqrt[\alpha]{\alpha}\exp(k_c)$.

Two versions of cluster contamination were considered. First, individuals were considered independently once the cluster of which they were members was contaminated. In this case, an individual i within cluster c is contaminated if $t_c + t_i < T$.

In the second version of cluster contamination, all individuals within the contaminated cluster were considered to be contaminated at the time when contamination first enters the cluster. The time at which cluster contamination occurs is generated in the same way as for other cluster models. However, no use is made of the individual contamination times within the cluster. Therefore, for a trial of length T, all individuals within cluster c are contaminated if $t_c < T$, otherwise no individuals in cluster c are contaminated.

Contamination levels

Two options for the amount of contamination to occur were considered. One option was to fix κ to a constant proportion for all individuals contaminated. The second option was to make κ dependent on the time at which contamination occurred.

Using the same Weibull contamination times as in the above description, the contamination time can be converted to a simple indicator without directly influencing the extent to which contamination takes place.

Under the option of a constant contamination effect, those patients who were contaminated during the period received a constant intervention effect, for example, $\kappa = 0.5$, $\kappa = 1$, that is, 50 or 100% of the intervention's true effect, respectively.

This constant effect simulation reflects the situation where contamination has an instantaneous effect, but the strength of that effect may be limited. For example, a constant contamination effect of 100% would be experienced by a control group patient who received and acted upon a reminder intervention for immunisation. Alternatively, a 50% effect might be achieved by a control group professional receiving instruction on a particular procedure from an intervention colleague. Awareness of the technique would produce an immediate effect but, since training in the procedure was informal and secondhand, an imperfect technique results, reducing the effect.

In the second option, the proportion of contamination was made dependent on the time at which contamination occurs. Given a trial duration of T and contamination occurring at time t, $\kappa = (T - t)/T$, that is, the proportion of contamination is equal to the proportion of time remaining in the study.

This time-dependent contamination illustrates a situation where the intervention has a gradual effect on the respondent, where the full effect requires exposure for the entire trial duration. For example, dietary advice for weight loss will only have an effect where the subject has time to implement the advice and change their eating habits: contamination at the beginning of the trial might enable the full effect to be received by members of the control group, but late contamination would have little or no effect.

Where the effect of the cluster contamination is dependent on time, the proportion κ of the intervention effect used to contaminate individual i within cluster c will be $[T - (t_c + t_i)]/T$, where t_c and t_i are the cluster and individual contamination times, respectively. This model mimics an intervention similar to the weight loss example given for individual time-dependent simulations; however, in order for an individual to be contaminated, their cluster must first be contaminated, which takes time t_c before the individual can be contaminated.

Implementation

The input parameters for the SAS simulation were as follows:

- Cycles: the number of trial simulations to perform

- k_c: as detailed above, for the clusters (cluster simulation program only)
- b: the parameter detailed above for the individuals
- ICC: the intracluster correlation coefficient (cluster simulation program only)
- NumClusts: number of clusters in the trial (cluster simulation program only)
- NumperClust: number of individuals per cluster (cluster simulation program only)
- Trialsize: total number of subjects in the trial (individual randomisation simulation program only)
- τ: true intervention effect
- k: the parameter k detailed above
- PropEff: proportion of the effect to be used for contamination (fixed effect/all or nothing simulation programs only).

Other parameters used in the model were either fixed or calculated from other parameters. For example, the total variance was set to 125, and the variance within and between groups was calculated from this value using the ICC.

The input parameters b, k_c and k were selected in order to provide a variety of contamination effects whilst minimising the number of simulation runs required. Three values of k_c were used, 0.01, 0.05 and 0.1, four values of b were used, 0 (representing contamination independent of individual's need), 0.01, 0.3 and 2, and three values of k were also used, 1, 5 and 20. Therefore, leaving all other inputs fixed, 36 simulations were produced.

Since 36 sets of simulations were produced for each size, ICC, proportion of effect and type of contamination dependence trial, the number of cycles chosen for initial analysis was chosen to be 100. This number was sufficiently large to enable results to be reliable, but necessarily small enough to obtain results within a reasonable timescale. A single set of 36 simulations for time-dependent, fixed and all or nothing contamination took approximately 2–3 hours for 100 cycles, whereas 1000 cycles took between 16 and 24 hours to run.

For each type of contamination dependence, a set of 36 cluster simulations with 100 cycles was performed for ICC = 0.2 and 0.01. For both of these ICC values, sets of simulations were performed for trials of the following dimensions:

- 20 clusters of five subjects
- five clusters of 20 subjects
- 10 clusters of 10 subjects.

Corresponding individual simulations with trial size 100, using the same values of k and b, were produced.

Larger scale simulations were also produced with the same ICCs. These simulations consisted of 10 clusters of 100 individuals and were compared with the results from an individually randomised simulation with trial size 1000.

The outcome of interest for each simulation was the bias through contamination and a comparison between the individual randomised design and the cluster randomised design. The bias was defined as the difference between the true intervention effect and the apparent intervention effect, that is, the mean difference between intervention and control at follow-up.

Parameters

For each of the three types of hazard functions (α = 1, 2, 3), the values of output by the simulation are the following:

- proportion contaminated (%)
- bias due to contamination (mean, SD, upper and lower 95% CI)
- follow-up education level (overall, contaminated subjects only and not contaminated subjects only, each of these at baseline and follow-up)
- contamination time (individual, cluster and overall, overall for those contaminated, overall for those not contaminated).

The parameter b controls the variance of the contamination times for individuals once awareness of the intervention is possible. As b increases, the contamination times of individuals with greatest need decrease, and the contamination times of individuals with least need increase. An illustration of the effect of k and b is given in *Figure 29*.

The parameter k_c causes a positive shift in the overall contamination times. Therefore, as k_c increases, there is no change to the spread of contamination times, but all contamination times are increased by a fixed constant ($\geq$0). Conversion of the parameter k_c into a more meaningful time elapsed and percentage of trial elapsed value is shown in *Table 13*.

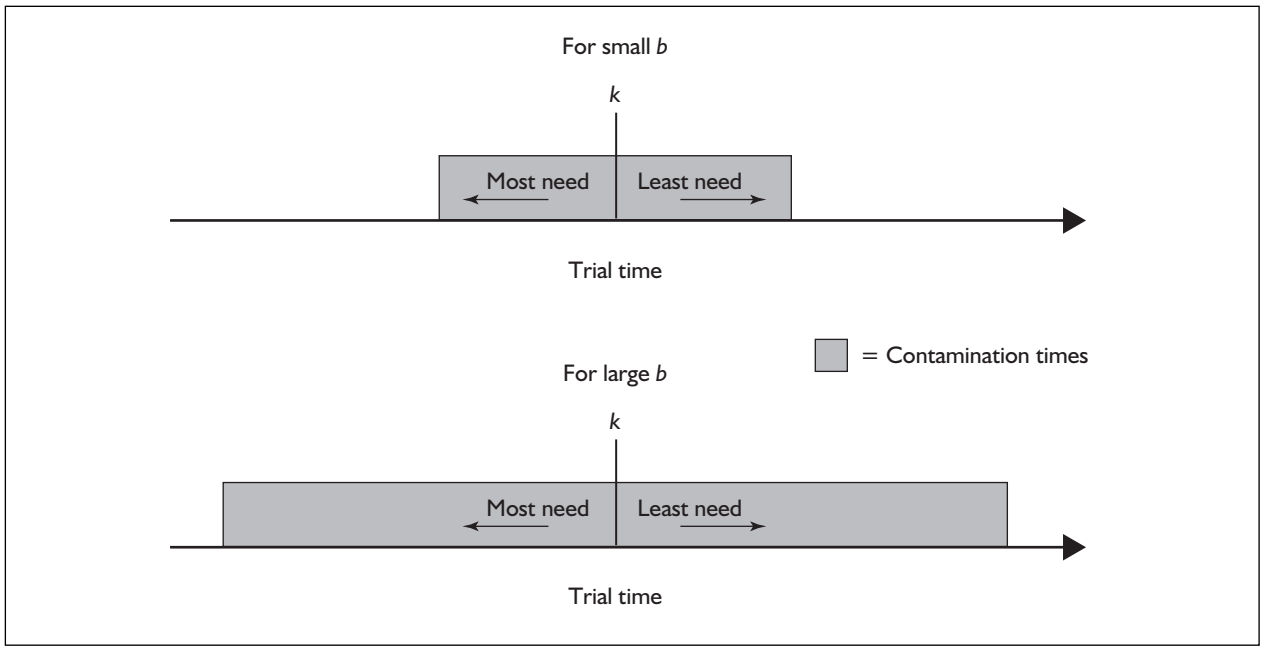

FIGURE 29 *Effect on time of contamination of altering* b *(education dependence of time of contamination) while holding* k *(mean time of contamination) constant*

TABLE 13 *Conversion of parameter* k_c *into expected time and percentage of trial elapsed*

	Exponential $\alpha = 1$		Linear Weibull $\alpha = 2$		Linear Weibull $\alpha = 3$	
k_c	Time (approx.)	% of trial	Time (approx.)	% of trial	Time (approx.)	% of trial
0.01	1.49	12	1.86	16	1.92	16
0.05	4.45	37	6.95	58	8.1	68
0.1	6.65	55	11.95	99	12.84	100

Results

The results from all simulations for all outputs are available from the first author. The figures provided are the bias through contamination for the individually randomised trial and the difference in bias through a cluster randomised trial in the same situation. A negative difference implies greater bias for the cluster randomised trial. This difference is also expressed as a percentage of the individually randomised simulation's effect. The percentage decrease in contamination time of the individual simulation compared with the cluster is also included, together with the baseline education level for both contaminated and not contaminated subjects.

b = 0 (no influence of baseline education on risk of contamination) all cluster members contaminated

When contamination of a cluster affected all cluster members there was a greater bias for cluster randomised trials compared to individually randomised in almost all scenarios. For a constant contamination level (i.e. $\kappa = 1$), as given in *Table 14*, the bias in both forms of trials was around 100 in many instances. For a high level of k (i.e. a long average individual contamination time) bias dropped off to zero for the individually randomised trial, but was frequently very high in the cluster randomised trial.

For time-dependent contamination level (i.e. κ depending on time), as given in *Table 15*, the results were not clear cut. There appeared to be an advantage in cluster randomisation only when $k = 1$ (i.e. a short average individual contamination time). This seemed particularly to be the case with a non-constant hazard of contamination, that is, risk of contamination increasing with time.

TABLE 14 *Bias if contamination independence* b = *0, all cluster members contaminated with fixed 100% contamination effect*

Number of clusters, number of members per cluster and ICC	***k***	**α = 1**				**α = 2**				**α = 3**			
		Individual simulation bias	**Difference in bias between individual and cluster (%)**			**Individual simulation bias**	**Difference in bias between individual and cluster (%)**			**Individual simulation bias**	**Difference in bias between individual and cluster (%)**		
			Cluster delay 1.49 (12%)	**Cluster delay 4.45 (37%)**	**Cluster delay 6.65 (55%)**		**Cluster delay 1.86 (16%)**	**Cluster delay 6.95 (58%)**	**Cluster delay 11.95 (99%)**		**Cluster delay 1.92 (16%)**	**Cluster delay 8.1 (68%)**	**Cluster delay 12.84 (100%)**
10 clusters with	1	100	0	21	80	100	0	28	98	100	0	25	100
10 per cluster and			(0)	(21)	(80)		(0)	(28)	(98)		(0)	(25)	(100)
ICC = 0.2	5	8	–92	–73	–14	2	–98	–74	–1	0	–100	–78	0
			–(1146)	–(913)	–(177)		–(4889)	–(3691)	–(59)		(NA)	(NA)	(NA)
	20	0	–100	–78	–20	0	–100	–71	–2	0	–100	–73	0
			(NA)	(NA)	(NA)		(NA)	(NA)	(NA)		(NA)	(NA)	(NA)
5 clusters with	1	100	0	20	80	100	0	27	98	100	0	24	99
20 per cluster and			(0)	(20)	(80)		(0)	(27)	(98)		(0)	(24)	(99)
ICC = 0.2	5	8	–92	–69	–9	2	–98	–68	0	0	–100	–72	–1
			–(1144)	–(864)	–(115)		–(4887)	–(3378)	(1)		(NA)	(NA)	(NA)
	20	0	–100	–81	–21	0	–100	–73	–2	0	–100	–76	–1
			(NA)	(NA)	(NA)		(NA)	(NA)	(NA)		(NA)	(NA)	(NA)
20 clusters with	1	100	–1	19	81	100	–1	26	98	100	–1	22	100
5 per cluster and			–(1)	(19)	(81)		–(1)	(26)	(98)		–(1)	(22)	(100)
ICC = 0.2	5	8	–92	–73	–11	2	–98	–72	0	0	–100	–76	–1
			–(1155)	–(906)	–(142)		–(4919)	–(3590)	–(15)		(NA)	(NA)	(NA)
	20	0	–100	–81	–18	0	–100	–74	–2	0	–100	–76	0
			(NA)	(NA)	(NA)		(NA)	(NA)	(NA)		(NA)	(NA)	(NA)
10 clusters with	1	100	1	18	82	100	0	25	98	100	0	22	100
10 per cluster and			(1)	(18)	(82)		(0)	(25)	(98)		(0)	(22)	(100)
ICC = 0.01	5	8	–92	–72	–12	2	–98	–72	0	0	–100	–76	0
			–(1151)	–(895)	–(145)		–(4905)	–(3586)	–(19)		(NA)	(NA)	(NA)
	20	0	–100	–79	–20	0	–100	–72	–3	0	–100	–75	–1
			(NA)	(NA)	(NA)		(NA)	(NA)	(NA)		(NA)	(NA)	(NA)

continued

TABLE 14 *Bias if contamination independence* b = *0, all cluster members contaminated with fixed 100% contamination effect (cont'd)*

Number of clusters, number of members per cluster and ICC	*k*	α = 1				α = 2				α = 3			
		Individual simulation bias	Difference in bias between individual and cluster (%)			Individual simulation bias	Difference in bias between individual and cluster (%)			Individual simulation bias	Difference in bias between individual and cluster (%)		
			Cluster delay 1.49 (12%)	Cluster delay 4.45 (37%)	Cluster delay 6.65 (55%)		Cluster delay 1.86 (16%)	Cluster delay 6.95 (58%)	Cluster delay 11.95 (99%)		Cluster delay 1.92 (16%)	Cluster delay 8.1 (68%)	Cluster delay 12.84 (100%)
5 clusters with 20 per cluster and ICC = 0.01	1	100	0 (0)	20 (20)	80 (80)	100	0 (0)	29 (29)	98 (98)	100	0 (0)	26 (26)	100 (100)
	5	8	–92 –(1147)	–73 –(913)	–14 –(169)	2	–98 –(4888)	–72 –(3601)	–1 –(57)	0	–100 (NA)	–77 (NA)	–1 (NA)
	20	0	–100 (NA)	–79 (NA)	–21 (NA)	0	–100 (NA)	–74 (NA)	–3 (NA)	0	–100 (NA)	–77 (NA)	–1 (NA)
20 clusters with 5 per cluster and ICC = 0.01	1	100	0 (0)	20 (20)	79 (79)	100	0 (0)	27 (27)	98 (98)	100	0 (0)	24 (24)	100 (100)
	5	8	–92 –(1148)	–72 –(894)	–11 –(137)	2	–98 –(4897)	–70 –(3514)	0 (5)	0	–100 (NA)	–75 (NA)	0 (NA)
	20	0	–100 (NA)	–80 (NA)	–20 (NA)	0	–100 (NA)	–73 (NA)	–3 (NA)	0	–100 (NA)	–76 (NA)	0 (NA)
10 clusters with 100 per cluster and ICC = 0.2	1	99	–2 –(2)	20 (20)	79 (80)	100	0 (0)	27 (27)	98 (98)	100	0 (0)	24 (24)	100 (100)
	5	8	–91 –(1115)	–70 –(854)	–13 –(159)	0	–99 –(24810)	–70 –(17590)	–3 –(664)	0	–100 (NA)	–74 (NA)	–1 (NA)
	20	0	–100 (NA)	–81 (NA)	–23 (NA)	0	–100 (NA)	–74 (NA)	–3 (NA)	0	–100 (NA)	–77 (NA)	–1 (NA)
10 clusters with 100 per cluster and ICC = 0.01	1	98	–2 –(2)	19 (19)	76 (78)	100	0 (0)	26 (26)	98 (98)	100	0 (0)	24 (24)	100 (100)
	5	8	–92 –(1107)	–73 –(879)	–12 –(150)	0	–100 –(24931)	–74 –(18568)	–2 –(468)	0	–100 (NA)	–77 (NA)	0 (NA)
	20	0	–100 (NA)	–80 (NA)	–19 (NA)	0	–100 (NA)	–72 (NA)	–3 (NA)	0	–100 (NA)	–75 (NA)	0 (NA)

NA, not applicable.

TABLE 15 *Bias if contamination independence* b = *0, all cluster members contaminated with time dependent contamination effect*

Number of clusters, number of members per cluster and ICC	***k***	**α = 1**				**α = 2**				**α = 3**			
		Individual simulation bias	**Difference in bias between individual and cluster (%)**			**Individual simulation bias**	**Difference in bias between individual and cluster (%)**			**Individual simulation bias**	**Difference in bias between individual and cluster (%)**		
			Cluster delay 1.49 (12%)	**Cluster delay 4.45 (37%)**	**Cluster delay 6.65 (55%)**		**Cluster delay 1.86 (16%)**	**Cluster delay 6.95 (58%)**	**Cluster delay 11.95 (99%)**		**Cluster delay 1.92 (16%)**	**Cluster delay 8.1 (68%)**	**Cluster delay 12.84 (100%)**
10 clusters with 10 per cluster and ICC = 0.2	1	82	–6 –(7)	30 (37)	70 (85)	74	–10 –(14)	43 (58)	73 (99)	72	–12 –(16)	47 (65)	72 (100)
	5	4	–84 –(2248)	–46 –(1228)	–6 –(167)	1	–84 –(12864)	–29 –(4485)	0 –(36)	0	–84 (NA)	–24 (NA)	0 (NA)
	20	0	–88 (NA)	–47 (NA)	–11 (NA)	0	–85 (NA)	–28 (NA)	–1 (NA)	0	–84 (NA)	–23 (NA)	0 (NA)
5 clusters with 20 per cluster and ICC = 0.2	1	82	–6 –(8)	34 (41)	71 (86)	74	–11 –(14)	45 (61)	74 (100)	72	–12 –(16)	49 (68)	72 (100)
	5	4	–85 –(2261)	–49 –(1309)	–6 –(157)	1	–84 –(12853)	–32 –(4944)	0 (51)	0	–84 (NA)	–27 (NA)	0 (NA)
	20	0	–87 (NA)	–52 (NA)	–9 (NA)	0	–84 (NA)	–32 (NA)	–1 (NA)	0	–84 (NA)	–26 (NA)	–1 (NA)
20 clusters with 5 per cluster and ICC = 0.2	1	82	–6 –(7)	32 (39)	71 (86)	74	–10 –(14)	44 (59)	73 (99)	72	–12 –(16)	48 (67)	72 (100)
	5	4	–84 –(2240)	–47 –(1254)	–6 –(170)	1	–84 –(12815)	–30 –(4654)	0 –(47)	0	–84 (NA)	–25 (NA)	0 (NA)
	20	0	–87 (NA)	–50 (NA)	–11 (NA)	0	–84 (NA)	–30 (NA)	–1 (NA)	0	–83 (NA)	–25 (NA)	0 (NA)
10 clusters with 10 per cluster and ICC = 0.01	1	82	–6 –(7)	31 (39)	72 (88)	74	–11 –(14)	44 (59)	73 (99)	72	–12 –(16)	48 (66)	72 (100)
	5	4	–84 –(2254)	–46 –(1236)	–8 –(200)	1	–84 –(12868)	–30 –(4548)	0 –(42)	0	–84 (NA)	–24 (NA)	0 (NA)
	20	0	–87 (NA)	–51 (NA)	–10 (NA)	0	–84 (NA)	–31 (NA)	–1 (NA)	0	–84 (NA)	–25 (NA)	0 (NA)

continued

TABLE 15 *Bias if contamination independence* b = *0, all cluster members contaminated with time dependent contamination effect (cont'd)*

Number of clusters, number of members per cluster and ICC	*k*	α = 1				α = 2				α = 3			
		Individual simulation bias	Difference in bias between individual and cluster (%)			Individual simulation bias	Difference in bias between individual and cluster (%)			Individual simulation bias	Difference in bias between individual and cluster (%)		
			Cluster delay 1.49 (12%)	Cluster delay 4.45 (37%)	Cluster delay 6.65 (55%)		Cluster delay 1.86 (16%)	Cluster delay 6.95 (58%)	Cluster delay 11.95 (99%)		Cluster delay 1.92 (16%)	Cluster delay 8.1 (68%)	Cluster delay 12.84 (100%)
5 clusters with	1	82	–6	33	71	74	–11	44	73	72	–12	48	72
20 per cluster and			–(8)	(40)	(87)		–(15)	(60)	(98)		–(16)	(67)	(99)
ICC = 0.01	5	4	–84	–44	–6	1	–84	–29	0	0	–84	–24	0
			–(2228)	–(1182)	–(153)		–(12769)	–(4412)	–(58)		(NA)	(NA)	(NA)
	20	0	–88	–51	–10	0	–85	–31	0	0	–84	–25	0
			(NA)	(NA)	(NA)		(NA)	(NA)	(NA)		(NA)	(NA)	(NA)
20 clusters with	1	82	–5	31	72	74	–10	43	74	72	–11	48	72
5 per cluster and			–(7)	(38)	(89)		–(13)	(59)	(99)		–(15)	(66)	(100)
ICC = 0.01	5	4	–84	–47	–7	1	–84	–30	0	0	–84	–25	0
			–(2244)	–(1256)	–(184)		–(12851)	–(4625)	(11)		(NA)	(NA)	(NA)
	20	0	–88	–49	–11	0	–85	–30	–1	0	–84	–24	0
			(NA)	(NA)	(NA)		(NA)	(NA)	(NA)		(NA)	(NA)	(NA)
10 clusters with	1	77	–11	28	67	71	–13	42	71	71	–13	47	70
100 per cluster and			–(14)	(36)	(87)		–(19)	(58)	(99)		–(19)	(66)	(99)
ICC = 0.2	5	4	–84	–47	–5	0	–85	–32	0	0	–84	–26	0
			–(2116)	–(1192)	–(114)		–(69498)	–(26312)	–(5)		(NA)	(NA)	(NA)
	20	0	–88	–50	–10	0	–85	–31	–1	0	–84	–25	0
			(NA)	(NA)	(NA)		(NA)	(NA)	(NA)		(NA)	(NA)	(NA)
10 clusters with	1	77	–11	26	67	71	–13	40	71	71	–14	45	71
100 per cluster and			–(14)	(34)	(87)		–(19)	(56)	(99)		–(19)	(64)	(100)
ICC = 0.01	5	4	–83	–45	–6	0	–84	–30	–1	0	–84	–24	0
			–(2088)	–(1134)	–(150)		–(69002)	–(24435)	–(577)		(NA)	(NA)	(NA)
	20	0	–87	–50	–12	0	–84	–30	–1	0	–84	–24	0
			(NA)	(NA)	(NA)		(NA)	(NA)	(NA)		(NA)	(NA)	(NA)

b = 0 (no influence of baseline education on risk of contamination) individual contamination post cluster contamination

For a fixed contamination effect, as given in *Table 16*, there appeared to be a small, moderate or potentially large advantage to cluster randomisation. For constant contamination ($\kappa = 0.5$), the reduction in bias from cluster randomisation was most apparent when $k = 1$ (i.e. a short average time to individual contamination), and k_c was relatively large (i.e. a long average time to cluster contamination), as one might expect. For time-dependent contamination, as given in *Table 17*, the pattern was the same but a greater advantage from cluster randomisation was apparent.

b = 0.3 (moderate influence of baseline education on risk of contamination) all cluster members contaminated

For a fixed contamination effect, as given in *Table 18*, there was a substantial increase in bias for the cluster randomised design in many situations. There was a modest advantage in the cluster randomised design only when time to cluster contamination was late, that is, $k_c = 0.1$. For time-dependent effect of contamination level, as given in *Table 19*, there was an advantage to cluster randomisation for shorter cluster contamination times and also when the hazard increased with time.

b = 0.3 (moderate influence of baseline education on risk of contamination) individual contamination post cluster contamination

For both constant contamination effect and time-dependent contamination effect, as given in *Tables 20* and *21*, respectively, there was a general advantage in the cluster randomised design, with some exceptions. The reduction in bias appeared greater for longer time to cluster contamination, for increasing hazard with time and contamination level dependent on time.

b = 2 (strong influence of baseline education on risk of contamination) all cluster members contaminated

For the constant-effect contamination model as given in *Table 22*, the cluster randomised design was only preferable for the latest cluster contamination ($k_c = 0.1$). The difference in bias appeared uniform across different hazard types. For time-dependent level of contamination, as given in *Table 23*, there was again a disadvantage for the cluster randomised design when the time to cluster contamination was short (i.e. for $k_c = 0.01$). Otherwise, the bias appeared to be less and this was more pronounced when the hazard function increased with time.

b = 2 (strong influence of baseline education on risk of contamination) individual contamination post cluster contamination

For constant contamination effect, as given in *Table 24*, there was a small to moderate advantage for cluster randomised design. This appeared to be independent of hazard function type. A similar pattern was seen for time-dependent contamination, as given in *Table 25*, but with a greater difference in bias in all situations.

Discussion

In general, the simulations showed that there are situations where cluster randomisation is clearly an unwise choice of design. The most noticeable of these was the situation where contamination of a cluster results in the contamination of all individuals in that cluster. Simulations suggest that individual randomisation should be used in this situation except in cases where the risk of cluster contamination is likely to be very small and that for individual contamination comparatively high. This benefit of individual randomisation is most pronounced for contamination effects that are constant over time.

The simulations suggested that, when contamination is not cluster-wide but filters through more slowly amongst individuals, a cluster randomised design can produce less biased results than individually randomised designs. This preference was most evident in situations where contamination of clusters was delayed until late in the trial, and in situations where a subject's baseline education is strongly related to their likelihood of being contaminated. The benefits of cluster randomisation are also more apparent when the intensity of contamination effect is time dependent, that is, when earlier contamination leads to a greater effect on an individual. Further, when the risk of contamination is not constant with time but increases with time (e.g. when contamination is communicated between individuals), the benefits of cluster randomisation are more apparent.

There appeared to be little variability in the difference in bias with respect to cluster study design, that is, with respect to varying cluster size or ICC.

TABLE 16 *Bias if contamination independence* b = *0, individual post-cluster contamination with constant 50% contamination effect*

Number of clusters, number of members per cluster and ICC	*k*	α = 1				α = 2				α = 3			
		Individual simulation bias	Difference in bias between individual and cluster (%)			Individual simulation bias	Difference in bias between individual and cluster (%)			Individual simulation bias	Difference in bias between individual and cluster (%)		
			Cluster delay 1.49 (12%)	Cluster delay 4.45 (37%)	Cluster delay 6.65 (55%)		Cluster delay 1.86 (16%)	Cluster delay 6.95 (58%)	Cluster delay 11.95 (99%)		Cluster delay 1.92 (16%)	Cluster delay 8.1 (68%)	Cluster delay 12.84 (100%)
10 clusters with	1	50	1	16	44	50	0	27	50	50	0	31	50
10 per cluster and			(3)	(33)	(87)		(0)	(53)	(100)		(0)	(61)	(99)
ICC = 0.2	5	4	1	2	4	1	1	1	0	0	0	0	−1
			(14)	(49)	(95)		(80)	(85)	(39)		(NA)	(NA)	(NA)
	20	0	0	0	0	0	0	0	0	0	0	0	0
			(NA)	(NA)	(NA)		(NA)	(NA)	(NA)		(NA)	(NA)	(NA)
5 clusters with	1	50	1	14	43	50	0	23	50	50	0	27	50
20 per cluster and			(3)	(28)	(85)		(0)	(47)	(99)		(0)	(55)	(100)
ICC = 0.2	5	4	0	2	4	1	1	1	1	0	0	0	0
			(10)	(60)	(91)		(64)	(54)	(93)		(NA)	(NA)	(NA)
	20	0	0	0	0	0	0	0	0	0	0	0	0
			(NA)	(NA)	(NA)		(NA)	(NA)	(NA)		(NA)	(NA)	(NA)
20 clusters with	1	50	1	16	42	50	0	26	49	50	0	30	50
5 per cluster and			(3)	(32)	(84)		(0)	(52)	(99)		(0)	(60)	(100)
ICC = 0.2	5	4	1	2	4	1	1	1	1	0	0	0	0
			(20)	(52)	(100)		(84)	(98)	(68)		(NA)	(NA)	(NA)
	20	0	0	0	0	0	0	0	0	0	0	0	0
			(NA)	(NA)	(NA)		(NA)	(NA)	(NA)		(NA)	(NA)	(NA)
10 clusters with	1	50	1	17	42	50	0	27	49	50	0	31	50
10 per cluster and			(3)	(34)	(84)		(0)	(54)	(99)		(0)	(62)	(100)
ICC = 0.01	5	4	1	2	3	1	1	1	1	0	0	0	0
			(18)	(49)	(87)		(86)	(97)	(96)		(NA)	(NA)	(NA)
	20	0	0	0	0	0	0	0	0	0	0	0	0
			(NA)	(NA)	(NA)		(NA)	(NA)	(NA)		(NA)	(NA)	(NA)

continued

TABLE 16 *Bias if contamination independence* b = *0, individual post-cluster contamination with constant 50% contamination effect (cont'd)*

Number of clusters, number of members per cluster and ICC	***k***	**α = 1**				**α = 2**				**α = 3**			
		Individual simulation bias	**Difference in bias between individual and cluster (%)**			**Individual simulation bias**	**Difference in bias between individual and cluster (%)**			**Individual simulation bias**	**Difference in bias between individual and cluster (%)**		
			Cluster delay 1.49 (12%)	**Cluster delay 4.45 (37%)**	**Cluster delay 6.65 (55%)**		**Cluster delay 1.86 (16%)**	**Cluster delay 6.95 (58%)**	**Cluster delay 11.95 (99%)**		**Cluster delay 1.92 (16%)**	**Cluster delay 8.1 (68%)**	**Cluster delay 12.84 (100%)**
5 clusters with	1	50	1	16	42	50	0	25	48	50	0	29	49
20 per cluster and			(3)	(31)	(83)		(0)	(51)	(97)		(0)	(59)	(99)
ICC = 0.01	5	4	0	2	4	1	1	1	1	0	0	0	0
			(11)	(44)	(89)		(74)	(83)	(97)		(NA)	(NA)	(NA)
	20	0	0	0	0	0	0	0	0	0	0	0	0
			(NA)	(NA)	(NA)		(NA)	(NA)	(NA)		(NA)	(NA)	(NA)
20 clusters with	1	50	1	15	42	50	0	26	49	50	0	30	50
5 per cluster and			(3)	(30)	(84)		(0)	(51)	(99)		(0)	(60)	(100)
ICC = 0.01	5	4	0	2	4	1	1	1	1	0	0	0	0
			(11)	(43)	(92)		(80)	(68)	(99)		(NA)	(NA)	(NA)
	20	0	0	0	0	0	0	0	0	0	0	0	0
			(NA)	(NA)	(NA)		(NA)	(NA)	(NA)		(NA)	(NA)	(NA)
10 clusters with	1	49	0	16	42	50	0	26	50	50	0	30	50
100 per cluster and			(1)	(32)	(86)		(0)	(53)	(99)		(0)	(61)	(100)
ICC = 0.2	5	4	0	2	4	0	0	0	0	0	0	0	0
			(9)	(47)	(93)		–(98)	(8)	(36)		(NA)	(NA)	(NA)
	20	0	0	0	0	0	0	0	0	0	0	0	0
			(NA)	(NA)	(NA)		(NA)	(NA)	(NA)		(NA)	(NA)	(NA)
10 clusters with	1	49	1	15	42	50	0	26	49	50	0	30	50
100 per cluster and			(1)	(31)	(86)		(0)	(52)	(99)		(0)	(60)	(100)
ICC = 0.01	5	4	1	2	4	0	0	0	0	0	0	0	0
			(21)	(53)	(94)		(89)	(75)	(43)		(NA)	(NA)	(NA)
	20	0	0	0	0	0	0	0	0	0	0	0	0
			(NA)	(NA)	(NA)		(NA)	(NA)	(NA)		(NA)	(NA)	(NA)

TABLE 17 *Bias if contamination independence* b = *0, individual post-cluster contamination with time-dependent contamination effect*

Number of clusters, number of members per cluster and ICC	*k*	α = 1				α = 2				α = 3			
		Individual simulation bias	Difference in bias between individual and cluster (%)			Individual simulation bias	Difference in bias between individual and cluster (%)			Individual simulation bias	Difference in bias between individual and cluster (%)		
			Cluster delay 1.49 (12%)	Cluster delay 4.45 (37%)	Cluster delay 6.65 (55%)		Cluster delay 1.86 (16%)	Cluster delay 6.95 (58%)	Cluster delay 11.95 (99%)		Cluster delay 1.92 (16%)	Cluster delay 8.1 (68%)	Cluster delay 12.84 (100%)
10 clusters with 10 per cluster and ICC = 0.2	1	82	17 (20)	46 (56)	75 (92)	74	18 (25)	59 (80)	74 (99)	72	18 (25)	64 (88)	72 (100)
	5	4	1 (23)	2 (60)	4 (98)	1	1 (97)	1 (76)	0 (48)	0	0 (NA)	0 (NA)	0 (NA)
	20	0	0 (NA)	0 (NA)	0 (NA)	0	0 (NA)	0 (NA)	0 (NA)	0	0 (NA)	0 (NA)	0 (NA)
5 clusters with 20 per cluster and ICC = 0.2	1	82	15 (18)	47 (58)	74 (90)	74	17 (23)	60 (81)	74 (99)	72	17 (23)	64 (88)	72 (100)
	5	4	0 (11)	2 (61)	3 (85)	1	1 (92)	1 (92)	0 –(13)	0	0 (NA)	0 (NA)	–1 (NA)
	20	0	0 (NA)	0 (NA)	0 (NA)	0	0 (NA)	0 (NA)	0 (NA)	0	0 (NA)	0 (NA)	0 (NA)
20 clusters with 5 per cluster and ICC = 0.2	1	82	16 (20)	47 (58)	75 (92)	74	18 (25)	60 (81)	74 (100)	72	18 (25)	64 (88)	72 (100)
	5	4	0 (11)	2 (54)	3 (88)	1	0 (43)	0 (45)	0 (70)	0	0 (NA)	0 (NA)	0 (NA)
	20	0	0 (NA)	0 (NA)	0 (NA)	0	0 (NA)	0 (NA)	0 (NA)	0	0 (NA)	0 (NA)	0 (NA)
10 clusters with 10 per cluster and ICC = 0.01	1	82	16 (20)	47 (58)	75 (92)	74	18 (24)	60 (81)	74 (99)	72	18 (24)	64 (89)	72 (100)
	5	4	1 (16)	3 (67)	3 (88)	1	1 (85)	1 (76)	1 (82)	0	0 (NA)	0 (NA)	0 (NA)
	20	0	0 (NA)	0 (NA)	0 (NA)	0	0 (NA)	0 (NA)	0 (NA)	0	0 (NA)	0 (NA)	0 (NA)

continued

TABLE 17 *Bias if contamination independence* b = *0, individual post-cluster contamination with time-dependent contamination effect (cont'd)*

Number of clusters, number of members per cluster and ICC	***k***	**α = 1**				**α = 2**				**α = 3**			
		Individual simulation bias	**Difference in bias between individual and cluster (%)**			**Individual simulation bias**	**Difference in bias between individual and cluster (%)**			**Individual simulation bias**	**Difference in bias between individual and cluster (%)**		
			Cluster delay 1.49 (12%)	**Cluster delay 4.45 (37%)**	**Cluster delay 6.65 (55%)**		**Cluster delay 1.86 (16%)**	**Cluster delay 6.95 (58%)**	**Cluster delay 11.95 (99%)**		**Cluster delay 1.92 (16%)**	**Cluster delay 8.1 (68%)**	**Cluster delay 12.84 (100%)**
5 clusters with 20 per cluster and ICC = 0.01	1	82	16 (19)	46 (56)	75 (91)	82	16 (19)	46 (56)	75 (91)	72	18 (24)	64 (89)	72 (100)
	5	4	1 (20)	2 (59)	3 (90)	4	1 (20)	2 (59)	3 (90)	0	0 (NA)	0 (NA)	0 (NA)
	20	0	0 (NA)	0 (NA)	0 (NA)	0	0 (NA)	0 (NA)	0 (NA)	0	0 (NA)	0 (NA)	0 (NA)
20 clusters with 5 per cluster and ICC = 0.01	1	82	16 (20)	47 (58)	75 (91)	74	18 (24)	60 (81)	74 (100)	72	18 (24)	64 (89)	72 (100)
	5	4	1 (23)	2 (54)	4 (94)	1	1 (82)	0 (59)	1 (99)	0	0 (NA)	0 (NA)	0 (NA)
	20	0	0 (NA)	0 (NA)	0 (NA)	0	0 (NA)	0 (NA)	0 (NA)	0	0 (NA)	0 (NA)	0 (NA)
10 clusters with 100 per cluster and ICC = 0.2	1	77	11 (15)	43 (56)	71 (92)	71	15 (21)	57 (80)	71 (100)	71	16 (22)	62 (88)	70 (99)
	5	4	1 (18)	2 (62)	3 (85)	0	0 –(113)	0 (62)	0 –(125)	0	0 (NA)	0 (NA)	0 (NA)
	20	0	0 (NA)	0 (NA)	0 (NA)	0	0 (NA)	0 (NA)	0 (NA)	0	0 (NA)	0 (NA)	0 (NA)
10 clusters with 100 per cluster and ICC = 0.01	1	77	12 (16)	44 (57)	71 (92)	71	16 (22)	58 (82)	71 (100)	71	16 (23)	63 (89)	71 (100)
	5	4	1 (25)	2 (59)	4 (96)	0	0 (77)	0 –(11)	0 (41)	0	0 (NA)	0 (NA)	0 (NA)
	20	0	0 (NA)	0 (NA)	0 (NA)	0	0 (NA)	0 (NA)	0 (NA)	0	0 (NA)	0 (NA)	0 (NA)

TABLE 18 *Bias if moderate contamination dependence* b = *0.3, all cluster members contaminated with fixed 100% contamination effect*

Number of clusters, number of members per cluster and ICC	*k*	α = 1				α = 2				α = 3			
		Individual simulation bias	Difference in bias between individual and cluster (%)			Individual simulation bias	Difference in bias between individual and cluster (%)			Individual simulation bias	Difference in bias between individual and cluster (%)		
			Cluster delay 1.49 (12%)	Cluster delay 4.45 (37%)	Cluster delay 6.65 (55%)		Cluster delay 1.86 (16%)	Cluster delay 6.95 (58%)	Cluster delay 11.95 (99%)		Cluster delay 1.92 (16%)	Cluster delay 8.1 (68%)	Cluster delay 12.84 (100%)
10 clusters with	1	62	–38	–18	43	59	–41	–14	56	58	–42	–18	57
10 per cluster and			–(61)	–(28)	(68)		–(71)	–(25)	(96)		–(73)	–(32)	(99)
ICC = 0.2	5	40	–60	–40	23	37	–63	–37	35	36	–64	–40	35
			–(150)	–(101)	(58)		–(173)	–(101)	(94)		–(178)	–(111)	(98)
	20	1	–99	–79	–18	1	–100	–73	–2	1	–100	–75	0
			–(6980)	–(5589)	–(1234)		–(9048)	–(6626)	–(222)		–(9484)	–(7185)	(29)
5 clusters with	1	62	–38	–20	42	59	–41	–16	56	58	–42	–21	57
20 per cluster and			–(61)	–(32)	(67)		–(71)	–(27)	(96)		–(73)	–(36)	(99)
ICC = 0.2	5	40	–60	–39	21	37	–64	–35	35	36	–64	–38	35
			–(150)	–(97)	(53)		–(173)	–(97)	(94)		–(178)	–(105)	(98)
	20	1	–98	–80	–17	1	–99	–74	–1	1	–99	–77	0
			–(6918)	–(5646)	–(1212)		–(8960)	–(6772)	–(85)		–(9391)	–(7290)	(20)
20 clusters with	1	62	–38	–18	43	59	–41	–15	56	58	–42	–19	57
5 per cluster and			–(61)	–(29)	(68)		–(71)	–(26)	(95)		–(73)	–(32)	(99)
ICC = 0.2	5	40	–60	–41	19	37	–63	–38	34	36	–64	–41	36
			–(150)	–(104)	(49)		–(173)	–(103)	(94)		–(177)	–(113)	(100)
	20	1	–98	–79	–19	1	–99	–72	–1	1	–99	–75	0
			–(6927)	–(5580)	–(1315)		–(8981)	–(6546)	–(113)		–(9413)	–(7139)	(39)
10 clusters with	1	62	–37	–19	44	59	–41	–16	57	58	–42	–19	58
10 per cluster and			–(60)	–(30)	(70)		–(70)	–(28)	(97)		–(72)	–(34)	(100)
ICC = 0.01	5	40	–60	–40	21	37	–63	–36	34	36	–64	–39	36
			–(149)	–(99)	(53)		–(172)	–(98)	(92)		–(177)	–(109)	(99)
	20	1	–99	–77	–18	1	–99	–70	–1	1	–99	–73	1
			–(6947)	–(5396)	–(1260)		–(8997)	–(6395)	–(74)		–(9430)	–(6942)	(61)

continued

TABLE 18 *Bias if moderate contamination dependence* b = *0.3, all cluster members contaminated with fixed 100% contamination effect (cont'd)*

Number of clusters, number of members per cluster and ICC	***k***	**α = 1**				**α = 2**				**α = 3**			
		Individual simulation bias	**Difference in bias between individual and cluster (%)**			**Individual simulation bias**	**Difference in bias between individual and cluster (%)**			**Individual simulation bias**	**Difference in bias between individual and cluster (%)**		
			Cluster delay 1.49 (12%)	**Cluster delay 4.45 (37%)**	**Cluster delay 6.65 (55%)**		**Cluster delay 1.86 (16%)**	**Cluster delay 6.95 (58%)**	**Cluster delay 11.95 (99%)**		**Cluster delay 1.92 (16%)**	**Cluster delay 8.1 (68%)**	**Cluster delay 12.84 (100%)**
5 clusters with	1	62	–38	–17	43	59	–41	–15	56	58	–42	–18	57
20 per cluster and			–(60)	–(28)	(70)		–(70)	–(26)	(95)		–(73)	–(32)	(99)
ICC = 0.01	5	40	–60	–39	19	37	–63	–35	35	36	–64	–38	36
			–(150)	–(97)	(47)		–(173)	–(96)	(94)		–(177)	–(106)	(99)
	20	1	–99	–80	–17	1	–99	–72	–2	1	–99	–76	1
			–(6940)	–(5666)	–(1200)		–(8988)	–(6562)	–(160)		–(9421)	–(7241)	(56)
20 clusters with	1	62	–37	–19	43	59	–41	–15	56	58	–42	–19	58
5 per cluster and			–(60)	–(30)	(69)		–(70)	–(26)	(96)		–(72)	–(33)	(99)
ICC = 0.01	5	40	–60	–40	20	37	–63	–36	34	36	–64	–40	36
			–(149)	–(99)	(50)		–(172)	–(99)	(93)		–(177)	–(112)	(99)
	20	1	–99	–79	–19	1	–99	–72	–2	1	–99	–75	0
			–(6942)	–(5590)	–(1363)		–(8995)	–(6586)	–(184)		–(9428)	–(7181)	(27)
10 clusters with	1	61	–39	–20	41	58	–43	–16	55	57	–43	–19	57
100 per cluster and			–(65)	–(33)	(67)		–(74)	–(28)	(96)		–(76)	–(34)	(100)
ICC = 0.2	5	40	–60	–39	19	37	–64	–35	35	36	–64	–39	36
			–(151)	–(99)	(48)		–(174)	–(96)	(95)		–(179)	–(108)	(99)
	20	1	–99	–79	–19	1	–99	–71	–2	1	–99	–74	1
			–(7887)	–(6308)	–(1500)		–(10349)	–(7447)	–(238)		–(11023)	–(8201)	(85)
10 clusters with	1	61	–39	–19	40	58	–42	–15	55	57	–43	–19	57
100 per cluster and			–(65)	–(32)	(66)		–(74)	–(27)	(95)		–(76)	–(33)	(99)
ICC = 0.01	5	40	–60	–39	19	37	–63	–35	34	36	–64	–38	36
			–(151)	–(99)	(48)		–(173)	–(97)	(94)		–(178)	–(107)	(99)
	20	1	–99	–80	–16	1	–99	–73	–2	1	–99	–76	1
			–(7883)	–(6351)	–(1313)		–(10344)	–(7577)	–(160)		–(11018)	–(8406)	(79)

TABLE 19 *Bias if moderate contamination dependence* b = *0.3, all cluster members contaminated with time-dependent contamination effect*

Number of clusters, number of members per cluster and ICC	*k*	α = 1				α = 2				α = 3			
		Individual simulation bias	Difference in bias between individual and cluster (%)			Individual simulation bias	Difference in bias between individual and cluster (%)			Individual simulation bias	Difference in bias between individual and cluster (%)		
			Cluster delay 1.49 (12%)	Cluster delay 4.45 (37%)	Cluster delay 6.65 (55%)		Cluster delay 1.86 (16%)	Cluster delay 6.95 (58%)	Cluster delay 11.95 (99%)		Cluster delay 1.92 (16%)	Cluster delay 8.1 (68%)	Cluster delay 12.84 (100%)
10 clusters with	1	56	–32	5	44	52	–32	21	51	51	–32	27	51
10 per cluster and			–(57)	(9)	(79)		–(61)	(41)	(97)		–(63)	(52)	(100)
ICC = 0.2	5	37	–52	–16	26	33	–52	1	32	32	–52	6	32
			–(141)	–(43)	(72)		–(157)	(2)	(97)		–(160)	(19)	(100)
	20	1	–88	–51	–10	1	–85	–31	–1	1	–84	–25	0
			–(8678)	–(5038)	–(970)		–(12077)	–(4444)	–(87)		–(12748)	–(3802)	(42)
5 clusters with	1	56	–31	8	46	52	–32	23	52	51	–32	28	51
20 per cluster and			–(56)	(14)	(82)		–(60)	(44)	(99)		–(62)	(54)	(100)
ICC = 0.2	5	37	–50	–16	27	33	–51	1	33	32	–51	6	32
			–(137)	–(45)	(73)		–(154)	(2)	(99)		–(158)	(19)	(100)
	20	1	–87	–51	–9	1	–84	–31	0	1	–84	–25	0
			–(8632)	–(5015)	–(866)		–(12005)	–(4426)	–(14)		–(12680)	–(3783)	(64)
20 clusters with	1	56	–32	5	46	52	–33	21	51	51	–33	27	51
5 per cluster and			–(57)	(10)	(82)		–(62)	(41)	(98)		–(64)	(52)	(99)
ICC = 0.2	5	37	–51	–13	27	33	–51	2	32	32	–51	8	32
			–(139)	–(37)	(74)		–(155)	(7)	(98)		–(159)	(24)	(100)
	20	1	–87	–50	–8	1	–84	–30	0	1	–84	–25	0
			–(8629)	–(4923)	–(811)		–(12027)	–(4341)	(69)		–(12704)	–(3715)	(52)
10 clusters with	1	56	–32	4	47	52	–32	20	51	51	–33	25	51
10 per cluster and			–(57)	(7)	(84)		–(62)	(39)	(98)		–(64)	(49)	(99)
ICC = 0.01	5	37	–51	–13	27	33	–52	2	33	32	–52	8	32
			–(140)	–(37)	(74)		–(156)	(7)	(98)		–(160)	(24)	(100)
	20	1	–86	–48	–10	1	–83	–29	–1	1	–83	–23	0
			–(8503)	–(4705)	–(1034)		–(11892)	–(4098)	–(88)		–(12584)	–(3511)	(39)

continued

TABLE 19 *Bias if moderate contamination dependence* b = *0.3, all cluster members contaminated with time-dependent contamination effect (cont'd)*

Number of clusters, number of members per cluster and ICC	***k***	**α = 1**				**α = 2**				**α = 3**			
		Individual simulation bias	**Difference in bias between individual and cluster (%)**			**Individual simulation bias**	**Difference in bias between individual and cluster (%)**			**Individual simulation bias**	**Difference in bias between individual and cluster (%)**		
			Cluster delay 1.49 (12%)	**Cluster delay 4.45 (37%)**	**Cluster delay 6.65 (55%)**		**Cluster delay 1.86 (16%)**	**Cluster delay 6.95 (58%)**	**Cluster delay 11.95 (99%)**		**Cluster delay 1.92 (16%)**	**Cluster delay 8.1 (68%)**	**Cluster delay 12.84 (100%)**
5 clusters with	1	56	–31	7	45	52	–31	23	51	51	–32	28	51
20 per cluster and			–(55)	(13)	(81)		–(60)	(44)	(98)		–(62)	(54)	(100)
ICC = 0.01	5	37	–51	–13	26	33	–51	3	32	32	–51	8	32
			–(138)	–(36)	(71)		–(155)	(8)	(97)		–(159)	(24)	(100)
	20	1	–86	–52	–10	1	–83	–32	0	1	–83	–26	0
			–(8509)	–(5194)	–(972)		–(11894)	–(4629)	–(61)		–(12594)	–(3957)	(58)
20 clusters with	1	56	–32	7	46	52	–32	22	51	51	–33	27	51
5 per cluster and			–(57)	(13)	(82)		–(62)	(43)	(98)		–(64)	(53)	(100)
ICC = 0.01	5	37	–51	–14	27	33	–51	2	32	32	–51	7	32
			–(139)	–(39)	(74)		–(155)	(6)	(98)		–(159)	(22)	(100)
	20	1	–86	–50	–9	1	–83	–30	0	1	–83	–24	0
			–(8539)	–(4899)	–(894)		–(11911)	–(4321)	(26)		–(12599)	–(3697)	(62)
10 clusters with	1	55	–32	4	44	52	–33	21	51	51	–33	27	51
100 per cluster and			–(59)	(7)	(80)		–(63)	(40)	(98)		–(64)	(52)	(100)
ICC = 0.2	5	35	–53	–18	26	32	–52	–1	31	31	–53	5	31
			–(150)	–(50)	(73)		–(164)	–(2)	(98)		–(167)	(16)	(99)
	20	1	–86	–50	–9	1	–83	–31	0	1	–83	–24	1
			–(8566)	–(4981)	–(864)		–(11462)	–(4193)	(34)		–(12126)	–(3566)	(90)
10 clusters with	1	55	–32	5	45	52	–32	22	51	51	–33	27	51
100 per cluster and			–(58)	(9)	(82)		–(62)	(41)	(98)		–(63)	(52)	(100)
ICC = 0.01	5	35	–52	–17	25	32	–52	0	31	31	–53	5	31
			–(148)	–(48)	(72)		–(163)	–(1)	(97)		–(167)	(17)	(100)
	20	1	–87	–50	–10	1	–84	–30	0	1	–83	–24	1
			–(8611)	–(4993)	–(990)		–(11491)	–(4171)	(32)		–(12138)	–(3542)	(92)

TABLE 20 *Bias if moderate contamination dependence* b = *0.3, individual post-cluster contamination with constant 50% contamination effect*

Number of clusters, number of members per cluster and ICC	*k*	α = 1				α = 2				α = 3			
		Individual simulation bias	Difference in bias between individual and cluster (%)			Individual simulation bias	Difference in bias between individual and cluster (%)			Individual simulation bias	Difference in bias between individual and cluster (%)		
			Cluster delay 1.49 (12%)	Cluster delay 4.45 (37%)	Cluster delay 6.65 (55%)		Cluster delay 1.86 (16%)	Cluster delay 6.95 (58%)	Cluster delay 11.95 (99%)		Cluster delay 1.92 (16%)	Cluster delay 8.1 (68%)	Cluster delay 12.84 (100%)
10 clusters with	1	30	0	7	25	28	0	9	28	28	0	9	28
10 per cluster and			(1)	(23)	(81)		(0)	(33)	(97)		(0)	(33)	(100)
ICC = 0.2	5	21	2	6	17	19	2	8	18	19	2	8	18
			(8)	(30)	(82)		(8)	(42)	(97)		(8)	(42)	(99)
	20	1	0	0	0	0	0	0	0	0	0	0	0
			–(10)	(58)	(55)		–(25)	(69)	(60)		–(26)	(81)	(57)
5 clusters with	1	30	1	9	26	28	1	11	28	28	1	10	28
20 per cluster and			(4)	(28)	(84)		(4)	(37)	(98)		(4)	(37)	(99)
ICC = 0.2	5	21	1	7	18	19	1	8	19	19	1	8	18
			(3)	(32)	(85)		(4)	(44)	(99)		(3)	(44)	(99)
	20	1	0	0	1	0	0	0	0	0	0	0	0
			(36)	–(41)	(94)		(39)	–(63)	(69)		(34)	–(66)	(66)
20 clusters with	1	30	1	7	25	28	1	10	28	28	1	9	28
5 per cluster and			(4)	(24)	(82)		(4)	(34)	(98)		(3)	(34)	(100)
ICC = 0.2	5	21	1	6	17	19	1	8	19	19	1	7	19
			(6)	(28)	(83)		(6)	(39)	(99)		(5)	(40)	(99)
	20	1	0	0	0	0	0	0	0	0	0	0	0
			(43)	(38)	(72)		(48)	(39)	(38)		(47)	(42)	(32)
10 clusters with	1	30	0	7	26	28	0	9	28	28	0	9	28
10 per cluster and			(0)	(24)	(85)		–(1)	(33)	(98)		–(1)	(32)	(100)
ICC = 0.01	5	21	1	6	18	19	1	7	19	19	1	7	19
			(5)	(28)	(85)		(5)	(38)	(100)		(5)	(39)	(99)
	20	1	0	0	0	0	0	0	0	0	0	0	0
			–(9)	(62)	(51)		–(26)	(99)	(47)		–(35)	(92)	(45)

continued

TABLE 20 *Bias if moderate contamination dependence* b = *0.3, individual post-cluster contamination with constant 50% contamination effect (cont'd)*

Number of clusters, number of members per cluster and ICC	***k***	**α = 1**				**α = 2**				**α = 3**			
		Individual simulation bias	**Difference in bias between individual and cluster (%)**			**Individual simulation bias**	**Difference in bias between individual and cluster (%)**			**Individual simulation bias**	**Difference in bias between individual and cluster (%)**		
			Cluster delay 1.49 (12%)	**Cluster delay 4.45 (37%)**	**Cluster delay 6.65 (55%)**		**Cluster delay 1.86 (16%)**	**Cluster delay 6.95 (58%)**	**Cluster delay 11.95 (99%)**		**Cluster delay 1.92 (16%)**	**Cluster delay 8.1 (68%)**	**Cluster delay 12.84 (100%)**
5 clusters with	1	30	1	8	25	28	1	10	28	28	1	10	28
20 per cluster and			(4)	(26)	(84)		(3)	(35)	(98)		(3)	(35)	(100)
ICC = 0.01	5	21	2	6	17	19	1	7	19	19	1	7	19
			(7)	(28)	(82)		(7)	(38)	(98)		(7)	(39)	(99)
	20	1	0	0	1	0	0	0	0	0	0	0	0
			–(15)	(34)	(87)		–(36)	(40)	(95)		–(48)	(44)	(90)
20 clusters with	1	30	1	7	25	28	1	9	28	28	1	9	28
5 per cluster and			(2)	(22)	(82)		(2)	(32)	(98)		(2)	(32)	(99)
ICC = 0.01	5	21	1	6	18	19	1	7	19	19	1	7	19
			(6)	(27)	(84)		(6)	(37)	(98)		(6)	(38)	(100)
	20	1	0	0	0	0	0	0	0	0	0	0	0
			–(12)	(43)	(61)		–(29)	(58)	(72)		–(32)	(61)	(69)
10 clusters with	1	30	1	7	24	29	1	9	28	28	1	9	28
100 per cluster and			(2)	(22)	(79)		(2)	(33)	(98)		(2)	(32)	(100)
ICC = 0.2	5	20	1	5	17	19	1	7	18	18	1	7	18
			(3)	(26)	(83)		(4)	(38)	(99)		(4)	(38)	(99)
	20	1	0	0	1	0	0	0	0	0	0	0	0
			–(51)	(66)	(93)		–(64)	(89)	(93)		–(68)	(94)	(92)
10 clusters with	1	30	0	7	25	29	0	9	28	28	0	9	28
100 per cluster and			(1)	(22)	(82)		(1)	(33)	(99)		(1)	(32)	(100)
ICC = 0.01	5	20	0	5	17	19	0	7	18	18	0	7	18
			(2)	(26)	(83)		(2)	(38)	(98)		(2)	(39)	(100)
	20	1	0	0	0	0	0	0	0	0	0	0	0
			(5)	(22)	(73)		(8)	(39)	(90)		(8)	(42)	(91)

TABLE 21 *Bias if moderate contamination dependence* b = *0.3, individual post-cluster contamination with time-dependent contamination effect*

Number of clusters, number of members per cluster and ICC	*k*	α = 1				α = 2				α = 3			
		Individual simulation bias	Difference in bias between individual and cluster (%)			Individual simulation bias	Difference in bias between individual and cluster (%)			Individual simulation bias	Difference in bias between individual and cluster (%)		
			Cluster delay 1.49 (12%)	Cluster delay 4.45 (37%)	Cluster delay 6.65 (55%)		Cluster delay 1.86 (16%)	Cluster delay 6.95 (58%)	Cluster delay 11.95 (99%)		Cluster delay 1.92 (16%)	Cluster delay 8.1 (68%)	Cluster delay 12.84 (100%)
10 clusters with 10 per cluster and ICC = 0.2	1	56	7 (12)	29 (52)	50 (90)	52	8 (15)	38 (72)	52 (99)	51	8 (15)	40 (78)	51 (100)
	5	37	6 (15)	20 (56)	34 (92)	33	6 (18)	25 (75)	33 (100)	32	6 (19)	26 (80)	32 (99)
	20	1	0 (28)	0 (18)	1 (75)	1	0 (26)	0 (19)	0 (49)	1	0 (25)	0 (23)	0 (45)
5 clusters with 20 per cluster and ICC = 0.2	1	56	8 (14)	28 (51)	52 (93)	52	9 (17)	37 (70)	52 (100)	51	9 (17)	39 (76)	51 (99)
	5	37	6 (15)	20 (54)	33 (91)	33	6 (19)	24 (72)	33 (98)	32	6 (19)	25 (78)	32 (99)
	20	1	0 –(19)	0 –(26)	1 (84)	1	0 –(34)	0 –(36)	1 (91)	1	0 –(38)	0 –(36)	1 (91)
20 clusters with 5 per cluster and ICC = 0.2	1	56	8 (13)	28 (51)	51 (91)	52	9 (16)	37 (71)	52 (100)	51	9 (17)	39 (77)	51 (100)
	5	37	7 (19)	20 (55)	33 (91)	33	7 (22)	25 (74)	33 (99)	32	7 (22)	26 (80)	32 (100)
	20	1	0 –(14)	1 (51)	1 (79)	1	0 –(29)	1 (71)	1 (82)	1	0 –(34)	1 (77)	1 (81)
10 clusters with 10 per cluster and ICC = 0.01	1	56	8 (14)	27 (49)	51 (90)	52	9 (17)	36 (69)	52 (99)	51	9 (17)	39 (75)	51 (100)
	5	37	6 (16)	19 (52)	33 (91)	33	6 (19)	24 (73)	33 (99)	32	6 (19)	25 (79)	32 (100)
	20	1	0 (36)	0 (47)	1 (80)	1	0 (37)	0 (57)	1 (87)	1	0 (39)	0 (60)	1 (87)

continued

TABLE 21 *Bias if moderate contamination dependence* b = *0.3, individual post-cluster contamination with time-dependent contamination effect (cont'd)*

Number of clusters, number of members per cluster and ICC	***k***	**α = 1**				**α = 2**				**α = 3**			
		Individual simulation bias	**Difference in bias between individual and cluster (%)**			**Individual simulation bias**	**Difference in bias between individual and cluster (%)**			**Individual simulation bias**	**Difference in bias between individual and cluster (%)**		
			Cluster delay 1.49 (12%)	**Cluster delay 4.45 (37%)**	**Cluster delay 6.65 (55%)**		**Cluster delay 1.86 (16%)**	**Cluster delay 6.95 (58%)**	**Cluster delay 11.95 (99%)**		**Cluster delay 1.92 (16%)**	**Cluster delay 8.1 (68%)**	**Cluster delay 12.84 (100%)**
5 clusters with 20 per cluster and ICC = 0.01	1	56	9 (16)	30 (54)	50 (89)	52	9 (16)	30 (54)	50 (89)	51	9 (17)	40 (78)	51 (100)
	5	37	6 (17)	21 (57)	33 (90)	33	6 (17)	21 (57)	33 (90)	32	6 (19)	26 (79)	32 (100)
	20	1	0 (33)	1 (87)	1 (90)	1	0 (33)	1 (87)	1 (90)	1	0 (32)	0 (62)	1 (84)
20 clusters with 5 per cluster and ICC = 0.01	1	56	8 (14)	29 (52)	50 (90)	52	9 (17)	37 (72)	52 (99)	51	9 (17)	40 (78)	51 (100)
	5	37	6 (16)	20 (54)	33 (90)	33	6 (19)	24 (73)	33 (99)	32	6 (19)	26 (79)	32 (100)
	20	1	0 (27)	0 (41)	1 (80)	1	0 (30)	0 (55)	1 (85)	1	0 (32)	0 (62)	1 (84)
10 clusters with 100 per cluster and ICC = 0.2	1	55	6 (10)	28 (50)	50 (91)	52	7 (14)	37 (71)	52 (100)	51	7 (14)	40 (77)	51 (100)
	5	35	6 (16)	19 (53)	32 (91)	32	6 (20)	23 (73)	32 (100)	31	6 (20)	25 (79)	31 (100)
	20	1	0 (35)	1 (81)	1 (85)	1	0 (42)	1 (93)	1 (69)	1	0 (43)	1 (86)	0 (66)
10 clusters with 100 per cluster and ICC = 0.01	1	55	7 (14)	27 (49)	49 (89)	52	9 (17)	36 (69)	52 (99)	51	9 (17)	39 (76)	51 (100)
	5	35	5 (14)	19 (54)	31 (89)	32	6 (18)	24 (74)	32 (99)	31	6 (18)	25 (80)	31 (100)
	20	1	0 (16)	1 (52)	1 (97)	1	0 (19)	1 (71)	1 (84)	1	0 (20)	1 (77)	1 (82)

TABLE 22 *Bias if strong contamination dependence* b = 2, *all cluster members contaminated with fixed 100% contamination effect*

Number of clusters, number of members per cluster and ICC	*k*	α = 1				α = 2				α = 3			
		Individual simulation bias	Difference in bias between individual and cluster (%)			Individual simulation bias	Difference in bias between individual and cluster (%)			Individual simulation bias	Difference in bias between individual and cluster (%)		
			Cluster delay 1.49 (12%)	Cluster delay 4.45 (37%)	Cluster delay 6.65 (55%)		Cluster delay 1.86 (16%)	Cluster delay 6.95 (58%)	Cluster delay 11.95 (99%)		Cluster delay 1.92 (16%)	Cluster delay 8.1 (68%)	Cluster delay 12.84 (100%)
10 clusters with	1	52	–49	–27	31	51	–49	–21	49	51	–49	–24	51
10 per cluster and			–(94)	–(53)	(60)		–(97)	–(41)	(97)		–(97)	–(47)	(99)
ICC = 0.2	5	49	–51	–30	31	48	–52	–24	47	48	–52	–26	48
			–(105)	–(61)	(63)		–(107)	–(49)	(96)		–(107)	–(55)	(100)
	20	38	–62	–44	16	37	–63	–38	34	37	–63	–41	37
			–(165)	–(117)	(42)		–(169)	–(102)	(92)		–(170)	–(111)	(99)
5 clusters with	1	52	–49	–28	29	51	–50	–24	49	51	–50	–26	50
20 per cluster and			–(95)	–(54)	(57)		–(98)	–(47)	(95)		–(98)	–(51)	(99)
ICC = 0.2	5	49	–52	–32	32	48	–52	–26	46	48	–52	–28	48
			–(106)	–(66)	(65)		–(108)	–(53)	(96)		–(108)	–(58)	(99)
	20	38	–62	–44	20	37	–63	–38	35	37	–63	–42	37
			–(166)	–(117)	(53)		–(170)	–(101)	(95)		–(171)	–(113)	(99)
20 clusters with	1	52	–49	–29	31	51	–49	–22	49	51	–49	–25	50
5 per cluster and			–(94)	–(56)	(61)		–(97)	–(44)	(95)		–(97)	–(49)	(99)
ICC = 0.2	5	49	–51	–32	29	48	–52	–24	46	48	–52	–27	48
			–(105)	–(66)	(60)		–(107)	–(51)	(95)		–(107)	–(57)	(99)
	20	38	–62	–43	18	37	–62	–37	35	37	–63	–40	36
			–(164)	–(114)	(47)		–(168)	–(100)	(94)		–(169)	–(108)	(99)
10 clusters with	1	52	–48	–30	30	51	–49	–24	47	51	–49	–27	50
10 per cluster and			–(94)	–(58)	(57)		–(96)	–(47)	(93)		–(97)	–(53)	(99)
ICC = 0.01	5	49	–51	–31	30	48	–51	–23	46	48	–51	–26	48
			–(104)	–(63)	(61)		–(106)	–(49)	(96)		–(107)	–(54)	(99)
	20	38	–62	–43	18	37	–63	–35	35	37	–63	–39	37
			–(165)	–(114)	(46)		–(169)	–(95)	(94)		–(170)	–(105)	(99)

continued

TABLE 22 *Bias if strong contamination dependence* b = 2, *all cluster members contaminated with fixed 100% contamination effect (cont'd)*

Number of clusters, number of members per cluster and ICC	***k***	**α = 1**				**α = 2**				**α = 3**			
		Individual simulation bias	**Difference in bias between individual and cluster (%)**			**Individual simulation bias**	**Difference in bias between individual and cluster (%)**			**Individual simulation bias**	**Difference in bias between individual and cluster (%)**		
			Cluster delay 1.49 (12%)	**Cluster delay 4.45 (37%)**	**Cluster delay 6.65 (55%)**		**Cluster delay 1.86 (16%)**	**Cluster delay 6.95 (58%)**	**Cluster delay 11.95 (99%)**		**Cluster delay 1.92 (16%)**	**Cluster delay 8.1 (68%)**	**Cluster delay 12.84 (100%)**
5 clusters with	1	52	–49	–32	33	51	–49	–24	49	51	–49	–28	50
20 per cluster and			–(94)	–(61)	(64)		–(97)	–(47)	(96)		–(97)	–(54)	(99)
ICC = 0.01	5	49	–51	–31	30	48	–52	–24	46	48	–52	–27	48
			–(105)	–(64)	(62)		–(107)	–(51)	(95)		–(107)	–(56)	(100)
	20	38	–62	–41	12	37	–63	–35	34	37	–63	–37	36
			–(165)	–(108)	(32)		–(169)	–(93)	(92)		–(170)	–(99)	(97)
20 clusters with	1	52	–49	–30	32	51	–49	–23	49	51	–49	–26	51
5 per cluster and			–(94)	–(57)	(61)		–(97)	–(45)	(96)		–(97)	–(51)	(100)
ICC = 0.01	5	49	–51	–32	29	48	–52	–26	46	48	–52	–29	48
			–(105)	–(65)	(59)		–(107)	–(54)	(95)		–(107)	–(60)	(100)
	20	38	–62	–43	19	37	–63	–36	35	37	–63	–39	37
			–(164)	–(114)	(50)		–(169)	–(98)	(93)		–(170)	–(105)	(99)
10 clusters with	1	52	–48	–31	31	51	–49	–25	48	51	–49	–28	50
100 per cluster and			–(94)	–(59)	(60)		–(96)	–(49)	(94)		–(96)	–(55)	(99)
ICC = 0.2	5	49	–51	–32	31	48	–52	–26	46	48	–52	–29	48
			–(105)	–(66)	(63)		–(108)	–(55)	(97)		–(108)	–(61)	(100)
	20	37	–63	–43	17	36	–64	–37	33	36	–64	–40	36
			–(171)	–(116)	(46)		–(175)	–(102)	(92)		–(176)	–(110)	(98)
10 clusters with	1	52	–48	–26	30	51	–18	49	–49	51	–21	50	–49
100 per cluster and			–(94)	–(51)	(58)		–(36)	(96)	–(96)		–(42)	(99)	–(96)
ICC = 0.01	5	49	–52	–32	28	48	–25	45	–52	48	–28	48	–52
			–(106)	–(66)	(58)		–(53)	(95)	–(109)		–(58)	(99)	–(109)
	20	37	(NA)	–45	17	36	–38	34	(NA)	36	–41	36	(NA)
			(NA)	–(121)	(48)		–(105)	(93)	(NA)		–(113)	(99)	(NA)

TABLE 23 *Bias if high contamination dependence* b = 2, *all cluster members contaminated with time-dependent contamination effect*

Number of clusters, number of members per cluster and ICC	*k*	α = 1				α = 2				α = 3			
		Individual simulation bias	Difference in bias between individual and cluster (%)			Individual simulation bias	Difference in bias between individual and cluster (%)			Individual simulation bias	Difference in bias between individual and cluster (%)		
			Cluster delay 1.49 (12%)	Cluster delay 4.45 (37%)	Cluster delay 6.65 (55%)		Cluster delay 1.86 (16%)	Cluster delay 6.95 (58%)	Cluster delay 11.95 (99%)		Cluster delay 1.92 (16%)	Cluster delay 8.1 (68%)	Cluster delay 12.84 (100%)
10 clusters with	1	51	–37	0	40	50	–35	20	50	50	–34	26	50
10 per cluster and			–(73)	(1)	(78)		–(69)	(39)	(99)		–(68)	(51)	(100)
ICC = 0.2	5	49	–38	–1	38	49	–36	18	47	48	–36	24	48
			–(78)	–(3)	(77)		–(74)	(37)	(98)		–(74)	(49)	(100)
	20	36	–52	–15	26	36	–49	4	35	36	–48	10	36
			–(142)	–(42)	(72)		–(136)	(11)	(98)		–(136)	(28)	(100)
5 clusters with	1	51	–37	2	40	50	–34	20	49	50	–34	26	50
20 per cluster and			–(73)	(5)	(79)		–(68)	(40)	(98)		–(68)	(52)	(100)
ICC = 0.2	5	49	–38	2	39	49	–36	21	47	48	–35	26	48
			–(78)	(4)	(79)		–(73)	(43)	(97)		–(72)	(54)	(99)
	20	36	–51	–14	26	36	–49	6	35	36	–48	11	36
			–(142)	–(39)	(72)		–(136)	(16)	(99)		–(136)	(32)	(100)
20 clusters with	1	51	–37	1	41	50	–34	20	49	50	–34	26	50
5 per cluster and			–(73)	(2)	(81)		–(68)	(40)	(98)		–(68)	(51)	(99)
ICC = 0.2	5	49	–38	–1	41	49	–36	18	48	48	–35	24	48
			–(77)	–(2)	(83)		–(73)	(36)	(99)		–(73)	(49)	(100)
	20	36	–51	–13	25	36	–49	6	35	36	–48	11	36
			–(141)	–(37)	(70)		–(136)	(16)	(97)		–(136)	(32)	(100)
10 clusters with	1	51	–37	–1	41	50	–34	18	49	50	–34	24	50
10 per cluster and			–(72)	–(2)	(81)		–(68)	(36)	(98)		–(68)	(48)	(100)
ICC = 0.01	5	49	–38	–1	38	49	–36	18	48	48	–36	24	48
			–(78)	–(1)	(77)		–(74)	(38)	(98)		–(73)	(50)	(100)
	20	36	–51	–14	26	36	–49	5	35	36	–48	11	35
			–(141)	–(38)	(73)		–(136)	(15)	(97)		–(135)	(31)	(99)

continued

TABLE 23 *Bias if high contamination dependence* b = 2, *all cluster members contaminated with time-dependent contamination effect (cont'd)*

Number of clusters, number of members per cluster and ICC	***k***	**α = 1**				**α = 2**				**α = 3**			
		Individual simulation bias	**Difference in bias between individual and cluster (%)**			**Individual simulation bias**	**Difference in bias between individual and cluster (%)**			**Individual simulation bias**	**Difference in bias between individual and cluster (%)**		
			Cluster delay 1.49 (12%)	**Cluster delay 4.45 (37%)**	**Cluster delay 6.65 (55%)**		**Cluster delay 1.86 (16%)**	**Cluster delay 6.95 (58%)**	**Cluster delay 11.95 (99%)**		**Cluster delay 1.92 (16%)**	**Cluster delay 8.1 (68%)**	**Cluster delay 12.84 (100%)**
5 clusters with 20 per cluster and ICC = 0.01	1	51	–36 –(71)	–2 –(3)	40 (79)	50	–31 –(60)	18 (36)	49 (98)	50	–33 –(67)	24 (49)	50 (99)
	5	49	–39 –(79)	–2 –(3)	37 (74)	49	–51 –(155)	18 (38)	48 (98)	48	–36 –(74)	24 (50)	48 (100)
	20	36	(NA) (NA)	–12 –(34)	25 (70)	36	–83 –(11894)	7 (18)	35 (98)	36	(NA) (NA)	12 (34)	36 (100)
20 clusters with 5 per cluster and ICC = 0.01	1	51	–36 –(72)	1 (2)	41 (81)	50	–34 –(67)	20 (40)	49 (98)	50	–34 –(67)	26 (52)	50 (100)
	5	49	–38 –(78)	–1 –(1)	40 (81)	49	–36 –(73)	18 (37)	48 (99)	48	–35 –(73)	24 (49)	48 (100)
	20	36	–52 –(142)	–14 –(39)	26 (71)	36	–49 –(136)	5 (14)	35 (97)	36	–48 –(136)	11 (30)	35 (99)
10 clusters with 100 per cluster and ICC = 0.2	1	51	–36 –(71)	2 (5)	39 (77)	50	–33 –(66)	21 (41)	50 (98)	50	–33 –(66)	26 (52)	50 (99)
	5	47	–41 –(86)	–5 –(10)	38 (80)	47	–38 –(81)	15 (32)	46 (99)	47	–38 –(81)	21 (45)	47 (100)
	20	36	–51 –(143)	–14 –(38)	27 (74)	35	–49 –(138)	5 (15)	35 (98)	35	–49 –(138)	11 (32)	35 (100)
10 clusters with 100 per cluster and ICC = 0.01	1	51	–36 –(71)	0 (0)	40 (78)	50	–34 –(67)	19 (39)	49 (98)	50	–34 –(67)	25 (51)	50 (100)
	5	47	–40 –(85)	–2 –(5)	37 (79)	47	–38 –(81)	17 (35)	46 (99)	47	–37 –(80)	22 (48)	47 (100)
	20	36	–53 –(147)	–13 –(36)	26 (73)	35	–50 –(140)	6 (18)	35 (98)	35	–49 –(139)	12 (33)	35 (100)

TABLE 24 *Bias if strong contamination dependence* b = 2, *individual post-cluster contamination with constant 50% contamination effect*

Number of clusters, number of members per cluster and ICC	*k*	α = 1				α = 2				α = 3			
		Individual simulation bias	Difference in bias between individual and cluster (%)			Individual simulation bias	Difference in bias between individual and cluster (%)			Individual simulation bias	Difference in bias between individual and cluster (%)		
			Cluster delay 1.49 (12%)	Cluster delay 4.45 (37%)	Cluster delay 6.65 (55%)		Cluster delay 1.86 (16%)	Cluster delay 6.95 (58%)	Cluster delay 11.95 (99%)		Cluster delay 1.92 (16%)	Cluster delay 8.1 (68%)	Cluster delay 12.84 (100%)
10 clusters with	1	26	0	6	21	26	0	8	25	26	0	7	25
10 per cluster and			(2)	(23)	(80)		(2)	(31)	(97)		(2)	(28)	(99)
ICC = 0.2	5	24	0	6	20	24	0	8	24	24	0	7	24
			(0)	(26)	(81)		–(1)	(33)	(99)		–(1)	(31)	(99)
	20	19	1	4	16	19	1	6	18	18	1	5	18
			(5)	(23)	(83)		(5)	(31)	(98)		(5)	(29)	(100)
5 clusters with	1	26	0	6	1	26	1	8	25	26	1	8	25
20 per cluster and			(1)	(25)	(94)		(2)	(30)	(96)		(3)	(29)	(99)
ICC = 0.2	5	24	0	6	21	24	1	7	23	24	1	7	24
			(1)	(23)	(80)		(6)	(30)	(97)		(6)	(28)	(100)
	20	19	0	4	20	19	1	5	18	18	1	4	18
			(1)	(19)	(83)		(7)	(26)	(98)		(6)	(23)	(100)
20 clusters with	1	26	0	5	21	26	0	7	25	26	0	7	25
5 per cluster and			(1)	(19)	(80)		(0)	(28)	(97)		(1)	(26)	(99)
ICC = 0.2	5	24	0	5	20	24	–1	7	24	24	–1	7	24
			–(2)	(22)	(81)		–(2)	(29)	(98)		–(2)	(27)	(100)
	20	19	0	5	15	19	0	6	18	18	0	5	18
			(0)	(26)	(82)		(0)	(32)	(96)		(0)	(30)	(99)
10 clusters with	1	26	0	6	21	26	0	8	25	26	0	7	26
10 per cluster and			(2)	(23)	(82)		(2)	(32)	(99)		(2)	(29)	(100)
ICC = 0.01	5	24	1	5	20	24	1	7	23	24	1	6	24
			(3)	(19)	(82)		(3)	(28)	(98)		(2)	(25)	(100)
	20	19	0	4	16	19	0	6	18	18	0	5	18
			–(1)	(23)	(83)		–(1)	(30)	(99)		–(2)	(29)	(99)

continued

TABLE 24 *Bias if strong contamination dependence* b = 2, *individual post-cluster contamination with constant 50% contamination effect (cont'd)*

Number of clusters, number of members per cluster and ICC	***k***	**α = 1**				**α = 2**				**α = 3**			
		Individual simulation bias	**Difference in bias between individual and cluster (%)**			**Individual simulation bias**	**Difference in bias between individual and cluster (%)**			**Individual simulation bias**	**Difference in bias between individual and cluster (%)**		
			Cluster delay 1.49 (12%)	**Cluster delay 4.45 (37%)**	**Cluster delay 6.65 (55%)**		**Cluster delay 1.86 (16%)**	**Cluster delay 6.95 (58%)**	**Cluster delay 11.95 (99%)**		**Cluster delay 1.92 (16%)**	**Cluster delay 8.1 (68%)**	**Cluster delay 12.84 (100%)**
5 clusters with	1	26	0	5	21	26	0	7	25	26	0	7	25
20 per cluster and			(2)	(20)	(82)		(1)	(28)	(97)		(2)	(27)	(99)
ICC = 0.01	5	24	1	5	20	24	0	8	24	24	0	7	24
			(2)	(22)	(81)		(2)	(32)	(98)		(2)	(28)	(99)
	20	19	0	4	15	19	0	6	18	18	0	5	18
			(2)	(23)	(79)		(2)	(33)	(97)		(2)	(30)	(100)
20 clusters with	1	26	0	5	21	26	0	7	25	26	0	7	26
5 per cluster and			(1)	(20)	(82)		(0)	(29)	(98)		(1)	(26)	(100)
ICC = 0.01	5	24	0	5	20	24	0	6	24	24	0	6	24
			(2)	(20)	(81)		(1)	(27)	(98)		(1)	(24)	(100)
	20	19	0	4	15	19	0	5	18	18	0	5	18
			(0)	(22)	(80)		(0)	(29)	(97)		(0)	(27)	(99)
10 clusters with	1	26	0	5	21	26	0	7	25	26	0	7	25
100 per cluster and			(0)	(21)	(80)		(0)	(28)	(96)		(0)	(26)	(98)
ICC = 0.2	5	24	0	6	20	24	0	7	24	24	0	7	24
			(1)	(23)	(83)		(1)	(30)	(99)		(1)	(27)	(100)
	20	18	0	3	15	18	0	5	18	18	0	4	18
			–(2)	(19)	(82)		–(2)	(28)	(99)		–(2)	(24)	(99)
10 clusters with	1	26	0	5	21	26	0	7	25	26	0	6	26
100 per cluster and			(2)	(19)	(82)		(2)	(27)	(98)		(2)	(25)	(100)
ICC = 0.01	5	24	0	5	19	24	0	7	23	24	0	6	24
			(0)	(22)	(79)		(0)	(30)	(97)		(0)	(27)	(100)
	20	18	0	4	14	18	0	5	18	18	0	5	18
			(0)	(21)	(78)		(0)	(28)	(97)		(0)	(26)	(99)

TABLE 25 *Bias if strong contamination dependence* b = 2, *individual post-cluster contamination with time-dependent contamination effect*

Number of clusters, number of members per cluster and ICC	*k*	α = 1				α = 2				α = 3			
		Individual simulation bias	Difference in bias between individual and cluster (%)			Individual simulation bias	Difference in bias between individual and cluster (%)			Individual simulation bias	Difference in bias between individual and cluster (%)		
			Cluster delay 1.49 (12%)	Cluster delay 4.45 (37%)	Cluster delay 6.65 (55%)		Cluster delay 1.86 (16%)	Cluster delay 6.95 (58%)	Cluster delay 11.95 (99%)		Cluster delay 1.92 (16%)	Cluster delay 8.1 (68%)	Cluster delay 12.84 (100%)
10 clusters with 10 per cluster and ICC = 0.2	1	51	7 (13)	25 (49)	45 (89)	50	8 (16)	35 (69)	50 (99)	50	8 (17)	37 (75)	50 (100)
	5	49	8 (16)	25 (50)	45 (91)	49	9 (19)	34 (69)	49 (100)	48	9 (19)	36 (75)	48 (99)
	20	36	6 (15)	18 (49)	33 (92)	36	7 (19)	25 (69)	36 (100)	36	7 (19)	27 (75)	35 (99)
5 clusters with 20 per cluster and ICC = 0.2	1	51	7 (15)	24 (47)	46 (91)	50	9 (17)	34 (68)	50 (100)	50	9 (18)	37 (74)	50 (99)
	5	49	7 (14)	25 (50)	44 (90)	49	8 (17)	34 (71)	48 (99)	48	9 (18)	37 (76)	48 (99)
	20	36	3 (9)	19 (53)	33 (91)	36	5 (13)	26 (72)	36 (100)	36	5 (14)	28 (78)	35 (99)
20 clusters with 5 per cluster and ICC = 0.2	1	51	6 (13)	25 (48)	45 (89)	50	8 (16)	34 (68)	50 (99)	50	8 (16)	37 (74)	50 (99)
	5	49	7 (15)	25 (51)	45 (92)	49	9 (18)	34 (70)	48 (100)	48	9 (18)	37 (76)	48 (100)
	20	36	4 (11)	18 (50)	32 (89)	36	5 (14)	25 (69)	35 (98)	36	5 (15)	27 (75)	35 (99)
10 clusters with 10 per cluster and ICC = 0.01	1	51	7 (13)	25 (49)	45 (88)	50	8 (16)	35 (69)	49 (98)	50	8 (17)	38 (75)	50 (99)
	5	49	8 (15)	24 (49)	44 (88)	49	9 (18)	34 (69)	48 (99)	48	9 (19)	37 (75)	48 (100)
	20	36	5 (14)	18 (50)	33 (90)	36	6 (17)	25 (70)	35 (99)	36	6 (18)	27 (76)	36 (100)

continued

TABLE 25 *Bias if strong contamination dependence* b = 2, *individual post-cluster contamination with time-dependent contamination effect (cont'd)*

Number of clusters, number of members per cluster and ICC	***k***	**α = 1**				**α = 2**				**α = 3**			
		Individual simulation bias	**Difference in bias between individual and cluster (%)**			**Individual simulation bias**	**Difference in bias between individual and cluster (%)**			**Individual simulation bias**	**Difference in bias between individual and cluster (%)**		
			Cluster delay 1.49 (12%)	**Cluster delay 4.45 (37%)**	**Cluster delay 6.65 (55%)**		**Cluster delay 1.86 (16%)**	**Cluster delay 6.95 (58%)**	**Cluster delay 11.95 (99%)**		**Cluster delay 1.92 (16%)**	**Cluster delay 8.1 (68%)**	**Cluster delay 12.84 (100%)**
5 clusters with 20 per cluster and ICC = 0.01	1	51	6 (11)	25 (48)	46 (90)	50	7 (14)	34 (68)	50 (100)	50	8 (16)	38 (75)	50 (100)
	5	49	8 (16)	25 (51)	44 (90)	49	9 (19)	34 (70)	48 (99)	48	9 (19)	37 (77)	48 (100)
	20	36	5 (14)	17 (48)	32 (89)	36	6 (17)	24 (67)	35 (99)	36	6 (16)	27 (75)	36 (100)
20 clusters with 5 per cluster and ICC = 0.01	1	51	6 (13)	25 (49)	46 (91)	50	8 (16)	35 (69)	50 (100)	50	8 (16)	38 (75)	50 (100)
	5	49	7 (15)	26 (53)	44 (90)	49	9 (18)	35 (72)	48 (99)	48	9 (19)	37 (77)	48 (100)
	20	36	5 (13)	18 (49)	33 (91)	36	6 (16)	24 (68)	36 (100)	36	6 (16)	27 (75)	36 (100)
10 clusters with 100 per cluster and ICC = 0.2	1	51	6 (12)	25 (49)	45 (89)	50	8 (15)	34 (68)	50 (99)	50	8 (16)	37 (74)	50 (99)
	5	47	5 (11)	23 (49)	42 (89)	47	7 (14)	32 (69)	46 (99)	47	7 (15)	35 (75)	46 (100)
	20	36	3 (9)	17 (48)	32 (90)	35	4 (13)	24 (69)	35 (99)	35	5 (13)	26 (75)	35 (100)
10 clusters with 100 per cluster and ICC = 0.01	1	51	6 (12)	25 (50)	46 (89)	50	8 (16)	35 (70)	50 (99)	50	8 (16)	38 (76)	50 (100)
	5	47	6 (12)	24 (50)	42 (90)	47	7 (15)	33 (70)	46 (99)	47	7 (16)	35 (76)	47 (100)
	20	36	5 (13)	18 (50)	32 (90)	35	6 (16)	25 (70)	35 (99)	35	6 (17)	27 (76)	35 (100)

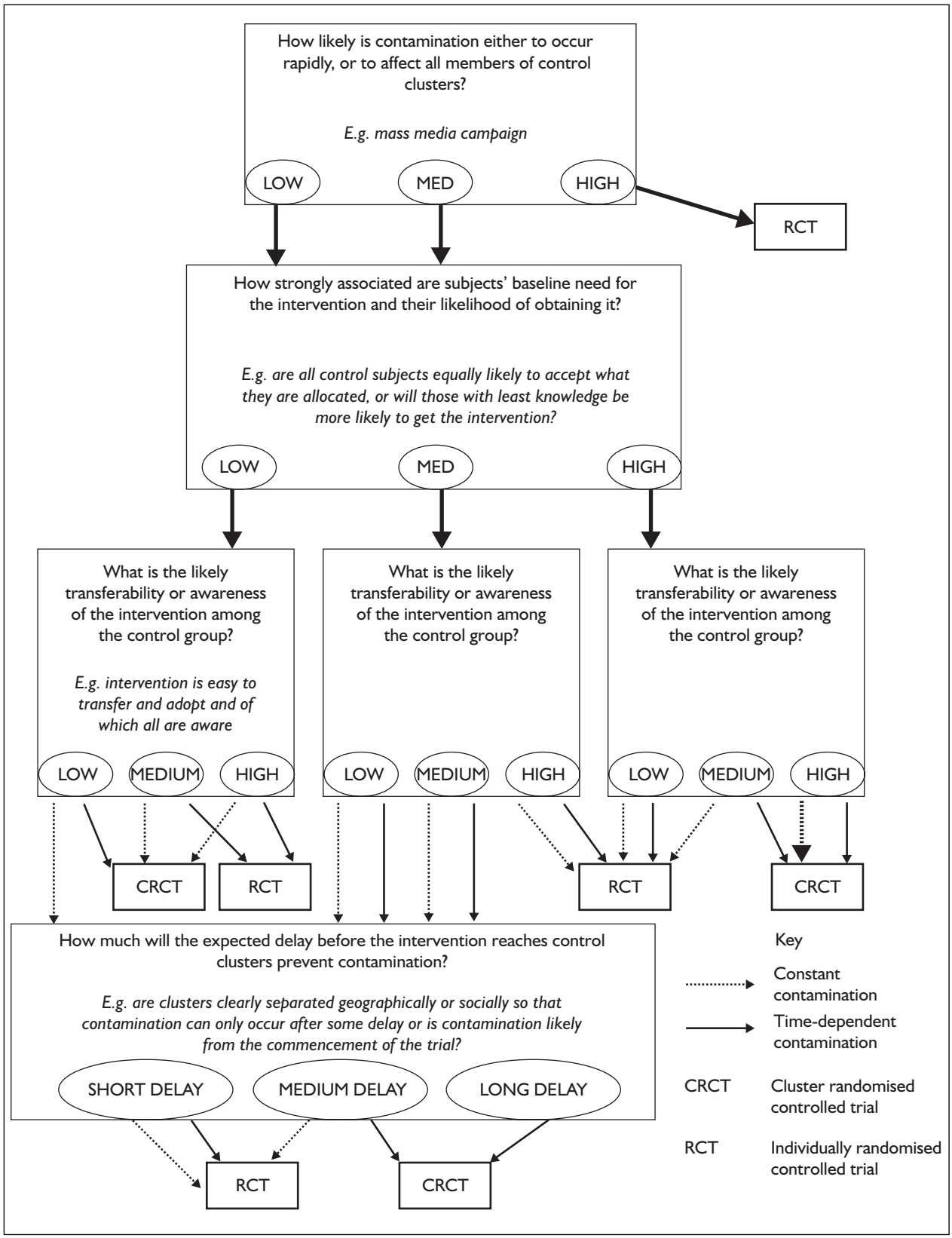

FIGURE 30 *Flowchart guide to cluster or individually randomised design*

The results of these simulations agree broadly with the results of our Delphi questionnaire. The Delphi responses concluded that, for cluster randomised designs, contamination was highly likely where both the intervention and control health professionals were part of the same workplace team (rank 2): a situation that may lead to early contamination of all or a large proportion of the cluster. A similar rapid contamination of a large proportion of a cluster might occur when a health professional moves from the intervention to the control group (rank 4).

The simulation parameters do not translate exactly to physical aspects of trial design. However, the results can be used as an approximate guide with regard to situations in which cluster randomisation is favourable and situations where it is not. The flowchart in *Figure 30* can be used as an approximate guide in making such a decision.

The first stage of the flow chart determines whether the proposed trial is similar to the 'all cluster members contaminated' simulation or whether individual contamination occurs after cluster contamination. The second stage is analogous to the parameter b and assesses the influence of baseline education on the likely response of subjects. The third stage relates to the parameter k determining the speed or ease with which contamination can occur in individuals. The final stage determines the delay in cluster contamination that the proposed design will incur.

These results are based on simulated data that may not accurately reflect what happens during real-world trials. It would therefore be valuable if future individual and cluster randomised trials of educational interventions could measure and analyse the process of contamination and its effects over time.

Chapter 5

Dealing with contamination in randomised controlled trials using Causal Average Effect Analysis

Background

Randomisation ensures baseline equivalence between two or more groups that are formed, apart from chance differences. It therefore controls for both observed and unobserved confounders and provides a basis for statistical inference. It does not address other biases that may occur after random allocation. In this report we consider the effects of one of these biases, namely contamination of one of the treatment groups.

Origins of randomised trial analysis

Modern randomised trials had their origins in work in the 1920s in the area of experimental agriculture by Fisher.[234] It was not until the late 1940s that randomisation was adapted, by Bradford Hill, for use in experimental medicine.[235] Although participants in medical trials are clearly equivalent to the agricultural plots studied in Fisher's work, the order of the randomisation procedures is slightly different. In agricultural research, experimental units are identified in advance of the trial being designed, the allocation structure is then developed and finally the pattern of treatment is imposed. The situation is usually reversed in medical research, with the randomisation sequence generated first and then the patients recruited.[236] Furthermore, the participants in the trial (i.e. plots of land or plants) cannot refuse the allocated therapy. Finally, in the context of this review, contamination between crops is less of a problem as application of fertiliser can be applied carefully so as to avoid contaminating control plots.

The analysis of agricultural trials is relatively straightforward with all allocated plots receiving the treatment they were allocated to and then being included in an ITT analysis.

An additional difference between agricultural and medical research is that medical researchers are interested in effects on individuals, whereas agricultural researchers are not. Twice as much grain can come from twice as many plants surviving, or from each plant doubling its output. The difference is irrelevant in agricultural research, but important in medical research. In contrast, although it is important to policy makers and public health evaluations to be interested in the group mean or group totals, the effects of treatment at an individual level are also of interest to the clinician and the patient. For example, if this patient takes a given treatment, as prescribed, what is the likely impact on their condition? Trials with contamination and that use ITT analysis do not answer this question. Rather, human trials answer the question as whether the **offer** of treatment to the intervention population is effective, which is rather different.

One approach that is widely used to deal with the problem of contamination is to use 'per protocol' or 'on-treatment' analysis. Per protocol analysis is where patients, who either do not adhere to their treatment or who access the experimental intervention, are discarded from the analysis. Because this approach results in a loss of statistical power due to a sample size reduction, some prefer to use an on-treatment approach. This is where contaminated participants are retained in the analysis but are assigned to a treatment-received group. However, for both of these analyses to be valid the participants who do not adhere have to be a random sample of all those participants who were offered treatment. This is rarely true.[237] Consequently, it is possible, indeed likely, that they differ in ways that could be associated with outcomes. If so, this will lead to bias. There are now classic examples which emphasise the pitfalls of these types of analyses.[238–240]

This drawback of per protocol or on-treatment analysis has long been recognised. Therefore, the use of ITT analysis was recommended as being the principal analytical approach by Bradford Hill in the seventh edition of his textbook *Principles of Medical Statistics*.[241] However, the concept of analysing participants as randomised was reported earlier.[242] The main reason for advocating ITT is that it maintains the baseline comparability achieved by randomisation, unlike per protocol and on-treatment analyses.[237] If the initial randomisation process is undermined, then

confounding can be introduced and consequently the internal validity of the results is questionable. In addition, an ITT analysis is completely objective because it guards against any conscious or unconscious decisions that have to be made if unexpected outcomes are observed. Many authors also note that as an ITT analysis essentially focuses on the effect of a change in treatment policy, it mirrors what would happen in actual clinical practice.[243] Consequently, comparing the outcomes of all those randomly allocated to treatment A with all patients allocated to treatment B allows policy decisions to be made on whether drug A could actually replace drug B in clinical practice.

Although an ITT analysis does, in principle, provide an unbiased estimate of offering treatment to all trial participants, it will, in the presence of non-adherence or contamination, result in a conservative or diluted estimate of the treatment effect for patients who adhered to treatment recommendations. ITT is a conservative approach because ITT sacrifices the Type II error (false negative, that is, power is reduced) rate while controlling for Type I error (false positive) rate. So, although ITT guards against selection bias, it does so at the cost of introducing 'dilution' bias and thus reduces the power of the study. Nevertheless, ITT will not be biased in the direction of showing an effect when there is no effect, apart from unusual circumstances such as dispensing errors leading to treatments accidentally being swapped.[244] However, the possibility of showing an effect when there is no effect (Type I error) cannot be excluded with per protocol or on-treatment analysis.

There are several approaches that could be undertaken to reduce the problem of contamination between treatment groups. Where contamination occurs because a treatment is easily passed on to the control group, such as a health education intervention, cluster randomisation can be undertaken. Physically separating groups of participants will reduce and sometimes abolish the threat of contamination, although this will not deal with the issue of non-adherence, which is a second form of contamination. There are a number of drawbacks to cluster randomisation. First, the sample size of the trial has to be increased to take into account the clustered nature of the data.[245] Second, cluster trials can be difficult to undertake properly and many suffer from methodological errors, such as recruitment bias, that can undermine the internal validity of the trial.[246–248] An alternative that has been suggested is to retain individual randomisation and accept some contamination. This has the advantage that for a given sample size a smaller difference between groups could be observed in an individually randomised trial compared with a cluster study. This smaller difference would take account of some dilution due to contamination. Indeed, the power advantages of individual randomisation are not eroded until contamination is fairly high (>30% in some instances).[249] There is a problem with this approach, which relates to the difference between statistical significance, clinical significance and cost-effectiveness. The advantages of individual randomisation that have been discussed so far relate to statistical significance. Even if the difference between the groups is statistically significant, it may not be clinically significant – the effect size is reduced due to dilution bias. Furthermore, even if clinically significant, the difference may not be cost-effective, but may have been cost-effective had we been able to observe the 'true' difference unbiased by contamination. Therefore, in the likely presence of contamination, some trialists may still feel justified in undertaking a cluster randomised controlled trial in order to gain a less precise but unbiased estimate instead of a more precise but diluted estimate, as would occur if ITT were used in an individual randomised trial where there was significant contamination. To retain individual randomisation when we know there has been some contamination, we require a statistical method that produces an unbiased estimate of effect.

Recently, a statistical method has been developed which is known as the Complier Average Causal Effect (CACE) or, in the economics literature, the Local Average Treatment Effect (LATE).[250–253] This is a measure of the average causal effect for the subpopulation of adherers and, as it preserves the initial randomisation it overcomes the problems faced by per protocol analyses. This approach has been advocated recently in the statistical literature,[254] but at present is not widely used. The method used for non-adherence can also be used for contamination, because both problems entail misallocation of exposures to trial subjects.

We will illustrate this approach with a trial of faecal-occult-blood screening for the prevention of colorectal cancer (CRC),[255] which is shown in *Figure 31*. The approach we use to deal with non-adherence could also be used to deal with contamination if the prevalence of contamination was known. This very large randomised trial showed a reduction of 15% in CRC mortality using

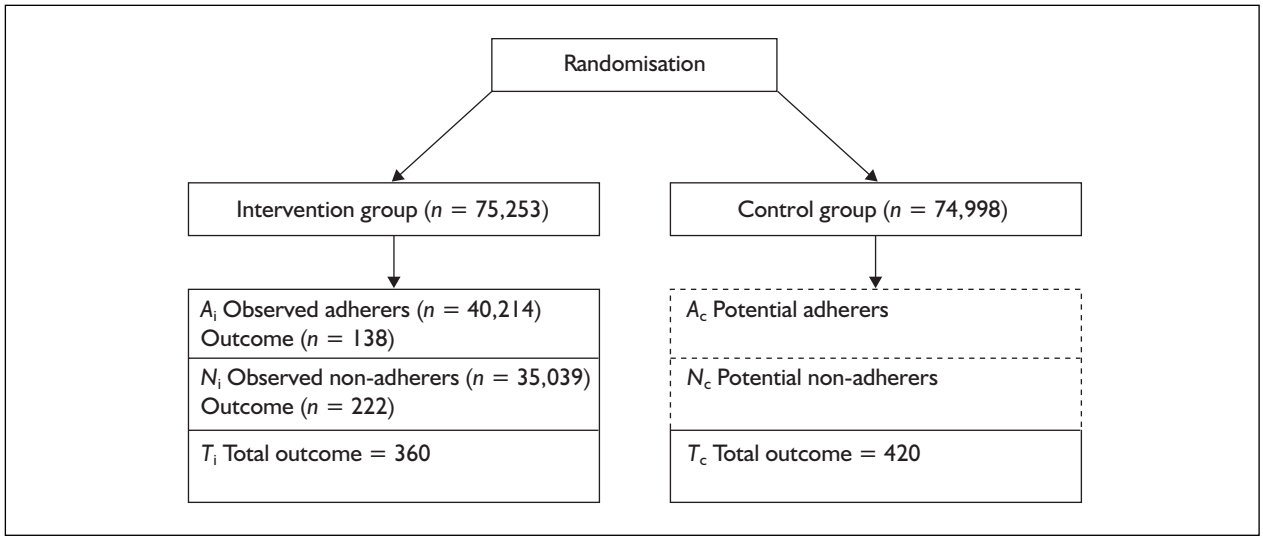

FIGURE 31 *Treatment profile of the faecal-occult-blood screening study*

ITT. However, there was a 47% non-adherence rate for the first screening round. Therefore, there is a fairly large dilution effect of screening mortality. Although the trial answers the question of 'What would be the reduction in CRC mortality if we offered population screening', it does not answer the question of what the reduction in mortality would be if those who were offered screening took up the offer. The authors of the study realised the problem of contamination and therefore undertook a per protocol analysis, where they found a much larger benefit on deaths due to CRC (39%). However, as the authors opted for a per protocol analysis, this may be an overestimate of the effect due to selection bias. In *Figure 31* we show the treatment profile of the faecal-occult-blood screening study and demonstrate how CACE analysis may work.[256]

From *Figure 31*, it can be seen that in the intervention group there are two subgroups of participants: those who take the treatment and those who do not. One of the first assumptions that we make about the CACE is that the control group, had they been offered the treatment, would have contained the same proportion of non-adherers. Because we are using random allocation, this statement must be true (except for chance imbalances). The second assumption of CACE is that merely being offered the treatment has no effect on outcomes. If we accept that both of these assumptions are true, then to obtain an unbiased estimate of any treatment effect all we need to do is to compare the observed outcomes in group A_i with the unobserved outcomes in group A_c. We can do this as follows.

From *Figure 31*, we can see that 47% of the participants in the intervention group refused screening and that 222 events occurred in this population (i.e. 0.63%). We can also observe that for the 40,214 participants that accepted the screening 138 had an event (i.e. 0.34%). For the control group, we cannot directly observe the event rates in the potential adherers and non-adherers groups. We do know, however, that the total number of participants with the outcome of interest was 420 in the control group. We can assume that the control group will also contain 47% of participants who, if offered screening, would refuse. If we assume that the offer of treatment has no effect on the outcome, then we can calculate that 222 of the 420 events must have occurred among the 34,920 potential non-adherers in the control group. This, then, leaves 198 remaining events that would have occurred in the potential adherers group. We can now compare the outcomes of those who accepted the screening with a similar group who would have accepted screening if they had been offered it. In *Table 26* the relative risks of the various approaches are compared.

As *Table 26* shows, the ITT analysis produces the highest relative risk and the per protocol approach the lowest. The CACE method produces an estimate that falls between these two extremes.

This method has been expanded to allow for the possibility of contamination, that is, use of the intervention in the control group.[257] To demonstrate this, let us suppose that 20% of the control group had in fact received faecal-occult-

TABLE 26 *Comparison of relative risks between ITT, per protocol and CACE analysis for the colorectal cancer screening study*

	Intervention (n = 75,253)	**Control (n = 74,998)**
Adherers (53%)	A_i 138/40,214 = 0.34%	A_c 198/39749 = 0.50%
Non-adherers (47%)	N_i 222/35,039 = 0.63%	N_c 222/35,249 = 0.63%
Total outcome	T_i 360/75,253 = 0.48%	T_c 420/74,998 = 0.56%
ITT analysis	T_i/T_c = 0.85 (0.48%/0.56%)	
Per protocol analysis	A_i/T_c = 0.61 (0.34%/0.56%)	
CACE analysis	A_i/A_c = 0.69 (0.34%/0.50%)	

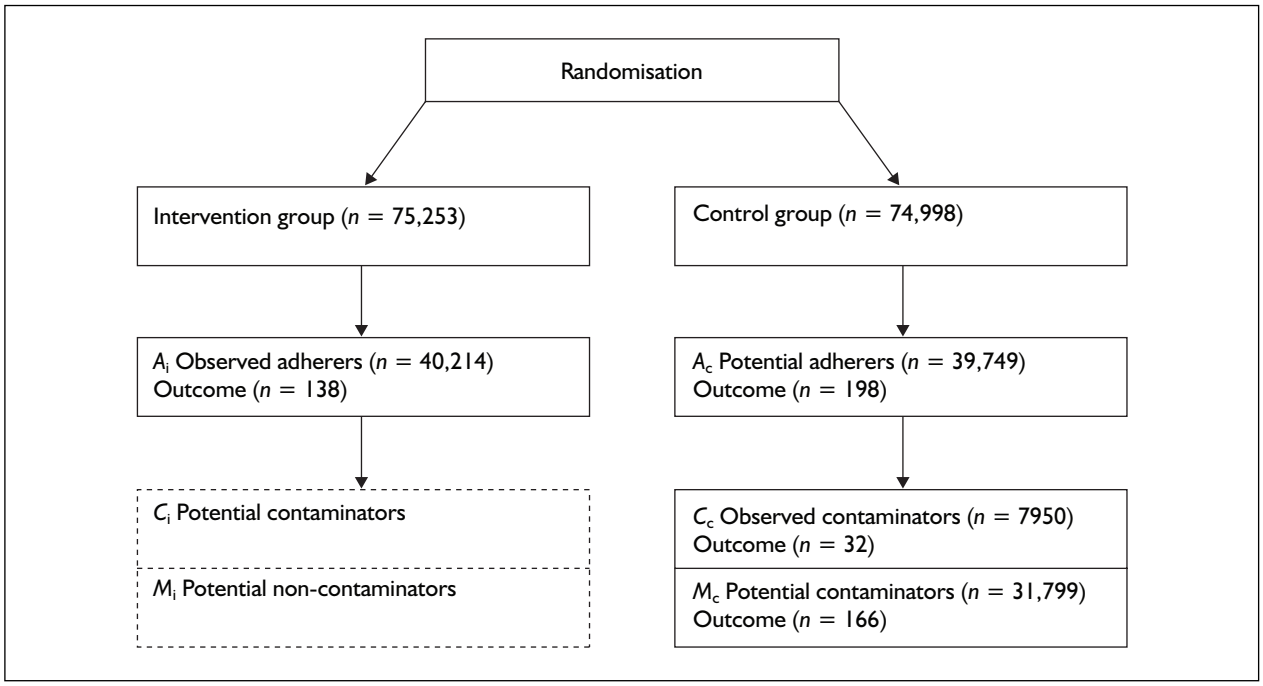

FIGURE 32 *Hypothetical treatment profile of the faecal-occult-blood screening study*

blood screening and also that 32 events occurred in this group. *Figure 32* shows a hypothetical extension of the profile of the faecal-occult-blood screening study.

From *Figure 32*, we can see that among the potential adherers in the control group there are two subgroups of participants: those who seek screening even though they have not been randomised to receive screening and those who do not. Let us assume that we observe that 20% of participants in fact received faecal-occult-blood screening in the control group and that 32 events occurred in this population (i.e. 0.4%). We further assume that we observe that for the 31,799 participants who did not seek screening in the control group 166 had an event (i.e. 0.52%). Using a similar argument as for the situation of non-adherence, we can assume that the intervention group will also contain 20% of participants who, if allocated to the control group, would seek the intervention. Consequently, we can calculate that 32 of the 138 events must have occurred among the potential contaminators in the intervention group. This then leaves 106 remaining events that would have occurred in the potential non-contaminators group.

Table 27 shows that the ITT analysis produces the highest relative risk and the CACE approach the lowest. The per protocol analysis produces an estimate that lies between the two.

This was the simplest approach to CACE, but many variables such as age will predict adherence. Using these predictors in a regression analysis will improve any point estimate because an unadjusted CACE estimate suffers from two sources of random variation: the initial random allocation and the random variation within the study arms. Also, using such predictors will enable a reduction in the variance surrounding the CACE estimators

TABLE 27 *Comparison of relative risks between ITT, per protocol and CACE analysis for the hypothetical extension of the colorectal cancer screening study*

	Intervention	Control
Non-contaminators (80%)	M_i 106/32,171 = 0.33%	M_c 166/31,799 = 0.52%
Contaminators (20%)	C_i 32/8,043 = 0.4%	C_c 32/7,950 = 0.4%
Overall total outcome	T_i 360/75,253 = 0.48%	T_c 420/74,998 = 0.56%
ITT analysis	T_i/T_c = 0.85 (0.48%/0.56%)	
Per protocol analysis	A_i/M_c = 0.66 (0.34%/0.52%)	
CACE analysis	M_i/M_c = 0.63 (0.33%/0.52%)	

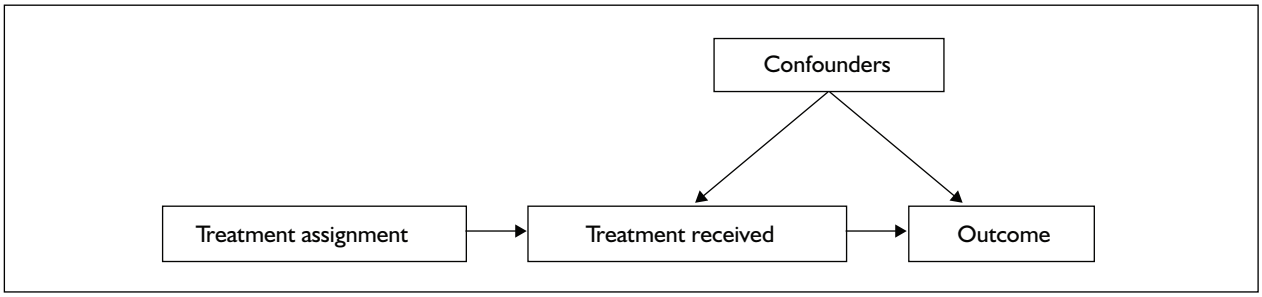

FIGURE 33 *Contamination or non-adherence (treatment received differs from treatment assignment) in a typical trial*

to be achieved; this is similar to the normal adjusted two group analysis done routinely in many randomised trials. There are a number of statistical approaches to implementing CACE, which do take account of the relationship between predictors of adherence and outcome. In the following sections, we have undertaken a review of the different methods.

Literature review of statistical methods to take account of contamination

Electronic searches were used to identify articles that reported statistical methods to take account of contamination in trials. A snowballing technique was used to supplement the electronic searches. References of studies that were included in the review were checked and the 'cited by' function was also used to find additional studies.

Results

The search identified a number of papers that described different methods or approaches to CACE. The key literature identified from the search is summarised below to give a brief overview of the history of the techniques; however, a full list of all of the included studies is available on request from the authors.

Bloom, in the field of educational research, was the first author we identified who reported methods of adjusting estimates of treatment effects to allow for non-adherence.[251] Bloom concentrates on the problem of 'no-shows', when some participants in the treatment group fail to take the treatment offered. In this simple instrumental variable (IV) approach, the treatment effect estimates are adjusted by considering the proportion of non-adherers in the treatment group. In the late 1980s and early 1990s, a number of articles were published[258–263] that incorporated or expanded on this early work by Bloom.[251] An important milestone in the formalisation of the IV technique was a publication by Angrist and colleagues in 1996.[250] They outlined a framework for the derivation of an estimator based on Rubin's Causal Model (RCM)[264,265] that allowed simple and easily interpretable assumptions to be stated, which were hidden in previous work (e.g. Bloom).

The IV approach has been widely used in economics when random allocation is not possible.[266] However, within an RCT with non-adherence, treatment assignment (that is, the groups to which the participants are randomised) provides a perfect IV for confounding control. *Figure 33* illustrates the problems of contamination in a typical trial.

We are interested in estimating the effect of the treatment on the outcome. However, when some participants fail to receive their allocated treatment, this relationship is likely to be

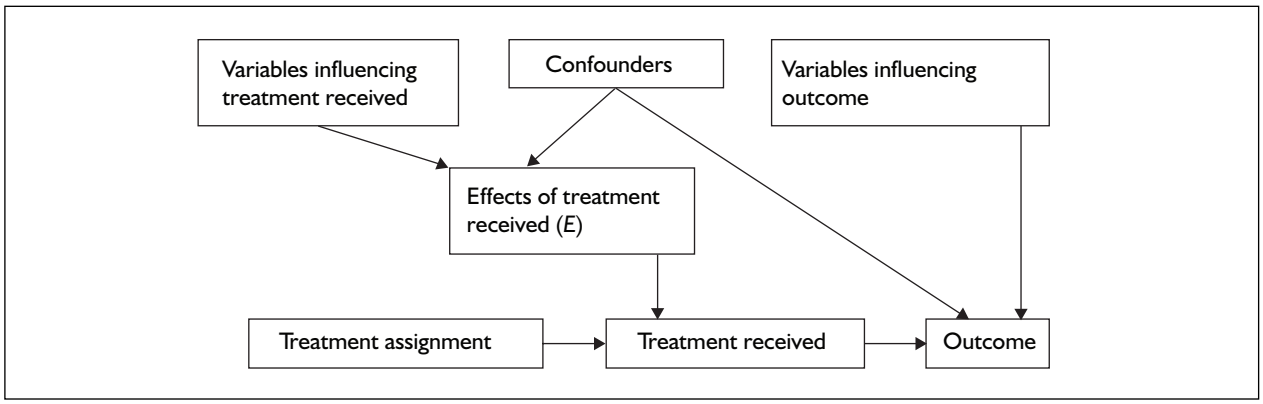

FIGURE 34 *An approach to the problems of contamination or non-adherence (treatment received differs from treatment assigned) in a typical trial*

confounded by variables that influence the treatment received and also the outcome. If it were possible to identify all of the confounders then the analysis could be conditioned on these variables, thereby adjusting for the selection effects that usually occur when on-treatment analysis is used. However, it is likely, indeed probable, that some of the confounders will be unmeasured or even unimagined. Hence it is unlikely that the effect of the treatment on the outcome can be measured in an unconfounded manner using traditional methods. Instrumental variable analysis requires the identification of a variable that is independent of all of the confounders, that is, associated with the treatment received and has no direct effect on the outcome itself.[267] Since the groups are randomised, the first assumption should be satisfied. We would also hope that the treatment the participants receive is affected by the treatment to which the participants are randomised; as there is some non-adherence, the treatment received would not be fully determined by the treatment assignment. Finally, we assume that merely being offered the treatment has no effect on outcomes. When these assumptions are met, the treatment assignment acts as an IV and the treatment effect can be estimated using the equation originally proposed by Bloom[251] or alternatively by implementing a two-stage least squares (2SLS) regression. The 2SLS regression can be performed in most statistical software; however, it involves regressing the treatment received on the treatment assigned and saving the predicted values, and then regressing the outcome on the predicted values. The estimand of interest is the predicted value coefficient.

If we have a binary outcome then the IV procedure has been shown to produce reasonable estimates using 2SLS regression, but bootstrapping is required to produce equivalent estimates for the standard error.[268] Sommer and Zeger were the first to apply the IV methods in the biostatistics literature for binary outcomes;[256] although they state that Tarwotjo and colleagues proposed the methods used.[263] They derived risk ratio estimators for the effect of the treatment among the adherers in the intervention group and the 'would be' treatment adherers in the control group. Cuzick and colleagues[257] extended the method proposed by Sommer and Zeger[256] to allow for the possibility of contamination and demonstrated this technique using data from screening and prevention trials.

Nagelkerke and colleagues presented an IV approach for dealing with non-adherence that is slightly different to the IV approach described above.[252] However, similar estimates are obtained to those when using Bloom's equation directly or by implementing a 2SLS regression, and also for odds ratios or risk ratios similar estimates to those obtained by the Sommer and Zeger method.

From *Figure 34*, we can see that the treatment assignment will affect the treatment received, which will in turn affect the outcome of the trial for each individual. However, the treatment received will also be affected by other factors; some of them will be random (e.g. error) and others will be confounding variables, which affect the treatment received and the outcome (e.g. age, distance). Nagelkerke and colleagues defined a variable *E*, which represents all of the other affects on the treatment received except the treatment assignment. This variable is used *in lieu* of having the real confounders and included in a multivariate analysis to adjust for the confounding.

This method has the advantage that most types of regression (linear, logistic, Poisson or Cox

proportional hazards model) can be used in the analysis. Irrespective of the method used, the treatment received variable and *E* (that is, all other effects on the treatment received except treatment assignment) are included as covariates in the analysis. To implement this method, a two-stage procedure is utilised. First, the treatment received variable is regressed on the treatment assignment variable and the residuals are saved; the residuals represent the variable *E*. Finally, the outcome is regressed on the treatment received variable and the saved residuals from the first regression. The estimand of interest is the coefficient of the treatment received variable, as in a conventional analysis.

Likelihood-based estimation is another area of research that has been explored to derive estimates of CACE. Imbens and Rubin were among the first to propose maximum likelihood (ML) methods using the expectation maximisation (EM) algorithm and Bayesian inference methods using the data augmentation (DA) algorithm for calculating CACE.[269,270] The ML method has subsequently been extended to incorporate covariates that are predictive of the outcome and adherence to treatment.[271] In these methods, the unknown adherence status in the control group is treated as missing data. However, it has also been proposed that the unknown adherence status of the control group could be treated as a latent variable within a structural equation model framework.[272] The adherence status is treated as a categorical latent variable and the ML–EM algorithm is used to estimate the unknown adherence status of each individual participant in the control group.

Application of the CACE approach to trial data

To assess the utility of the CACE approach, we applied one of the methods to two recently published randomised trials, one looking at the role of hip protectors for hip fracture prevention and the other looking at dietary intervention for the treatment of irritable bowel syndrome (IBS). We chose these two trials for the following reasons: first, we had access to the individual level data for each trial; second, both trials suffered from a high level of non-adherence. In this section, we emphasise methods to deal with participants who are randomised to receive the intervention but who fail to adhere to the treatment and subsequently 'access' the control group treatment. As was noted in the section 'Direct evidence of contamination reported in trials of educational interventions' (p. 9), a large number of trials report some degree of non-adherence whereas very few report contamination. The methods we show here to deal with non-adherence could also be used to deal with contamination. Finally, we wanted an example with a continuous outcome measure and another with a binary measure.

In this work, we primarily used Nagelkerke and colleagues' IV approach to calculate the CACE[252] and used SPSS to implement this. Nagelkerke and colleagues' method is the simplest to apply with commonly available statistical methods and software (e.g. SPSS), and takes account of non-random non-adherence. Other approaches using slightly different statistical techniques generated similar results when applied to these two trials.

Trial 1: hip protector trial

The primary care hip protector trial is the largest published randomised trial of hip protectors to date.[273] The main aim of this study was to assess whether hip protectors used among women living in the community and at high risk of hip fracture led to a reduction in hip fracture. The study included 4169 women, aged 70 years and older, with one or more risk factors for hip fracture (i.e. low body weight, current smoker, a prior fracture, family history of hip fracture). The women in the intervention group were mailed three pairs of hip protectors with instructions on how to use them and the control group received routine care. Adherence with the hip protectors was poor, with only 38% of participants reporting that they wore them on a daily basis at 12 months. A total of 4129 participants were analysed in this paper after listwise deletion of cases that had missing data. The hip fracture rate among the control group was 2.4% and among the intervention group 2.8% (i.e. no difference between the groups). A key criticism of the trial is that the observed lack of benefit of the hip protectors could have been because of poor adherence, which masked an important effect among women who did use the hip protectors as instructed.

In *Table 28* we show the results from the analyses of the hip protector data using three different approaches, that is ITT, per protocol and Nagelkerke and colleagues' CACE approach. Although none of the results were statistically significant, we can see that for the unadjusted analysis the odds of experiencing a hip fracture were slightly elevated in the protector group for the ITT analysis. In the perprotocol analysis, the risk of hip fracture appears to be reduced slightly

TABLE 28 *Hip protector trial*

	Odds ratio	Lower 95% CI	Upper 95% CI	*p*
Unadjusted				
ITT	1.21	0.81	1.81	0.355
Per protocol	1.14	0.63	2.04	0.664
CACE	1.64	0.57	4.73	0.358
Adjusted[a]				
ITT	1.19	0.79	1.78	0.401
Per protocol	1.10	0.61	1.98	0.751
CACE	1.50	0.55	4.09	0.425

[a] For outcome confounders, adjustments were age, prior fracture, history of falling and volunteer status, whereas for predictors of adherence we adjusted for weight, smoking status and volunteer status. Adjustments were made using individual patient data.

relative to the ITT analysis, from 1.21 to 1.14. However, the CACE results show a relative increased risk of hip fracture, which is greater than with either ITT or per protocol analysis (i.e. 1.64).

In *Table 28* we also present adjusted results. In theory, we can make CACE more precise if we identify not only the usual covariates for outcome but also predictors of contamination. In this instance, however, despite having statistically significant predictors of adherence, the width of the adjusted CIs for the adjusted CACE analysis, although smaller than the unadjusted CACE CIs, was still substantially wider than the ITT CIs.

By using the CACE approach, we can see that even among women who claim to have been wearing hip protectors 'most of the time' there is no evidence of any benefit. Therefore, in this instance the CACE approach strengthens the usual ITT method of confirming a lack of benefit of hip protectors, which suggests that non-adherence was not an explanation for the negative findings of this trial.

Trial 2: dietary trial for irritable bowel syndrome

The aim of the second randomised trial was to assess the therapeutic potential of dietary elimination based on the presence of IgG antibodies to food among outpatients with IBS.[274] Participants were randomised to receive either a true diet that excluded all foods to which they had raised IgG antibodies or a sham diet excluding the same number of foods but not those to which they had antibodies. Adherence with the diet was better than the hip protector trial, as 70% of participants reported that they adhered 'moderately well' or 'completely' to the diet at 12 weeks. Nevertheless, a substantial proportion of participants did not adhere to the diet as instructed. Therefore, the ITT estimates are likely to underestimate the benefit of dietary change on the symptoms of IBS for those who adhere to the diet.

Change between groups in IBS symptom severity score at 12 weeks, after adjusting for baseline severity, was the main outcome measure. *Table 29* shows the results from the IBS trial using the three different approaches. Whereas the ITT analysis suggests a small, statistically significant effect of the true diet, the per protocol analysis suggests a somewhat larger effect, which is highly statistically significant. The CACE estimate lies between the ITT and per protocol estimates. Note also in this analysis that whereas both CACE and ITT are statistically significant in their unadjusted analyses, only CACE comes close to statistical significance in the adjusted analysis.

In this instance, the clinician can now advise the patient that if they adhere to the dietary recommendations they can expect an approximately 50-point improvement in symptom severity. This contrasts with the ITT approach, which would have underestimated the benefit for the adherent participant, and the per protocol estimate, which overestimates its effectiveness. Note, however, that CACE produces the widest CIs of the three approaches as it has two sources of sample variance.

Summary of findings of examples

The application of CACE to these two trials produced estimates of effect that we would expect with complete adherence. For the IBS expected study, our prior assumption was that ITT would underestimate the diet's effectiveness whereas per

TABLE 29 *IBS trial data*

	Estimate	Lower 95% CI	Upper 95% CI	*p*	Effect size
Adjusted for baseline severity only					
ITT	33.67	0.51	66.83	0.047	0.36
Per protocol	60.48	23.96	97.00	0.001	0.66
CACE	49.22	4.14	94.31	0.033	0.56
Adjusted[a]					
ITT	26.32	–6.74	59.38	0.118	0.30
Per protocol	53.35	15.50	91.20	0.006	0.60
CACE	40.47	–1.43	82.37	0.058	0.48

[a] For outcomes, we adjusted for baseline severity, symptom duration, proton pump inhibitor use, constipation predominant group, total non-colonic features at visit 1 and the anxiety items of the Hospital Anxiety and Depression questionnaire, whereas for adherence, we adjusted for baseline severity, sex, symptom duration and whether the participant was taking antispasmodic drugs.

protocol analysis would overestimate the treatment's effects. CACE, as expected, produced an estimate that fell between these. In contrast, in the hip protector trial, the CACE method showed an increased risk of hip fracture using hip protectors, albeit not statistically significant.

Note that one important feature in these two examples is that CACE tends to produce relatively wide CIs. Thus, if we were in a situation of having to choose between cluster randomisation and individual allocation, then the CACE approach may not offer much improvement in precision for a given sample size. Nevertheless, choosing an individual randomised trial with an *a priori* specification of undertaking a CACE analysis rather than a cluster trial lets us avoid some of the methodological difficulties associated with undertaking cluster trials.

It seems to us that by using CACE methods we can avoid the trade-off between individual and cluster randomisation in terms of a biased estimate. Both approaches however lose precision with a given sample size. There will be trade-offs between the methods depending upon the size of ICC, the cluster size and the likely contamination rate. Furthermore, there is a trade-off in terms of variables that can predict adherence. If we can identify strong predictors of adherence then some of the imprecision of CACE can be avoided. In the following section we have undertaken a simulation study describing some of these likely trade-offs.

Simulation study

As noted earlier, one solution to the potential problem of contamination is to randomise clusters rather than individuals. However, when clusters are randomised we suffer reduced power due to the design effect, or variance inflation factor. The design effect describes the extent to which the sample size must be increased in order to obtain the same power. It is given by

$$D_{\text{eff}} = 1 + (m - 1) \times \text{ICC}$$

where m is the mean cluster size. It can be seen that if ICC is zero, the design effect is necessarily zero. We also have potential problems when implementing the design due to issues such as inability to conceal allocation. The CACE method highlighted above demonstrates one way to take account of the problem of contamination; however, a disadvantage of the CACE approach is that it often produces wider CIs than an ITT analysis and can do so for the per protocol analysis. Hence we wanted to investigate the trade-off between performing an individual randomised trial and accepting the fact that there will be some contamination, and the subsequent effect on the precision of the estimate. Monte Carlo simulations were used to investigate the use of Nagelkerke and colleagues approach[252] as an alternative to using cluster allocation when contamination, through non-adherence, is suspected.

Monte Carlo simulation can reveal two different effects on parameter estimates from different treatments. First would be the increased variability in parameter estimates when using one method as opposed to another. The first effect will be the variability of the estimates – in this the expected value of the estimated parameter may be equal to the true parameter value in all methods of analysis, but the variability may be greater in one method. The second effect is bias in the

TABLE 30 *Effect of contamination on the power of a study*

Contamination (%)	Power with Nagelkerke method	
	Sample size = 630, true difference = 0.2	Sample size = 158, true difference = 0.4
0	0.80	0.80
10	0.74	0.72
20	0.64	0.61
30	0.54	0.54

parameter estimates – the expected value of the parameter estimate is different (higher or lower) than the true (population) value. It would be expected that the treatment received analysis would frequently give biased parameter estimates. The other effect to examine is on the *p*-values of the results. Here, the interest is in examining the Type I error rate and power (Type II error rate). When the null result is the case, we would expect to find a Type I error rate equal to 5%, that is, we make a false positive 5% of the time (given that we are using $\alpha = 0.05$). If the resultant error rate is lower than 5%, then the test employed is over-conservative (and has less power); if the error rate is higher than this, the test is not sufficiently stringent. We would expect that the treatment received approach would have an inflated Type I error rate, and that ITT analysis would have a reduced Type I error rate. We would also expect that the IV approach would give a type I error rate which was approximately correct. Finally, examination of the *p*-values can give us information about the power of the test. Those tests which are conservative with their Type I error rate tend also to be less powerful than those tests which are not conservative, hence we expect that the treatment received approach will be more powerful than ITT analysis. A major outcome of this investigation will be to investigate the power of the IV approach under a wide range of circumstances.

Methods

The individual randomised trial was implemented as follows:

- We described a set of hypothetical populations in terms of the parameter values. Hence we defined an effect size for the treatment, a sample size and a level of contamination.
- We generated 1000 samples that had the specified parameter values.
- The samples were then analysed and the results were compared with the known parameter values from the populations. Data were analysed using three different methods: treatment received, ITT and Nagelkerke and colleagues' CACE method.[252]

The cluster randomised trial was implemented as follows:

- We described a set of hypothetical populations in terms of the parameter values. Hence we defined an effect size for the treatment, a sample size and ICC.
- We generated 1000 samples that had the specified parameter values.
- The samples were then analysed and the results were compared with the known parameter values from the populations.

For a CACE approach to deliver an unbiased estimate compared with per protocol analysis, a covariate must correlate with adherence and outcome, otherwise both analyses will give the same result. In this instance, we used the baseline test score as the appropriate covariate that predicts both outcome and contamination. The correlation coefficient that was used was 0.45 for the outcome prediction. This value was taken from the observed value in the IBS study.

Results

In *Table 30* we show the effect of contamination on the power of a study to detect a given effect size with increasing rates of contamination. As can be seen, with increasing contamination the power of the study to detect a true difference between the groups declines. The issue that confronts the trialist is whether an alternative approach, namely cluster randomisation, can be used that avoids contamination **and** preserves study power.

In *Table 31* we illustrate the trade-off that contamination has on the sample size needed to detect a fixed effect size. As can be seen, as contamination increases the sample size required for 80% power increases. *Table 31* also shows the effect on sample size of switching to a cluster design. Using a CACE approach, unless the expected contamination exceeds 30%, retains a

TABLE 31 *Sample sizes needed with cluster randomised trial and no contamination versus individual randomised trial using CACE to adjust for contamination*

True difference	Cluster size	Total sample size needed to have 80% power to detect the true difference between groups for a cluster randomised trial with ICC = 0.04	Contamination (%)	Total sample size needed to have 80% power to detect the true difference between groups for an individual randomised trial using CACE and including a covariate	Contamination effect
0.2	10	1080	0	630	1
	30	1740	10	756	1.20
	50	2400	20	890	1.41
	100	4000	30	1090	1.73
0.4	10	280	0	158	1
	30	480	10	190	1.20
	50	600	20	230	1.46
	100	1000	30	276	1.75

sample size advantage over the cluster randomised design despite a relatively small cluster (i.e. 10) and reasonably small ICC.

These results are very similar to those found by both Torgerson[249] and Slymen and Hovell,[275] who advocated using ITT analysis and accepting some contamination to show a diluted effect size.

Discussion

Contamination, through non-adherence, is a common problem in RCTs and creates problems at the analysis stage. Although it is widely recommended that the primary analysis of an RCT should be using ITT, investigators often supplement this with a per protocol analysis. The main problem with a per protocol approach is that as participants self-select themselves into the two groups, the initial randomisation is undermined and this consequently violates the basis for statistical inference. In addition, the participants who adhere to the treatment usually have different characteristics from the participants who do not, and this may introduce bias.

We have demonstrated the use of the CACE approach to calculate the treatment effects in two recently published RCTs with some degree of non-adherence in the intervention group. A disadvantage of the CACE approach is that it produces wider CIs than an ITT analysis and can do so for the per protocol analysis, as demonstrated in the IBS example. However, in the examples presented we did not include really strong predictors of contamination. If there are strong predictors of treatment contamination, these can be included in the analysis and can in some instances substantially reduce the width of the CIs. Nevertheless, the price of increasing the risk of a Type II error compared with a per protocol approach is, we feel, worth paying as it is better to be approximately correct than precisely wrong, which can be the outcome of a per protocol analysis.

An alternative approach to using individual randomisation and then using a CACE analytical approach is to use cluster allocation. In our simulation study we have shown that despite CACE losing power relative to per protocol analysis, it still has a power advantage over the cluster design even when clusters are fairly small and contamination relatively high. We recommend, therefore, that if contamination can be easily and accurately measured and is less than 30%, trialists should consider using individually randomised designs as opposed to cluster methods. In some instances this is not possible. For example, Craven and colleagues undertook a study looking at boosting children's self-worth through praise.[276] In this study the authors used a split plot design whereby they randomised at the level of the classroom and then within the intervention classes they randomised at the level of the pupil. They found that although the teacher was instructed not to give the intervention to the control children in the class, they were often unable to do this. This resulted in substantial dilution within the intervention classes and therefore the only reasonable solution would be to cluster randomise. In this instance and in similar types of studies (e.g. education of physicians in a new treatment technique or guideline), CACE

analysis would not offer any advantages as it would not be possible to measure contamination accurately at an individual level in order to control for it in the analysis.

We suggest that although the ITT approach should usually remain the primary analysis, some form of CACE analysis should be performed as a secondary analysis instead of a per protocol analysis (or similar alternatives). There may also be instances where CACE could be the primary analysis in the presence of high levels of contamination, when the main research question concerns the individual patient.

Chapter 6
Conclusions

Contamination is often assumed to be a problem in controlled trials of educational interventions. It is commonly given as a reason for using cluster randomisation, or as a reason why effectiveness estimates were small or not statistically significant. However, we have found few studies in which contamination was recorded and quantified, and no original trials in which it was adjusted for statistically. Nevertheless, substantial experience of educational trials in health exists, and there is consensus on when contamination is more or less likely to be a problem. This study provides guidance on situations where contamination is more or less likely, how to avoid it when designing and conducting trials and how to account for it when analysing results.

When designing a controlled trial of an educational intervention in which contamination is possible, one should systematically consider the characteristics of the intervention, of the study population and sample and of the geographical, social and organisational contexts of the study. First, can one avoid it during the trial by restricting educational materials or by asking intervention subjects to keep it to themselves? How easily can the intervention be transmitted to control subjects, as a whole or in parts? Is the intervention unusual or desirable, so that control subjects will hear about it and seek it? Will trial participants actively disseminate it? If one were to randomise clusters rather than individuals, how might contamination still occur? Could the intervention still reach control clusters and, if so, how rapidly? Would it reach all subjects in control clusters, or only a few, and how rapidly? Will control subjects with unfavourable characteristics at baseline be more, less or equally likely to be contaminated? Our simulations in Chapter 4 suggest that, with different combinations of these factors, the relative advantages of cluster and individual randomisation vary. In many situations cluster randomisation will reduce bias, but it may increase bias, especially if entire clusters are rapidly contaminated.

The choice between cluster and individual randomisation should not only be based on concerns about contamination. As stated by several respondents to our Delphi study, the choice of cluster randomisation is often primarily a logical consequence of the level or unit of inference of the trial, rather than a method of avoiding contamination. For example, if an intervention to improve clinical practice is targeted at doctors, but effects are inferred by comparing patients' outcomes, then it makes sense to randomise groups of patients managed by the same doctors, who receive or do not receive an intervention. In that way, doctors (and their patients) in intervention and control groups are comparable at baseline and any difference at follow-up can be attributed to the effect of the intervention on the doctors. If, instead, individual patients are randomised to doctors who happened to have received the intervention or not, then one cannot assume that the doctors are comparable. Hence cluster randomisation may be appropriate even if contamination is not a problem.

If substantial contamination occurs and one knows which control subjects have been contaminated, then CACE or IV analysis seems likely to be more valid than ITT or per protocol analysis, and is more powerful than the latter. However, ITT analysis, based on the groups to which subjects have been initially randomised, should arguably remain the primary method of analysis, with CACE analysis the secondary method. This is because of the assurance that randomisation provides that the groups being compared were identical, on average, regarding all characteristics, including those that were not measured. Statistical adjustment is unlikely to provide the same assurance. However, if contamination is found to have been common and intense, then CACE will probably give less biased effect estimates. If a high degree of contamination is expected before a trial, so that CACE is likely to be most appropriate, protocols should specify this in advance to be the primary analytic method, so as to avoid scepticism about having performed multiple analyses to obtain a desired result.

In trials in which contamination is likely, one should wherever possible record, for each subject, whether and how it occurred. The most obvious way is to add such questions to

follow-up questionnaires. If outcome is assessed without personal contact, there may be creative alternatives. For example, in a screening trial reported in Chapter 2, control patients' records were examined to see whether they had received the screening test. In a cluster randomised nurse education trial we (MOB and others) were concerned that trained nurses would be transferred from intervention to control clinics, so we checked the employment records of control clinics for the trial period. None had.

In summary, contamination is often a potential problem but it may not be. It may be avoided during the design or carrying out of the trial. If it cannot be avoided it can be adjusted for, provided that it is found. Therefore, contamination should be recorded and reported in randomised trials of educational interventions.

Acknowledgements

We are grateful to the National Health Service Research and Development National Coordinating Centre for Research Methodology for funding this research, to Miranda Mugford who helped with study design, to the 37 respondents to the Delphi study for their expert views, Galen Ives and Richard Lindle of Priority Research Ltd for administration of the Delphi questionnaires and the Vocational Dental Practitioners Trials Group for permission to report their results. The Health Services Research Unit is funded by the Scottish Executive Health Department of the Chief Scientist's Office.

Contribution of authors

All authors contributed substantially to planning the study and to writing the report. MR Keogh-Brown (Research Associate), L Shepstone (Reader in Medical Statistics) and MO Bachmann (Professor of Health Services Research) simulated contamination in cluster and individually randomised trials and wrote that chapter. MR Keogh-Brown, MO Bachmann, S Miles (Research Associate) and A Howe (Professor of Primary Care) conducted the Delphi study and wrote that chapter. MR Keogh-Brown, F Song (Reader in Evidence Synthesis), CR Ramsay (Senior Statistician) and MO Bachmann analysed evidence from published reviews and wrote that chapter. C Hewitt (Research Fellow), JNV Miles (Behavioural Scientist) and DJ Torgerson (Director of York Trials Unit) conducted the CACE analyses and wrote that chapter. MR Keogh-Brown and MO Bachmann wrote the Introduction and Conclusions. D Elborne (Professor of Healthcare Evaluation), I Harvey (Professor of Epidemiology and Public Health), MJ Campbell (Professor, Medical Statistics Group, ScHARR) contributed to the study design and interpretation of results. All contributors read and approved the final manuscript. MR Keogh-Brown was the principal investigator and MO Bachmann was the grant-holder.

References

1. Donner A, Klar N. *Design and analysis of cluster randomization trials in health research*. Oxford: Oxford University Press; 2000.
2. Slymen DJ. Hovell MF. Cluster versus individual randomization in adolescent tobacco and alcohol studies: illustrations for design decisions. *Int J Epidemiol* 1997;**26**:765–71.
3. Torgerson DJ. Contamination in trials: is cluster randomisation the answer? *BMJ* 2001;**322**:355–7.
4. Zelen M. A new design for randomized controlled trials. *N Engl J Med* 1979;**300**:1242–5.
5. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;**7**(27).
6. Labarre R, Jequier JC, Shephard RJ, Lavallee H, Rajic M. Testing for inter-group contamination in a controlled longitudinal study of added physical education. *J Sports Med Phys Fitness* 1994;**34**:403–6.
7. Courneya KS, Friedenreich CM, Quinney HA, Fields Al, Jones LW, Fairey AS, A randomised trial of exercise and quality of life in colorectal cancer survivors. *Eur J Cancer Care* 2003;**12**:347–57.
8. Courneya KS, Freidenreich CM, Quinney HA, Fields, ALA, Jones LW, Fairey AS. Predictors of adherence and contamination in a randomized trial of exercise in colorectal cancer survivors. *Psycho-Oncology* 2004;**13**:857–66.
9. Courneya KS, Friedenreich CM, Sela RA, Quinney HA, Rhodes RE, Handman M. The group psychotherapy and home-based physical exercise (group–hope) trial in cancer survivors: physical fitness and quality of life outcomes. *Psycho-Oncology* 2003;**12**:357–74.
10. Ross MW, Chatterjee NS, Leonard L, A community level syphilis prevention programme: outcome data from a controlled trial. *Sex Transm. Infect.* 2004;**80**:100–4.
11. Goel V, Cohen MM, Kaufert P, MacWilliam L. Assessing the extent of contamination in the Canadian National Breast Screening Study. *Am J Prev Med* 1998;**15**:206–11.
12. Tilgren P, Dignan M, Michielutte R. Assessment of contamination in a trial of community-based cancer education. *Am J Health Behav* 1998; **24**:292–7.
13. Stewart-Brown S, Patterson J, Mockford C, Barlow J, Klimes I, Pyper C. Impact of a general practice based group parenting programme: quantitative and qualitative results from a controlled trial at 12 months. *Arch Disease Child* 2004;**89**:519–25.
14. Grimshaw JM, Thomas RE, Maclennan G, Fraser C, Ramsey C, Vale L, *et al*. Guideline dissemination and implementation strategies: a systematic review of effectiveness and efficiency. *J Epidemiol Community Health* 2003;**57**(Suppl I):A1–21.
15. Anderson FA Jr, Wheeler HB, Goldberg RI, Hosmer DW, Forcier A, Patwardhan NA. Changing clinical practice. Prospective study of the impact of continuing medical education and quality assurance programs on use of prophylaxis for venous thromboembolism. *Arch Intern Med* 1994; **154**:669–77.
16. University of Newcastle upon Tyne. *North of England study of standards and performance in general practice*. Report No. 40. Newcastlle upon Tyne: Health Care Research Unit, University of Newcastle upon Tyne; 1990.
17. University of Newcastle upon Tyne. North of England Study of Standards and Performance in General Practice. An overview of the study. Report No. 50. Newcastle upon Tyne: Centre for Health Services Research, Ambulatory Care Programme, University of Newcastle upon Tyne; 1991.
18. University of Newcastle upon Tyne. Medical audit in general practice. I: effects on doctors' clinical behaviour for common childhood conditions. North of England Study of Standards and Performance in General Practice. *BMJ* 1992; **304**:1480–4.
19. University of Newcastle upon Tyne. Medical audit in general practice. II: effects on health of patients with common childhood conditions. North of England Study of Standards and Performance in General Practice. *BMJ* 1992;**304**:1484–8.
20. Diabetes Integrated Care Evaluation Team. Integrated care for diabetes: clinical, psychosocial, and economic evaluation. Diabetes Integrated Care Evaluation Team. *BMJ* 1994;**308**:1208–12.
21. Agency for Health Care Policy and Research. *CCQE–AHCPR guideline criteria project. Building and applying a guideline-based performance measurement system: develop, apply, and evaluate medical review criteria and educational outreach based upon practice guidelines*. Final Project Report. Report No. 97-N002-97-N003, U-97. Rockville, MD: Department of Health and Human Services (US), Public Health Service, Agency for Health Care Policy and Research; 1996.

22. Aubin M, Vezina L, Maziade J, Robitaille NM. Control of arterial hypertension: effectiveness of an intervention performed by family practitioners. *Can Fam Physician* 1994;**40**:1742–52.

23. Aubin M, Vezina L, Fortin JP, Bernard PM. Effectiveness of a program to improve hypertension screening in primary care. *CMAJ* 1994;**150**:509–15.

24. Aucott JN, Pelecanos E, Dombrowski R, Fuehrer SM, Laich J, Aron DC. Implementation of local guidelines for cost-effective management of hypertension. A trial of the firm system. *J Gen Intern Med* 1996;**11**:139–46.

25. Auleley GR, Ravaud P, Giraudeau B, Kerboull L, Nizard R, Massin P, *et al*. Implementation of the Ottawa ankle rules in France. A multicenter randomised controlled trial. *JAMA* 1997; **277**:1935–9.

26. Avorn J, Soumerai SB, Everitt DE, Ross-Degnan D, Beers MH, Sherman D, *et al*. A randomized trial of a program to reduce the use of psychoactive drugs in nursing homes. *N Engl J Med* 1992;**327**:168–73.

27. Barnett GO, Winickoff RN, Morgan MM, Zielstorff RD. A computer-based monitoring system for follow-up of elevated blood pressure. *Med Care* 1983;**21**:400–9.

28. Bearcroft PW, Small JH, Flower CD. Chest radiography guidelines for general practitioners: a practical approach. *Clin Radiol* 1994;**49**:56–8.

29. Becker DM, Gomez EB, Kaiser DL, Yoshihasi A, Hodge RH Jr. Improving preventive care at a medical clinic: how can the patient help? *Am J Prev Med* 1989;**5**:353–9.

30. Bejes C, Marvel MK. Attempting the improbable: offering colorectal cancer screening to all appropriate patients. *Fam Pract Res J* 1992; **12**:83–90.

31. Belcher DW. Implementing preventive services. Success and failure in an outpatient trial. *Arch Intern Med* 1990;**150**:2533–41.

32. Boekeloo BO, Becker DM, Levine DM, Belitsos PC, Pearson TA. Strategies for increasing house staff management of cholesterol with inpatients. *Am J Prev Med* 1990;**6**:51–9.

33. Boissel JP, Collet JP, Alborini A, Cordel JC, Fllsnoel J, Gillet J, *et al*. Education program for general practitioners on breast and cervical cancer screening: a randomized trial. PRE.SA.GF Collaborative Group. *Rev Epidemiol Santé Publique* 1995;**43**:541–7.

34. Browner WS, Baron RB, Solkowitz S, Adler LJ, Gullion DS. Physician management of hypercholesterolemia. A randomized trial of continuing medical education. *West J Med* 1994; **161**:572–8.

35. Bryce FP, Neville RG, Crombie IK, Clark RA, McKenzie P. Controlled trial of an audit facilitator in diagnosis and treatment of childhood asthma in general practice. *BMJ* 1995;**310**:838–42.

36. Neville RG, Clark RA. *The Childhood Asthma Project 1990–93. A report on a controlled trial of an audit facilitator on the diagnosis and treatment of childhood asthma in general practice*. Dundee: Department of General Practice, University of Dundee; 1995.

37. Buchsbaum DG, Buchanan RG, Lawton MJ, Elswick RK Jr, Schnoll SH. A program of screening and prompting improves short-term physician counseling of dependent and nondependent harmful drinkers. *Arch Intern Med* 1993;**153**:1573–7.

38. Buffington J, Bell KM, LaForce FM. A target-based model for increasing influenza immunizations in private practice. Genesee Hospital Medical Staff. *J Gen Intern Med* 1991;**6**:204–9.

39. Burack RC, Gimotty PA, George J, Stengle W, Warbasse L, Moncrease A. Promoting screening mammography in inner-city settings: a randomized controlled trial of computerized reminders as a component of a program to facilitate mammography. *Med Care* 1994;**32**:609–24.

40. Burack RC, Gimotty PA, George J, Simafl MS, Dews P, Moncrease A. The effect of patient and physician reminders on use of screening mammography in a health maintenance organization. Results of a randomized controlled trial. *Cancer* 1996;**78**:1708–21.

41. Callahan CM, Hendrie HC, Dittus RS, Brater DC, Hui SL, Tierney WM. Improving treatment of late life depression in primary care: a randomized clinical trial. *J Am Geriatr Soc* 1994;**42**:839–46.

42. Chambers CV: Balaban DJ, Carlson BL, Ungemack JA, Grasberger DM. Microcomputer-generated reminders. Improving the compliance of primary care physicians with mammography screening guidelines. *J Fam Pract* 1989;**29**:273–80.

43. Chambers CV: Balaban DJ, Carlson BL, Grasberger DM. The effect of microcomputer-generated reminders on influenza vaccination rates in a university-based family practice center. *J Am Board Fam Pract* 1991;**4**:19–26.

44. Chassin MR, McCue SM. A randomized trial of medical quality assurance. Improving physicians' use of pelvimetry. *JAMA* 1986;**256**:1012–16.

45. Cheney C, Ramsdell JW. Effect of medical records' checklists on implementation of periodic health measures, *Am J Med* 1987;**83**:129–36.

46. Cohen Dl, Littenberg B, Wetzel C, Neuhauser D. Improving physician compliance with preventive medicine guidelines. *Med Care* 1982;**20**:1040–5.

47. Cohen SJ, Christen AG, Katz BP, Drook CA, Davis BJ, Smith DM, *et al*. Counseling medical and dental patients about cigarette smoking: the

impact of nicotine gum and chart reminders. *Am J Public Health* 1987;**77**:313–16.

48. Cohen SJ, Stookey GK, Katz BP, Drook CA, Smith DM. Encouraging primary care physicians to help smokers quit. A randomized, controlled trial. *Ann Intern Med* 1989;**110**:648–52.
49. Cohen SJ. Implementing smoking cessation protocols in medical and dental practices. *Tob Control* 1997;**6** (Suppl 1):S24–6.
50. Cowan JA, Heckerling PS, Parker JB. Effect of a fact sheet reminder on performance of the periodic health examination: a randomized controlled trial. *Am J Prev Med* 1992;**8**:104–9.
51. Danchaivijitr S, Chokloikaew S, Tangtrakool T, Waitayapiches S. Does indication sheet reduce unnecessary urethral catheterization? *J Med Assoc Thai* 1992;**75**:1–5.
52. de Burgh S, Mant A, Mattick RP, Donnelly N, Hall W, Bridges-Webb C. A controlled trial of educational visiting to improve benzodiazepine prescribing in general practice. *Aust J Public Health* 1995;**19**:142–8.
53. De Santis G, Harvey KJ, Howard D, Mashford ML, Moulds RF. Improving the quality of antibiotic prescription patterns in general practice. The role of educational intervention. *Med J Aust* 1994; **160**:502–5.
54. Deeb LC, Pettijohn FP, Shirah JK, Freeman G. Interventions among primary-care practitioners to improve care for preventable complications of diabetes. *Diabetes Care* 1988;**11**:275–80.
55. Del Mar CB, Green AC. Aid to diagnosis of melanoma in primary medical care. *BMJ* 1995; **310**:492–5.
56. Dickey LL, Petitti D. A patient-held minirecord to promote adult preventive care. *J Fam Pract* 1992; **34**:457–63.
57. Carney PA, Dietrich AJ, Keller A, Landgraf J, O'Connor GT. Tools, teamwork, and tenacity: an office system for cancer prevention. *J Fam Pract* 1992;**35**:388–94.
58. Carney PA, Dietrich AJ, Freeman DH Jr, Mott LA. A standardized-patient assessment of a continuing medical education program to improve physicians' cancer-control clinical skills. *Acad Med* 1995; **70**:52–8.
59. Dietrich AJ, Barrett J, Levy D, Carney-Gersten P. Impact of an educational program on physician cancer control knowledge and activities. *Am J Prev Med* 1990;**6**:346–52.
60. Dietrich AJ, O'Connor GT, Keller A, Carney PA, Levy D, Whaley FS. Cancer: improving early detection and prevention. A community practice randomised trial. *BMJ* 1992;**304**:687–91.
61. Dietrich AJ, Sox CH, Tosteson TD, Woodruff CB. Durability of improved physician early detection of cancer after conclusion of intervention support. *Cancer Epidemiol Biomarkers Prev* 1994;**3**:335–40.
62. Diwan VK, Wahlstrom R, Tomson G, Beermann B, Sterky G, Eriksson B. Effects of 'group detailing' on the prescribing of lipid-lowering drugs: a randomized controlled trial in Swedish primary care. *J Clin Epidemiol* 1995;**48**:705–11.
63. Wahlstrom R. Bridging the gap between guidelines and clinical practice. An educational intervention in primary care. Dissertation. Stockholm: Karolinska Institutet; 1997.
64. Dranitsaris G, Warr D, Puodziunas A. A randomized trial of the effects of pharmacist intervention on the cost of antiemetic therapy with ondansetron. *Support Care Cancer* 1995;**3**:183–9.
65. Elliott TE, Murray DM, Oken MM, Johnson KM, Braun BL, Elliott BA, *et al*. Improving cancer pain management in communities: main results from a randomized controlled trial. *J Pain Sympt Manage* 1997;**13**:191–203.
66. Emslie C, Grimshaw J, Templeton A. Do clinical guidelines improve general practice management and referral of infertile couples? *BMJ* 1993; **306**:1728–31.
67. Evans AT, Rogers LQ, Peden JG Jr, Seelig CB, Layne RD, Levine MA, *et al*. Teaching dietary counseling skills to residents: patient and physician outcomes. The CADRE Study Group. *Am J Prev Med* 1996;**12**:259–65.
68. Evans D, Mellins R, Lobach K, Ramos-Bonoan C, Pinkett-Heller M, Wiesemann S, *et al*. Improving care for minority children with asthma: professional education in public health clinics. *Pediatrics* 1997;**99**:157–64.
69. Feder G, Griffiths C, Highton C, Eldridge S, Spence M, Southgate L. Do clinical guidelines introduced with practice based education improve care of asthmatic and diabetic patients? A randomised controlled trial in general practices in east London. *BMJ* 1995;**311**:1473–8.
70. Feder G, Griffiths C. *The dissemination of asthma and diabetes guidelines to inner-city general practice. An evaluation in east London non-training general practices. Hackney Collaborative Clinical Guidelines Project*. London: Department of General Practice and Primary Care, Medical College of St Bartholomew's and the Royal London Hospitals; 1995.
71. Fender GR, Prentice A, Gorst T, Nixon RM, Duffy SW, Day NE, *et al*. Randomised controlled trial of educational package on management of menorrhagia in primary care: the Anglia menorrhagia education study. *BMJ* 1999; **318**:1246–50.
72. Fletcher SW, Harris RP, Gonzalez JJ, Degnan D, Lannin DR, Strecher VJ, *et al*. Increasing

mammography utilization: a controlled study. *J Nat Cancer Inst* 1993;**85**:112–20.

73. Flynn BS, Gavin P, Worden JK, Ashikaga T, Gautam S, Carpenter J. Community education programs to promote mammography participation in rural New York State. *Prev Med* 1997;**26**:102–8.
74. Fowkes FG, Davies ER, Evans KT, Green G, Hartley G, Hugh AE, *et al*. Multicentre trial of four strategies to reduce use of a radiological test. *Lancet* 1986;**15**:367–70.
75. Fox S, Tsou CV, Klos DS. An intervention to increase mammography screening by residents in family practice. *J Fam Pract* 1985;**20**:467–71.
76. Frame PS, Zimmer JG, Werth PL, Hall WJ, Eberly SW. Computer-based vs manual health maintenance tracking. A controlled trial. *Arch Fam Med* 1994;**3**:581–8.
77. Freeborn DK, Shye D, Mullooly JP, Eraker S, Romeo J. Primary care physicians' use of lumbar spine imaging tests: effects of guidelines and practice pattern feedback. *J Gen Intern Med* 1997; **12**:619–25.
78. Gans KM, Lapane KL, Lasater TM, Carleton RA. Effects of intervention on compliance to referral and lifestyle recommendations given at cholesterol screening programs. *Am J Prev Med* 1994; **10**:275–82.
79. Girotti MJ, Fodoruk S, Irvine-Meek J, Rotstein OD. Antibiotic handbook and pre-printed perioperative order forms for surgical antibiotic prophylaxis: do they work? *Can J Surg* 1990; **33**:385–8.
80. Goldberg HI, Wagner EH, Fihn SD, Martin DP, Horowitz CR, Christensen DB, *et al*. A randomized controlled trial of CQI teams and academic detailing: can they alter compliance with guidelines? *Jt Comm J Qual Improve* 1998; **24**:130–42.
81. Gonzalez JJ, Ranney J, West J. Nurse-initiated health promotion prompting system in an internal medicine residents' clinic. *South Med J* 1989; **82**:342–4.
82. Grady KE, Lemkau JP, Lee NR, Caddell C. Enhancing mammography referral in primary care. *Prev Med* 1997;**26**:791–800.
83. Hartmann P, Bott U, Grosser M, Kronsbein P, Jorgens V. Effects of peer-review groups on physicians' practice. *Eur J Gen Pract* 1995;**1**:107–12.
84. Hay JA, Maldonado L, Weingarten SR, Ellrodt AG. Prospective evaluation of a clinical guideline recommending hospital length of stay in upper gastrointestinal tract hemorrhage. *JAMA* 1997; **278**:2151–6.
85. Headrick LA, Speroff T, Pelecanos HI, Cebul RD. Efforts to improve compliance with the National Cholesterol Education Program guidelines. Results of a randomized controlled trial. *Arch Intern Med* 1992;**152**:2490–6.
86. Herfindal ET, Bernstein LR, Kishi DT. Effect of clinical pharmacy services on prescribing on an orthopedic unit. *Am J Hosp Pharm* 1983; **40**:1945–51.
87. Herman CJ, Speroff T, Cebul RD. Improving compliance with immunization in the older adult: results of a randomized cohort study. *J Am Geriatr Soc* 1994;**42**:1154–9.
88. Herman CJ, Speroff T, Cebul RD. Improving compliance with breast cancer screening in older women. Results of a randomized controlled trial. *Arch Intern Med* 1995;**155**:717–22.
89. Hillman AL, Ripley K, Goldfarb N, Nuamah I, Weiner J, Lusk E. Physician financial incentives and feedback: failure to increase cancer screening in Medicaid managed care. *Am J Public Health* 1998;**88**:1699–701.
90. Hopkins JA, Shoemaker WC, Greenfield S, Chang PC, McAuliffe T, Sproat RW. Treatment of surgical emergencies with and without an algorithm. *Arch Surg* 1980;**115**:745–50.
91. Hopkins JA, Shoemaker WC, Chang PC, Schluchter M, Greenfield S. Clinical trial of an emergency resuscitation algorithm. *Crit Care Med* 1983;**11**:621–9.
92. Hueston WJ, Stiles MA. Effects of physician prompters on long-term screening test behaviors. *Fam Pract Res J* 1994;**14**:251–9.
93. Hulscher ME, van Drenth BB, van der Wouden JC, Mokkink HG, van Weel C, Grol RP. Changing preventive practice: a controlled trial on the effects of outreach visits to organise prevention of cardiovascular disease. *Qual Health Care* 1997; **6**:19–24.
94. Hulscher MEJL. Implementing prevention in a general practice: a study on cardiovascular disease. *Dissertation*. The Hague: CIP-Gegevens; 1998.
95. van Drenth BB, Hulscher ME, Mokkink HG, van de Lisdonk EH, van der Wouden JC, Grol RPTM. Effects of outreach visits by trained nurses on cardiovascular risk-factors recording in general practice: a controlled trial. *Eur J Gen Pract* 1997;**3**:90–5.
96. van Drenth BB. Organising cardiovascular preventive care in general practice. Dissertation. Wageningen: Ponsen and Looijen; 1998.
97. Jones DL, Kroenke K, Landry FJ, Tomich DJ, Ferrel RJ. Cost savings using a stepped-care prescribing protocol for nonsteroidal anti-inflammatory drugs. *JAMA* 1996;**275**:926–30.
98. Calkins E, Katz LA, Karuza J, Wagner A. The small group consensus process for changing physician practices: influenza vaccination. *HMO Pract* 1995;**9**:107–10.

99. Karuza J, Calkins E, Feather J, Hershey CO, Katz L, Majeroni B. Enhancing physician adoption of practice guidelines. Dissemination of influenza vaccination guideline using a small-group consensus process. *Arch Intern Med* 1995; **155**:625–32.

100. Katon W, Yon Korff M, Lin E, Bush T, Ormel J. Adequacy and duration of antidepressant treatment in primary care. *Med Care* 1992; **30**:67–76.

101. Katon W, Yon Korff M, Lin E, Walker E, Simon GE, Bush T, *et al*. Collaborative management to achieve treatment guidelines. Impact on depression in primary care. *JAMA* 1995; **273**:1026–31.

102. Katon W, Yon Korff M, Lin E, Simon G, Walker E, Bush T, *et. al*. Collaborative management to achieve depression treatment guidelines. *J Clin Psychiatry* 1997;**58**:20–3.

103. Katon W, Robinson P, Yon Korff M, Lin E, Bush T, Ludman E, *et al*. A multifaceted intervention to improve treatment of depression in primary care. *Arch Gen Psychiatry* 1996;**53**:924–32.

104. Kong GK, Belman MJ, Weingarten S. Reducing length of stay for patients hospitalized with exacerbation of COPD by using a practice guideline. *Chest* 1997;**111**:89–94.

105. Landefeld CS, Anderson PA. Guideline-based consultation to prevent anticoagulant-related bleeding. A randomized, controlled trial in a teaching hospital. *Ann Intern Med* 1992; **116**:829–37.

106. Landis SE, Hulkower SD, Pierson S. Enhancing adherence with mammography through patient letters and physician prompts. A pilot study. *N C Med J* 1992;**53**:575–8.

107. Lee TH, Pearson SD, Johnson PA, Garcia TB, Weisberg MC, Guadagnoli E, *et al*. Failure of information as an intervention to modify clinical management. A time-series trial in patients with acute chest pain. *Ann Intern Med* 1995;**122**:434–7.

108. Leviton LC, Goldenberg RL, Baker CS, Schwartz RM, Freda MC, Fish LJ, *et al*. Methods to encourage the use of antenatal corticosteroid therapy for fetal maturation: a randomized controlled trial. *JAMA* 1999;**281**:46–52.

109. Lin EH, Katon WJ, Simon GE, Yon Korff M, Bush TM, Rutter CM, *et al*. Achieving guidelines for the treatment of depression in primary care: is physician education enough? *Med Care* 1997; **35**:831–42.

110. Linn BS. Continuing medical education. Impact on emergency room burn care. *JAMA* 1980; **244**:565–70.

111. Litzelman DK, Slemenda CW, Langefeld CD, Hays LM, Welch MA, Bild DE, *et al*. Reduction of lower extremity clinical abnormalities in patients with non-insulin-dependent diabetes mellitus. A randomized, controlled trial. *Ann Intern Med* 1993;**119**:36–41.

112. Litzelman DK, Dittus RS, Miller ME, Tierney WM. Requiring physicians to respond to computerized reminders improves their compliance with preventive care protocols. *J Gen Intern Med* 1993; **8**:311–17.

113. Lobach DF, Hammond WE. Development and evaluation of a Computer-Assisted Management Protocol (CAMP): improved compliance with care guidelines for diabetes mellitus. *Proc Annu Symp Comput Appl Med Care* 1994;787–91.

114. Lobach DF, Hammond WE. Computerized decision support based on a clinical practice guideline improves compliance with care standards. *Am J Med* 1997;**102**:89–98.

115. Lobach DF. Electronically distributed, computer-generated, individualized feedback enhances the use of a computerized practice guideline. *Proc AMIA Annual Fall Symp* 1996;493–7.

116. Lomas J, Enkin M, Anderson GM, Hannah WJ, Vayda E, Singer J. Opinion leaders vs audit and feedback to implement practice guidelines. Delivery after previous cesarean section. *JAMA* 1991;**265**:2202–7.

117. Lomas J. Making clinical policy explicit. Legislative policy making and lessons for developing practice guidelines. *Int J Technol Assess Health Care* 1993;**9**:11–25.

118. MacCosbe PE, Gartenberg G. Modifying empiric antibiotic prescribing: experience with one strategy in a medical residency program. *Hosp Formulary* 1985;**20**:986–8.

119. Manfredi C, Czaja R, Freels S, Trubitt M, Warnecke R, Lacey L. Prescribe for health. Improving cancer screening in physician practices serving low-income and minority populations. *Arch Fam Med* 1998;**7**:329–37.

120. Margolis CZ, Warshawsky SS, Goldman L, Dagan O, Wirtschafter D, Pliskin JS. Computerized algorithms and pediatricians' management of common problems in a community clinic. *Acad Med* 1992;**67**:282–4.

121. Marton KI, Tul V, Sox HC Jr. Modifying test-ordering behavior in the outpatient medical clinic. A controlled trial of two educational interventions. *Arch Intern Med* 1985;**145**:816–21.

122. Mayefsky JH, Foye HR. Use of a chart audit: teaching well child care to paediatric house officers. *Med Educ* 1993;**27**:170–4.

123. Mazzuca SA, Vinicor F, Einterz RM, Tierney WM, Norton JA, Kalasinksi LS. Effects of the clinical environment on physicians' response to postgraduate medical education. *Am Educ Res J* 1990;**27**:473–88.

124. McA1ister NH, Covvey HD, Tong C, Lee A, Wigle ED. Randomised controlled trial of computer assisted management of hypertension in primary care. *BMJ* 1986;**293**:670–4.

125. McDonald CJ. Use of a computer to detect and respond to clinical events: its effect on clinician behavior. *Ann Intern Med* 1976;**84**:162–7.

126. McDonald CJ, Wilson GA, McCabe GP Jr. Physician response to computer reminders. *JAMA* 1980;**244**:1579–81.

127. McDonald CJ, Hui SL, Smith DM, Tierney WM, Cohen SJ, Weinberger M, *et al*. Reminders to physicians from an introspective computer medical record. A two-year randomized trial. *Ann Intern Med* 1984;**100**:130–8.

128. McDonald CJ, Hui SL, Tierney WM. Effects of computer reminders for influenza vaccination on morbidity during influenza epidemics. *MD Computing* 1992;**9**:304–12.

129. McPhee SJ, Bird JA, Jenkins CN, Fordham D. Promoting cancer screening. A randomized, controlled trial of three interventions. *Arch Intern Med* 1989;**149**:1866–72.

130. Bird JA, McPhee SJ, Jenkins C, Fordham D. Three strategies to promote cancer screening. How feasible is wide-scale implementation? *Med Care* 1990;**28**:1005–12.

131. Fordham D, McPhee SJ, Bird JA, Rodnick JE, Detmer WM. The Cancer Prevention Reminder System. *MD Computing* 1990;**7**:289–95.

132. McPhee SJ, Bird JA, Fordham D, Rodnick JE, Osborn EH. Promoting cancer prevention activities by primary care physicians. Results of a randomized, controlled trial. *JAMA* 1991; **266**:538–44.

133. Meador KG, Taylor JA, Thapa PB, Fought RL, Ray WA. Predictors of antipsychotic withdrawal or dose reduction in a randomized controlled trial of provider education. *J Am Geriatr Soc* 1997; **45**:207–10.

134. Mesters I, Meertens R, Kok G, Parcel GS. Effectiveness of a multidisciplinary education protocol in children with asthma (0–4 years) in primary health care. *J Asthma* 1994;**31**:347–59.

135. Moore M, Siu A, Partridge JM, Hays RD, Adams J. A randomized trial of office-based screening for common problems in older persons. *Am J Med* 1997;**102**:371–8.

136. Morgan M, Studney DR, Barnett GO, Winickoff RN. Computerized concurrent review of prenatal care. *Qual Rev Bull* 1978;**4**:33–6.

137. Morrison J. *Improving quality of referral: a cost-effectiveness evaluation of clinical guidelines for infertility management across the interface*. Final Report. Project No. 2–09. Glasgow: Department of General Practice, University of Glasgow; 1999.

138. Morrissey JP, Harris RP, Kincade-Norburn J, McLaughlin C, Garrett JM, Jackman AM, *et al*. Medicare reimbursement for preventive care. Changes in performance of services, quality of life, and health care costs. *Med Care* 1995;**33**:315–31.

139. Nattinger AB, Panzer RJ, Janus J. Improving the utilization of screening mammography in primary care practices. *Arch Intern Med* 1989;**149**:2087–92.

140. Nilasena DS, Lincoln MJ. A computer-generated reminder system improves physician compliance with diabetes preventive care guidelines. *Proc Annu Symp Comput Appl Med Care* 1995;640–5.

141. Norton PG, Dempsey LJ. Self-audit: its effect on quality of care. *J Fam Pract* 1985;**21**:289–91.

142. Oakeshott P, Kerry SM, Williams JE. Randomized controlled trial of the effect of the Royal College of Radiologists' guidelines on general practitioners' referrals for radiographic examination. *Br J Gen Pract* 1994;**44**:197–200.

143. Ockene IS, Hebert JR, Ockene JK, Merriam PA, Hurley TG, Saperia GM. Effect of training and a structured office practice on physician-delivered nutrition counseling: the Worcester-Area Trial for Counseling in Hyperlipidemia (WATCH). *Am J Prev Med* 1996;**12**:252–8.

144. Onion CWR. Changes in medical practice following superficial and deep processing of evidence: a controlled experiment in clinical guideline implementation. *Dissertation*. University of Liverpool; 1997.

145. Ornstein SM, Garr DR, Jenkins RG, Rust PF, Arnon A. Computer-generated physician and patient reminders. Tools to improve population adherence to selected preventive services. *J Fam Pract* 1991;**32**:82–90.

146. Overhage JM, Tierney WM, McDonald CJ. Computer reminders to implement preventive care guidelines for hospitalized patients. *Arch Intern Med* 1996;**156**:1551–6.

147. Overhage JM, Tierney WM, Zhou XH, McDonald CJ. A randomized trial of 'corollary orders' to prevent errors of omission. *J Am Med Inform Assoc* 1997;**4**:364–75.

148. Hargraves JL, Palmer RH, Orav EJ, Wright EA. Practice characteristics and performance of primary care practitioners. *Med Care* 1996; **34** (9 Suppl):SS67–76.

149. Palmer RH, Louis TA, Hsu LN, Peterson HF, Rothrock JK, Strain R, *et al*. A randomized controlled trial of quality assurance in sixteen ambulatory care practices. *Med Care* 1985; **23**:751–70.

150. Palmer RH. Does quality assurance improve ambulatory care? Implementing a randomized controlled trial in 16 group practices. *J Ambul Care Manage* 1986;**9**:1–15.

151. Palmer RH, Louis TA, Peterson HF, Rothrock JK, Strain R, Wright EA. What makes quality assurance effective? Results from a randomized, controlled trial in 16 primary care group practices. *Med Care* 1996;**34** (9 Suppl):SS29–39.

152. Perez-Cuevas R, Guiscafre H, Munoz O, Reyes H, Tome P, Libreros V, *et al*. Improving physician prescribing patterns to treat rhinopharyngitis. Intervention strategies in two health systems of Mexico. *Soc Sci Med* 1996;**42**:1185–94.

153. Peterson GM, Bergin JK, Nelson BJ, Stanton LA. Improving drug use in rheumatic disorders. *J Clin Pharm Ther* 1996;**21**:215–20.

154. Pierce C, Goldstein M, Suozzi K, Gallaher M, Dietz, Stevenson J. The impact of the standards for pediatric immunization practices on vaccination coverage levels. *JAMA* 1996; **276**:626–30.

155. Prislin MD, Vandenbark MS, Clarkson QD. The impact of a health screening flow sheet on the performance and documentation of health screening procedures. *Fam Med* 1986;**18**:290–2.

156. Rabin DL, Boekeloo BO, Marx ES, Bowman MA, Russell NK, Willis AG. Improving office-based physician's prevention practices for sexually transmitted diseases. *Ann Intern Med* 1994; **121**:513–19.

157. Raisch DW, Bootman JL, Larson LN, McGhan WF. Improving antiulcer agent prescribing in a health maintenance organization. *Am J Hosp Pharm* 1990; **47**:1766–73.

158. Ray WA, Blazer DG, Schaffner W; Federspiel CF, Fink R. Reducing long-term diazepam prescribing in office practice. A controlled trial of educational visits. *JAMA* 1986;**256**:2536–9.

159. Ray WA, Blazer DG, Schaffner W, Federspiel CF. Reducing antipsychotic drug prescribing for nursing home patients: a controlled trial of the effect of an educational visit. *Am J Public Health* 1987;**7**:1448–50.

160. Ray WA, Taylor JA, Meador KG, Lichtenstein MJ, Griffin MR, Fought R, *et al*. Reducing antipsychotic drug use in nursing homes. A controlled trial of provider education. *Arch Intern Med* 1993;**153**:713–21.

161. Robie PW. Improving and sustaining outpatient cancer screening by medicine residents. *South Med J* 1988;**81**:902–5.

162. Robinson MB, Thompson E, Black NA. Evaluation of the effectiveness of guidelines, audit and feedback: improving the use of intravenous thrombolysis in patients with suspected acute myocardial infarction. *Int J Qual Health Care* 1996; **8**:211–22.

163. Rogers JL, Raring OM. The impact of a computerized medical record summary system on incidence and length of hospitalization. *Med Care* 1979;**17**:618–30.

164. Rogers JL, Raring OM, Wortman PM, Watson RA, Goetz JP. Medical information systems: assessing impact in the areas of hypertension, obesity and renal disease. *Med Care* 1982;**20**:63–74.

165. Rogers JL, Raring OM, Goetz JP. Changes in patient attitudes following the implementation of a medical information system. *Qual Rev Bull* 1984;**10**:65–74.

166. Rokstad K, Straand J, Fugelli P. Can drug treatment be improved by feedback on prescribing profiles combined with therapeutic recommendations? A prospective, controlled trial in general practice. *J Clin Epidemiol* 1995; **48**:1061–8.

167. McDowell I, Newell C, Rosser W. Comparison of three methods of recalling patients for influenza vaccination. *CMAJ* 1986;**135**:991–7.

168. McDowell I, Newell C, Rosser W. Computerized reminders to encourage cervical screening in family practice. *J Fam Pract* 1989;**28**:420–4.

169. McDowell I, Newell C, Rosser W. A randomized trial of computerized reminders for blood pressure screening in primary care. *Med Care* 1989; **27**:297–305.

170. Rosser WW, McDowell I, Newell C. Use of reminders for preventive procedures in family medicine. *CMAJ* 1991;**145**:807–14.

171. Rosser WW, Hutchison BG, McDowell I, Newell C. Use of reminders to increase compliance with tetanus booster vaccination. *CMAJ* 1992; **146**:911–17.

172. Rossi RA, Every NR. A computerized intervention to decrease the use of calcium channel blockers in hypertension. *J Gen Intern Med* 1997;**12**:672–8.

173. Safran C, Rind DM, Davis RB, Ives D, Sands DZ, Currier J, *et al*. Guidelines for management of HIV infection with computer-based patient's record. *Lancet* 1995;**346**:341–6.

174. Safran C, Rind DM, Davis RB, Sands DZ, Caraballo E, Rippel K, *et al*. A clinical trial of a knowledge-based medical record. *Medinfo* 1995; **8**:1076–80.

175. Sanazaro PJ, Worth RM. Concurrent quality assurance in hospital care. Report of a study by Private Initiative in PSRO. *N Engl J Med* 1978; **298**:1171–7.

176. Schectman JM, Elinsky EG, Pawlson LG. Effect of education and feedback on thyroid function testing strategies of primary care clinicians. *Arch Intern Med* 1991;**151**:2163–6.

177. Schmidt I, Claesson CB, Westerholm B, Nilsson LG, Svarstad BL. The impact of regular multidisciplinary team interventions on

psychotropic prescribing in Swedish nursing homes. *J Am Geriatr Soc* 1998;**46**:77–82.

178. Schreiner DT, Petrusa ER, Rettie CS, Kluge RM. Improving compliance with preventive medicine procedures in a house staff training program. *South Med J* 1988;**81**:1553–7.

179. Shojania KG, Yokoe D, Platt R, FiskioJ, Ma'luf N, Bates DW. Reducing vancomycin use utilizing a computer guideline: results of a randomized controlled trial. *J Am Med Inform Assoc* 1998; **5**:554–62.

180. Smeele IJM, Grol RPTM, Van Schayck CP, Van den Bosch WJHM, Van den Hoogen HJM, Muris JWM. Can small group education and peer review improve care for patients with asthma/chronic obstructive pulmonary disease? *Qual Health Care* 1999;**8**:92–8.

181. Smith DH, Christensen DB, Stergachis A, Holmes G. A randomized controlled trial of a drug use review intervention for sedative hypnotic medications. *Med Care* 1998;**36**:1013–21.

182. Somkin CP, Hiatt RA, Hurley LB, Gruskin E, Ackerson L, Larson P. The effect of patient and provider reminders on mammography and Papanicolaou smear screening in a large health maintenance organization. *Arch Intern Med* 1997; **157**:1658–64.

183. Sommers LS, Sholtz R, Shepherd RM, Starkweather DB. Physician involvement in quality assurance. *Med Care* 1984;**22**:1115–38.

184. Soumerai SB, Salem-Schatz S, Avorn J, Casteris CS, Ross-Degnan D, Popovsky MA. A controlled trial of educational outreach to improve blood transfusion practice. *JAMA* 1993;**270**:961–6.

185. Soumerai SB, McLaughlin TJ, Gurwitz J, Guadagnoli E, Hauptman P. A randomized trial of opinion leader education plus performance feedback to improve quality of care for acute MI. *Abstract Book/Association for Health Services Research* 1997;**14**:101–2.

186. Soumerai SB, McLaughlin TJ, Gurwitz JH, Guadagnoli E, Hauptman PJ, Borbas C, *et al*. Effect of local medical opinion leaders on quality of care for acute myocardial infarction: a randomized controlled trial. *JAMA* 1998; **279**:1358–63.

187. Sulmasy DP, Terry PB, Faden RR, Levine DM. Long-term effects of ethics education on the quality of care for patients who have do-not-resuscitate orders. *J Gen Intern Med* 1994;**9**:622–6.

188. Tape TG, Campbell JR. Computerized medical records and preventive health care: success depends on many factors. *Am J Med* 1993; **94**:619–25.

189. Tape TG, Givner N, Wigton RS, Seelig CB, Patil K, Campbell JR. Process in ambulatory care: a controlled clinical trial of computerized records. *Proc Annu Symp Comput Appl Med Care* 1988;749–52.

190. Thomas JC, Moore A, Qualls PE. The effect on cost of medical care for patients treated with an automated clinical audit system. *J Med Syst* 1983; **7**:307–13.

191. Turner BJ, Day SC, Borenstein B. A controlled trial to improve delivery of preventive care: physician or patient reminders? *J Gen Intern Med* 1989;**4**:403–9.

192. Turner RC, Waivers LE, O'Brien K. The effect of patient-carried reminder cards on the performance of health maintenance measures. *Arch Intern Med* 1990;**150**:645–7.

193. Turner RC, Peden JG Jr, O'Brien K. Patient-carried card prompts vs computer-generated prompts to remind private practice physicians to perform health maintenance measures. *Arch Intern Med* 1994;**154**:1957–60.

194. Urban N, Taplin SH, Taylor VM, Peacock S, Anderson G, Conrad D, *et al*. Community organization to promote breast cancer screening among women aged 50–75. *Prev Med* 1995; **24**:477–84.

195. Van der Weijden T. Effect of implementation of cholesterol guidelines on performance. A randomised controlled trial in 20 general practices. *Dissertation*. University of Maastricht; 1997.

196. Van der Weijden T, Grol RPTM, Knottnerus JA. Feasibility of a national cholesterol guideline in daily practice. A randomized controlled trial in 20 general practices. *Int J Qual Health Care* 1999; **11**:131–7.

197. van Essen GA, Kuyvenhoven MM, de Melker RA. Implementing the Dutch College of General Practitioner's guidelines for influenza vaccination: an intervention study. *Br J Gen Pract* 1997;**47**:25–9.

198. Vissers MC, Hasman A, van der Linden CJ. Protocol processing system (ProtoVIEW) to support residents at the emergency ward. *Comput Methods Programs Biomed* 1995;**48**:53–8.

199. Vissers MC, Biert J, van der Linden CJ, Hasman A. Effects of a supportive protocol processing system (ProtoVIEW) on clinical behaviour of residents in the accident and emergency department. *Comput Methods Programs Biomed* 1996;**49**:177–84.

200. Vissers MC, Hasman A, van der Linden CJ. Impact of a protocol processing system (ProtoVIEW) on clinical behaviour of residents and treatment. *Int J Bio-Med Comput* 1996;**42**:143–50.

201. Watson MC. The development, implementation and evaluation of prescribing guidelines in general practice. *Dissertation*. Primary Health Care and Epidemiology, University of Bristol; 1998.

202. Watson MC, Gunnell DJ, Peters TJ, Brookes ST, Sharp DJ. *Randomised controlled trial comparing two implementation strategies for prescribing guidelines in general practice*. Bristol: Department of Clinical and Social Medicine, University of Bristol; 1999.

203. Weingarten MA, Bazel D, Shannon HS. Computerized protocol for preventive medicine: a controlled self-audit in family practice. *Fam Pract* 1989;**6**:120–4.

204. Weingarten SR, Riedinger MS, Conner L, Lee TH, Hoffman I, Johnson B, *et al*. Practice guidelines and reminders to reduce duration of hospital stay for patients with chest pain. An interventional trial. *Ann Intern Med* 1994;**120**:257–63.

205. Weingarten S, Riedinger M, Conner L, Johnson B, Ellrodt AG. A practice guideline to reduce hospital costs for chest pain patients: two prospective controlled interventional trials using different dissemination strategies. *Abstract Book/Association for Health Services Research* 1994;**11**:114–15.

206. Weingarten S, Riedinger M, Conner L, Johnson B, Ellrodt AG. Reducing lengths of stay in the coronary care unit with a practice guideline for patients with congestive heart failure. Insights from a controlled clinical trial. *Med Care* 1994; **32**:1232–43.

207. Weingarten SR, Riedinger MS, Hobson Po Noah MS, Johnson B, Giugliano G, *et al*. Evaluation of a pneumonia practice guideline in an interventional trial. *Am J Respir Crit Care Med* 1996;**153**:1110–15.

208. Wilson DM, Taylor DW, Gilbert JR, Best JA, Lindsay EA, Willms DG, *et al*. A randomized trial of a family physician intervention for smoking cessation. *JAMA* 1988;**260**:1570–4.

209. Winickoff RN, Coltin KL, Morgan MM, Buxbaum RC, Bamett GO. Improving physician performance through peer comparison feedback. *Med Care* 1984;**22**:527–34.

210. Winickoff RN, Wilner S, Neisuler R, Barnett GO. Limitations of provider interventions in hypertension quality assurance. *Am J Public Health* 1985;**75**:43–6.

211. Wirtschafter DD, Sumners J, Jackson JR, Brooks CM, Turner M. Continuing medical education using clinical algorithms. A controlled-trial assessment of effect on neonatal care. *Am J Dis Child* 1986;**140**:791–7.

212. Worrall G, Angel J, Chaulk P, Clarke C, Robbins M. Effectiveness of an educational strategy to improve family physicians' detection and management of depression: a randomized controlled trial. *CMAJ* 1999;**161**:37–40.

213. Costanza ME, Zapka JG, Harris DR, Hosmer D, Barth R, Gaw VP, *et al*. Impact of a physician intervention program to increase breast cancer screening. *Cancer Epidemiol Biomarkers Prev* 1992;**1**:581–9.

214. Zapka JG, Costanza ME, Harris DR, Hosmer D, Stoddard A, Barth R, *et al*. Impact of a breast cancer screening community intervention. *Prev Med* 1993;**22**:34–53.

215. Zenni EA, Robinson TN. Effects of structured encounter forms on pediatric house staff knowledge, parent satisfaction, and quality of care. A randomized, controlled trial. *Arch Pediatr Adolesc Med* 1996;**150**:975–80.

216. L'Abbé KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. *Ann Intern Med* 1987; **107**:224–33.

217. Song, F. Exploring heterogeneity in meta-analysis: is the L'Abbé plot useful? *J Clin Epidemiol* 1999; **52**:725–30.

218. Riemsma RP, Pattenden J, Bridle C, Sowden AJ, Mather L, Watt IS, *et al*. A systematic review of the effectiveness of interventions based on the stages-of-change approach to promote individual behaviour change. *Health Technol Assess* 2002;**6**(24).

219. Wang WD. Feasibility and effectiveness of a stages of-change model in cigarette smoking cessation counseling. *J Formos Med Assoc* 1994;**93**:752–7.

220. DiClemente CC, Prochaska JO, Fairhurst SK, Velicer WF, Velasquez MM, Rossi JS. The process of smoking cessation: an analysis of precontemplation, contemplation, and preparation stages-of-change. *J Consult Clin Psychol* 1991;**59**:295–304.

221. Dijkstra A, De Vries H, Roijackers J. Targeting smokers with low readiness to change with tailored and nontailored self-help materials. *Prev Med* 1999;**28**:203–11.

222. Butler CC, Rollnick S, Cohen D, Bachmann M, Russell I, Stott N. Motivational consulting versus brief advice for smokers in general practice: a randomized trial. *Br J Gen Pract* 1999;**49**:611–16.

223. Gritz ER, Carr CR, Rapkin D, Abemayor E, Chang LJ, Wong WK, *et al*. Predictors of long-term smoking cessation in head and neck cancer patients. *Cancer Epidemiol Biomarkers Prev* 1993; **2**:261–70.

224. Resnicow K, Royce J, Vaughan R, Orlandi MA, Smith M. Analysis of a multicomponent smoking cessation project: what worked and why. *Prev Med* 1997;**26**:373–81.

225. Pallonen UE, Velicer WF, Prochaska JO, Rossi JS, Bellis JM, Tsoh JY, *et al*. Computer-based smoking cessation interventions in adolescents: description, feasibility, and six-month follow-up findings. *Subst Use Misuse* 1998;**33**:935–65.

226. Velicer WF, Prochaska JO, Fava JL, Laforge RG, Rossi JS. Interactive versus noninteractive interventions and dose-response relationships for

stage-matched smoking cessation programs in a managed care setting. *Health Psychol* 1999;**18**:21–8.

227. Pallonen UE, Leskinen L, Prochaska JO, Willey CJ, Kaariainen R, Salonen JT. A 2-year self-help smoking cessation manual intervention among middle-aged Finnish men: an application of the transtheoretical model. *Prev Med* 1994;**23**:507–14.

228. Berman BA, Gritz ER, Braxton-Owens H, Nisenbaum R. Targeting adult smokers through a multi-ethnic public school system. *J Cancer Educ* 1995;**10**:91–101.

229. Lennox AS, Bain N, Taylor RJ, McKie L, Donnan PT. Stages-of-change training for opportunistic smoking intervention by the primary healthcare team. Part I: randomised controlled trial of the effect of training on patient smoking outcomes and health professional behaviour as recalled by patients. *Health Educ J* 1998;**57**:140–9.

230. Sinclair HK, Silcock J, Bond CM, Lennox AS, Winfield AJ. The cost-effectiveness of intensive pharmaceutical intervention in assisting people to stop smoking. *Int J Pharm Pract* 1999;**7**:107–12.

231. Morgan GD, Noll EL, Orleans CT, Rimer BK, Amfoh K, Bonney G. Reaching midlife and older smokers: tailored interventions for routine medical care. *Prev Med* 1996;**25**:346–54.

232. Adler M, Ziglio E, editors. *Gazing into the oracle: the Delphi method and its application to social policy and public health*. London: Jessica Kingsley Publishers; 1996.

233. Fowles J. *Handbook of futures research*. Westport, CT: Greenwood Press; 1978.

234. Fisher R. The arrangement of field experiments. *J Min Ag GB* 1926;**33**:503–13.

235. Medical Research Council. Streptomycin treatment of pulmonary tuberculosis. *BMJ* 1948;**ii**:769–82.

236. Senn S. Added values – controversies concerning randomization and additivity in clinical trials. *Stat Med* 2004;**23**:3729–53.

237. Lee YJ, Ellenberg JH, Hirtz DG, Nelson KB. Analysis of clinical trials by treatment actually received – is it really an option. *Stat Med* 1991; **10**:1595–605.

238. Coronary Drug Project Research Group. Influence of adherence to treatment and response of cholesterol on mortality in the coronary drug project. *N Engl J Med* 1982;**303**:1038–41.

239. Sherry S. The Anturane Reinfarction Trial. *Circulation* 1980;**62**:73–8.

240. Temple R, Pledger GW. The FDAs critique of the Anturane Reinfarction Trial. *N Engl J Med* 1980; **303**:1488–92.

241. Hill AB. *Principles of medical statistics*. 7th ed. London: The Lancet; 1961.

242. Bjerkelund C. The effect of long term treatment with dicoumarol in myocardial infarction; a controlled clinical study. *Acta Med Scand* 1957; **158** (Suppl 330):1–212.

243. Lewis JA, Machin D. Intention to treat – who should use ITT? *Br J Cancer* 1993;**68**:647–50.

244. Dallal G. *Intention-to-treat Analysis*. Medford: Tufts University; 2004.

245. Murray D. *Design and analysis of group randomization trials*. New York: Oxford University Press; 1998.

246. Farrin A, Russell I, Torgerson D, Underwood M. Differential recruitment in a cluster randomized trial in primary care: the experience of the UK Back pain, Exercise, Active management and Manipulation (UK BEAM) feasibility study. *Clin Trials* 2005;**2**:119–124.

247. Puffer S, Torgerson D, Watson J. Evidence for risk of bias in cluster randomised trials: review of recent trials published in three general medical journals. *BMJ* 2003;**327**:785–9.

248. Hahn S, Puffer S, Torgerson D, Watson J. Methodological bias in cluster randomised trials. *BMC Med Res Methodol* 2005;**5**:10.

249. Torgerson DJ. Contamination in trials: is cluster randomisation the answer? *BMJ* 2001;**322**:355–7.

250. Angrist JD, Imbens GW, Rubin DB. Identification of causal effects using instrumental variables. *J Am Stat Assoc* 1996;**91**:444–55.

251. Bloom HS. Accounting for no-shows in experimental evaluation designs. *Eval Rev* 1984; **8**:225–46.

252. Nagelkerke N, Fidler V, Bersen R, Borgdorff M. Estimating treatment effects in randomised clinical trials in the presence of non-compliance. *Stat Med* 2000;**19**:1849–64.

253. Dunn G. The challenge of patient choice and nonadherence to treatment in randomized controlled trials of counseling or psychotherapy. *Understanding Statistics* 2002;**1**:19–29.

254. Altman DG, Schulz KF, Moher D, Egger M, Davidoff F, Elbourne D, *et al*. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001;**134**:663–94.

255. Hardcastle JD, Chamberlain JO, Robinson MH, Moss SM, Amar SS, Balfour TW, *et al*. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. *Lancet* 1996;**348**:1472–7.

256. Sommer A, Zeger SL. On estimating efficacy from clinical trials. *Stat Med* 1991;**10**:45–52.

257. Cuzick J, Edwards R, Segnan N. Adjusting for non-compliance and contamination in randomized clinical trials. *Stat Med* 1997;**16**:1017–29.

258. Efron B, Feldman D. Compliance as an explanatory variable in clinical trials. *J Am Stat Assoc* 1991;**86**:9–17.

259. Newcombe R. Explanatory and pragmatic estimates of the treatment effect when deviations from allocated treatment occur. *Stat Med* 1988; **7**:1179–86.

260. Permutt T, Hebel JR. Simultaneous-equation estimation in a clinical trial of the effect of smoking on birth-weight. *Biometrics* 1989; **45**:619–22.

261. Robins JM, Tsiatis AA. Correcting for noncompliance in randomized trials using rank preserving structural failure time models. *Commun Stat Theory Methods* 1991;**20**:2609–31.

262. Robins J, Tsiatis AA. Semiparametric estimation of an accelerated failure time model with time-dependent covariates. *Biometrika* 1992;**79**:311–19.

263. Tarwotjo I, Sommer A, West KP, Djunaedi E, Mele L, Hawkins B. Influence of participation on mortality in a randomized trial of vitamin-A prophylaxis. *Am J Clin Nutr* 1987;**45**:1466–71.

264. Holland PW. Statistics and causal inference. *J Am Stat Assoc* 1986;**81**:945–60.

265. Rubin D. Estimating causal effects of treatments in randomized and nonrandomized studies. *J Educ Psychol* 1974;**66**:688–701.

266. McClellan M, McNeil BJ, Newhouse JP. Does more intensive treatment of acute myocardial infarction in the elderly reduce mortality? Analysis using instrumental variables. *JAMA* 1994;**272**:859–66.

267. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol* 2000;**29**:722–9.

268. Dunn G, Maracy M, Tomenson B. Estimating treatment effects from randomized clinical trials with noncompliance and loss to follow-up: the role of instrumental variable methods. *Stat Methods Med Res* 2005;**14**:369–95.

269. Imbens GW, Rubin DB. Estimating outcome distributions for compliers in instrumental variables models. *Rev Econ Stud* 1997;**64**:555–74.

270. Imbens GW, Rubin DB. Bayesian inference for causal effects in randomized experiments with noncompliance. *Ann Stat* 1997;**25**:305–27.

271. Little RJ, Yau LHY. Statistical techniques for analyzing data from prevention trials: treatment of no-shows using Rubin's causal model. *Psychol Methods* 1998;**3**:147–59.

272. Jo B, Bengt OM. Modeling of intervention effects with noncompliance: a latent variable modeling approach for randomized trials. In Marcoulides G, Schumacker R, editors. *Advanced structural equation modeling: new developments and techniques*. Mahwah, NJ: Lawrence Erlbaum Associates; 2001. pp. 57–87.

273. Birks YF, Porthouse J, Addie C, Loughney K, Saxon L, Baverstock M, *et al*. Randomized controlled trial of hip protectors among women living in the community. *Osteoporos Int* 2004; **15**:701–6.

274. Atkinson W, Sheldon TA, Shaath N, Whorwell PJ. Food elimination based on IgG antibodies in irritable bowel syndrome: a randomised controlled trial. *Gut* 2004;**53**:1459–64.

275. Slymen D, Hovell M. Cluster versus individual randomization in adolescent tobacco and alcohol studies: illustrations for design decisions. *Int J Epidemiol* 1997;**26**:765–71.

276. Craven RG, Marsh HW, Debus RL, Jayasinghe U. Diffusion effects: control group contamination threats to the validity of teacher-administered interventions. *J Educ Psychol* 2001;**93**:639–45.

Appendix 1

First-round Delphi questionnaire

School of Medicine, Health Policy and Practice
University of East Anglia

DELPHI questionnaire – Contamination in trials of educational intervention

Expressing your views:
The first round of this exercise seeks to determine your views as to the extent to which different types of educational intervention study designs may be at risk of contamination. Some of the questions may seem obvious, but as little previous work addresses this issue we need to see how much consensus exists among experts. We ask you both to rank your agreement, and to add free comment about issues which you consider important but which the format of ranking has not allowed you to address.
Given that many elements in a study may vary, we ask you to consider each answer as if other factors were constant, that is, 'all other things being equal'. So for example, when rating the risks of contamination in trials of interventions aimed at improving skills, knowledge, behaviour or attitudes, you need to abstract this from other features of the intervention such as complexity of intervention or size of target audience. PLEASE CONSIDER THE ISSUE UNDER QUESTION AS IF IT WAS THE ONLY COMPONENT WHICH COULD CAUSE CONTAMINATION.

(Click here for some definitions that you must read.)

Part 1 – Questions:

Q1 Educational interventions can focus on improving knowledge, altering skill set, changing behaviours, or changing attitudes. For each of the following educational interventions please state the likeliness of transfer of the intervention to controls:

	Highly likely	Moderately likely	Neither likely nor unlikely	Moderately unlikely	Highly unlikely	No comment/ not sure
Transfer of new knowledge						
Transfer of new skills						
Altering behaviours						
Changing attitudes						

Q2 Interventions can be defined as Simple, Modest and Complex. Please tell us for each of the following designs the likeliness of contamination:						
	Highly likely	Moderately likely	Neither likely nor unlikely	Moderately unlikely	Highly unlikely	No comment/ not sure
Simple – elements stand alone (a leaflet, a video)						
Modest – some elements stand alone (the leaflet is backed up by a face to face training event)						
Complex – multiple interdependent parts (trial participants meet monthly to practise their technical skills on a simulator hosted in a protected research environment)						

Q3 Different communities of people may be involved in studies. Assuming that each named category provides participants to both trial and control cohorts, what do you think the likelihood is that contamination could occur if the intervention is:						
	Highly likely	Moderately likely	Neither likely nor unlikely	Moderately unlikely	Highly unlikely	No comment/ not sure
Targeted at patients						
Targeted at health professionals						
Targeted at any section of the general population						

Q4 Please tell us the likelihood of contamination in the following situations where there is movement between intervention and control locations:						
	Highly likely	Moderately likely	Neither likely nor unlikely	Moderately unlikely	Highly unlikely	No comment/ not sure
Staff looking after patients in the intervention arm move to look after patients in the control arm						
Patients at the intervention arm setting moving to the control arm setting						

(If necessary click here to refer back to the definitions.)

Q5 Please tell us the likeliness of contamination where the controls and trials are:						
	Highly likely	Moderately likely	Neither likely nor unlikely	Moderately unlikely	Highly unlikely	No comment/ not sure
From the same geographical site						
From nearby sites in same community						
Geographically separated						

Q6 Please tell us the likeliness of contamination where the controls and trials are:						
	Highly likely	Moderately likely	Neither likely nor unlikely	Moderately unlikely	Highly unlikely	No comment/ not sure
From same social networks						
From overlapping social networks						
Unlikely to have social networks in common						

(If necessary click here to refer back to the definitions.)

Q7 How likely is contamination where the intervention has:						
	Highly likely	Moderately likely	Neither likely nor unlikely	Moderately unlikely	Highly unlikely	No comment/ not sure
Little desirability e.g. a set of statistical exercises to help patients understand risk						
Medium desirability						
High desirability e.g. a free video on exercise for weight loss						

Q8 Please tell us the likeliness of contamination where the intervention is:						
	Highly likely	Moderately likely	Neither likely nor unlikely	Moderately unlikely	Highly unlikely	No comment/ not sure
Easy to transfer						
Difficult to transfer						

Q9 Please tell us the likeliness of contamination where:						
	Highly likely	Moderately likely	Neither likely nor unlikely	Moderately unlikely	Highly unlikely	No comment/ not sure
No efforts are made to educate participants against contamination and how it could occur						
Participants are educated to avoid contamination						

(If necessary click here to refer back to the definitions.)

Q10 In the following types of study design how likely is it that contamination will occur?						
	Highly likely	Moderately likely	Neither likely nor unlikely	Moderately unlikely	Highly unlikely	No comment/ not sure
Individually randomised parallel group trial						
Cluster randomised trial						
Before/after comparisons						
Repeated time series						

Q11 Are there any situations which usually require cluster randomisation to avoid contamination? (please specify which in the box below)

(If necessary click here to refer back to the definitions.)

Q12 Educational studies may use different types of media to convey new information. For the following types of media how likely is it that contamination will occur?						
	Highly likely	Moderately likely	Neither likely nor unlikely	Moderately unlikely	Highly unlikely	No comment/ not sure
Written information e.g. booklet						
Audiovisual e.g. video, CD-ROM						
Media output (TV, radio)						
Training event (single attendance)						
Training programme (multiple attendances)						
Use of specialist resources (models, simulators)						
Computer reminders (preset 'popups' on computer)						
Other						

Q13 Different communities of people may be involved in studies. Assuming that each named category provides participants to both trial and control cohorts, please tell us the likeliness of contamination for each of the following:						
	Highly likely	Moderately likely	Neither likely nor unlikely	Moderately unlikely	Highly unlikely	No comment/ not sure
Patients living in the same residence						
Patients living in the same geographical locality						
Patients with shared social networks						
Patients whose healthcare comes from the same practice						

(If necessary click here to refer back to the definitions.)

Q14 Similarly, if health professionals are the focus of studies, please tell us the likeliness of contamination for each of the following:						
	Highly likely	Moderately likely	Neither likely nor unlikely	Moderately unlikely	Highly unlikely	No comment/ not sure
Health professionals sharing a place of employment						
Health professionals in the same workplace team						
Health professionals in the same clinical directorate or equivalent						
Health professionals sharing an employer						

Q15 To what extent do you agree that the following methods will be effective in avoiding contamination:						
	Strongly agree	Tend to agree	Neither agree nor disagree	Tend to disagree	Strongly disagree	Don't know/NA
Avoid allocation of subjects to a less desirable arm (preference studies)						
Geographical/social separation of subjects						
Restriction on medium of intervention						
Education of participants to avoid transfer of intervention						

(If necessary click here to refer back to the definitions.)

Part 2 – Qualitative stage (please give examples/references from your own work if possible)
How do you think researchers can protect against contamination in controlled educational intervention studies?

Under which circumstances would you employ a particular study design specifically to avoid contamination?

How would you know if contamination were occurring in a study?

(If necessary click here to refer back to the definitions.)

Part 3 – Comments about Round 1 of the Delphi
Do you have any comments about the responses you have given in the questionnaire?

Are there any ways in which we could increase the clarity of the questionnaire?

Do you accept our definitions of the relevant concepts? If not, how could we improve them?

Are there any key areas in contamination of educational interventions which we have not covered?

Additional information:
Please enter your details below (all responses will be kept anonymous)
Name:
Email address:
Job title:
Relevant interest:

Thank you for completing this survey. Click on the submit answers button to *submit* your answers.

This may take a few seconds.

Appendix 2

Second-round Delphi questionnaire

School of Medicine, Health Policy and Practice
University of East Anglia

DELPHI questionnaire - Contamination in trials of educational intervention

Final Round

Section 1 - Questions:
The Round 1 responses are below the selection boxes as percentages, your Round 1 response is in the leftmost column of the table beneath the factor under consideration. Please indicate your response in the light of this information. You can stick with your original response or change your mind, both are valid.

Q1 Educational interventions can focus on improving knowledge, altering skill set, changing behaviours, or changing attitudes. For each of the following educational interventions please state the likeliness of transfer of the intervention to controls:

	Highly likely	Moderately likely	Neither likely nor unlikely	Moderately unlikely	Highly unlikely	No comment / not sure
Transfer of new knowledge	○ 22%	○ 49%	○ 8%	○ 14%	○ 3%	○ 5%
Transfer of new skills	○ 5%	○ 24%	○ 19%	○ 41%	○ 5%	○ 5%
Altering behaviours	○ 8%	○ 24%	○ 24%	○ 27%	○ 8%	○ 8%
Changing attitudes	○ 8%	○ 22%	○ 27%	○ 22%	○ 14%	○ 8%

(If necessary click here to refer back to the definitions.)

Q2 Interventions can be defined as Simple, Modest and Complex. Please tell us for each of the following designs the likeliness of contamination:

	Highly likely	Moderately likely	Neither likely nor unlikely	Moderately unlikely	Highly unlikely	No comment / not sure
Simple – elements stand alone (a leaflet, a video)	○ 27%	○ 35%	○ 22%	○ 8%	○ 3%	○ 5%
Modest - some elements stand alone (the leaflet is backed up by a face to face training event)	○ 3%	○ 46%	○ 32%	○ 14%	○ 0%	○ 5%
Complex – multiple interdependent parts (trial participants meet monthly to practice their technical skills on a simulator hosted in a protected research environment).	○ 5%	○ 22%	○ 19%	○ 30%	○ 19%	○ 5%

Q3 Different communities of people may be involved in studies. Assuming that each named category provides participants to both trial and control cohorts, what do you think the likelihood is that contamination could occur if the intervention is:

	Highly likely	Moderately likely	Neither likely nor unlikely	Moderately unlikely	Highly unlikely	No comment / not sure
Targeted at patients	○ 3%	○ 41%	○ 16%	○ 32%	○ 0%	○ 5%
Targeted at health professionals	○ 35%	○ 38%	○ 14%	○ 5%	○ 0%	○ 8%
Targeted at any section of the general population	○ 5%	○ 27%	○ 22%	○ 24%	○ 11%	○ 8%

Q4 Please tell us the likelihood of contamination in the following situations where there is movement between intervention and control locations:	Highly likely	Moderately likely	Neither likely nor unlikely	Moderately unlikely	Highly unlikely	No comment / not sure
Staff looking after patients in the intervention arm move to look after patients in the control arm	49%	38%	5%	0%	3%	5%
Patients at the intervention arm setting moving to the control arm setting	22%	46%	14%	11%	0%	8%

(If necessary click here to refer back to the definitions.)

Q5 Please tell us the likeliness of contamination where the controls and trials are:	Highly likely	Moderately likely	Neither likely nor unlikely	Moderately unlikely	Highly unlikely	No comment / not sure
From the same geographical site	35%	46%	11%	0%	0%	8%
From nearby sites in same community	3%	43%	30%	16%	0%	8%
Geographically separated	0%	5%	5%	46%	35%	8%

Q6 Please tell us the likeliness of contamination where the controls and trials are:	Highly likely	Moderately likely	Neither likely nor unlikely	Moderately unlikely	Highly unlikely	No comment / not sure
From same social networks	57%	32%	3%	0%	0%	8%
From overlapping social networks	16%	49%	19%	5%	0%	11%
Unlikely to have social networks in common	0%	5%	5%	35%	46%	8%

(If necessary click here to refer back to the definitions.)

Q7 How likely is contamination where the intervention has:	Highly likely	Moderately likely	Neither likely nor unlikely	Moderately unlikely	Highly unlikely	No comment / not sure
Little desirability e.g. a set of statistical exercises to help patients understand risk	0%	5%	27%	32%	27%	8%
Medium desirability	0%	32%	49%	8%	0%	11%
High desirability e.g. a free video on exercise for weight loss	35%	41%	11%	3%	0%	11%

Q8 Please tell us the likeliness of contamination where the intervention is:	Highly likely	Moderately likely	Neither likely nor unlikely	Moderately unlikely	Highly unlikely	No comment / not sure
Easy to transfer	30%	46%	5%	3%	0%	16%
Difficult to transfer	0%	14%	8%	38%	24%	16%

Q9 Please tell us the likeliness of contamination where:	Highly likely	Moderately likely	Neither likely nor unlikely	Moderately unlikely	Highly unlikely	No comment / not sure
No efforts are made to educate participants against contamination and how it could occur	11%	49%	22%	8%	0%	11%
Participants are educated to avoid contamination	0%	24%	16%	32%	11%	16%

(If necessary click here to refer back to the definitions.)

Q10 In the following types of study design how likely is it that contamination will occur?	Highly likely	Moderately likely	Neither likely nor unlikely	Moderately unlikely	Highly unlikely	No comment / not sure
Individually randomised parallel group trial	8%	51%	19%	8%	3%	11%
Cluster randomised trial	0%	14%	14%	54%	11%	8%
Before/after comparisons	3%	32%	14%	22%	8%	22%
Repeated time series	16%	22%	22%	11%	3%	27%

(If necessary click here to refer back to the definitions.)

Q11 Educational studies may use different types of media to convey new information. For the following types of media how likely is it that contamination will occur?	Highly likely	Moderately likely	Neither likely nor unlikely	Moderately unlikely	Highly unlikely	No comment / not sure
Written information e.g. booklet	22%	41%	22%	11%	0%	5%
Audiovisual e.g. video, CD-ROM	14%	54%	19%	8%	0%	5%
Media output (TV, radio)	38%	43%	14%	0%	0%	5%
Training event (single attendance)	0%	22%	24%	35%	11%	8%
Training programme (multiple attendances)	3%	24%	11%	38%	16%	8%
Use of specialist resources (models, simulators)	0%	16%	16%	35%	24%	8%
Computer reminders (preset 'popups' on computer)	5%	14%	22%	22%	16%	22%

Q12 Different communities of people may be involved in studies. Assuming that each named category provides participants to both trial and control cohorts, please tell us the likeliness of contamination for each of the following:	Highly likely	Moderately likely	Neither likely nor unlikely	Moderately unlikely	Highly unlikely	No comment / not sure
Patients living in the same residence	73%	19%	3%	0%	0%	5%
Patients living in the same geographical locality	5%	49%	32%	8%	0%	5%
Patients with shared social networks	16%	70%	5%	3%	0%	5%
Patients whose health care comes from the same practice	8%	57%	22%	8%	0%	5%

(If necessary click here to refer back to the definitions.)

Q13 Similarly, if health professionals are the focus of studies, please tell us the likeliness of contamination for each of the following:	Highly likely	Moderately likely	Neither likely nor unlikely	Moderately unlikely	Highly unlikely	No comment / not sure
Health professionals sharing a place of employment	41%	43%	8%	3%	0%	5%
Health professionals in the same workplace team	70%	24%	0%	0%	0%	5%
Health professionals in the same clinical directorate or equivalent	19%	54%	11%	5%	0%	11%
Health professionals sharing an employer	3%	38%	32%	11%	3%	14%

Q14 To what extent do you agree that the following methods will be effective in avoiding contamination:	Strongly agree	Tend to agree	Neither agree nor disagree	Tend to disagree	Strongly disagree	Don't know / N/A
Avoid allocation of subjects to a less desirable arm (preference studies)	3%	22%	19%	32%	14%	11%
Geographical / social separation of subjects	16%	76%	3%	0%	0%	5%
Restriction on medium of intervention	5%	43%	27%	11%	3%	11%
Education of participants to avoid transfer of intervention	11%	35%	27%	19%	0%	8%

(If necessary click here to refer back to the definitions.)

Section 2

We now apply these general assessments to two specific types of educational intervention. These are examples of 1) a trial of a patient-targeted intervention and 2) a trial of a health professional-targeted intervention. Both interventions could lead to contamination because of their ease of transferability, high desirability, and simplicity. Taking the interventions as given, we ask you rank the contextual factors most likely to aggravate contamination.

For each example we provide details of factors likely to lead to contamination as ranked by experts in Round 1 (1 = most likely to lead to contamination, 8 = least likely to lead to contamination).

Please read the descriptions of the intervention, and then consider whether or not you agree with the ordering of factors. Each factor should be considered independently of the other factors as some may be more applicable in cluster randomised trials and others in individually randomised trials. If you do not agree with the ordering below please supply a new ordering.

Example 1: Patient Targeted Intervention

Intervention patients are supplied with a free workout video to aid weight loss. Control patients receive no intervention.

I agree with presented ranking (please tick here) ☐
Or, provide own ranking below

Rank	Description of factor	New Rank
1	Health professionals administering the intervention are in the same workplace team as health professionals administering the control	
2	Patients receiving the intervention live in the same residence as patients receiving the control	
3	Patients receiving the intervention are from the same social network as patients receiving the control	
4	Staff move from intervention arm to control arm	
5	Health professionals in the intervention arm sharing a place of employment with health professionals in the control arm	
6	Patients receiving the intervention are from the same geographical site as patients receiving the control	
7	Patients move from intervention arm to control arm	
8	No effort is made to educate participants against contamination	

Example 2: Health Professional Targeted Intervention

Intervention health professionals receive a CD based computer program for managing diabetes in primary care using NICE guidelines. Control health professionals receive no intervention. Patient outcome is used as measure of effect.

I agree with presented ranking (please tick here) ☐
Or, provide own ranking below

Rank	Description of factor	New Rank
1	Health professionals receiving the intervention are in the same workplace team as health professionals receiving the control	
2	Patients in the intervention group live in the same residence as patients in the control group	
3	Patients in the intervention group are from the same social network as patients in the control group	
4	Staff move from intervention arm to control arm	
5	Health professionals receiving the intervention share a place of employment with health professionals receiving the control	
6	Patients in the intervention group are from the same geographical site as patients in the control group	
7	Patients move from intervention arm to control arm	
8	No effort is made to educate participants against contamination	

Thank you for completing this survey. Click on the submit answers button to *submit* your answers.

This may take a few seconds.

Submit answers

Appendix 3

Delphi questionnaire definitions

School of Medicine, Health Policy and Practice
University of East Anglia

DELPHI questionnaire – Contamination in trials of educational interventions

Some definitions:

Contamination: an intervention intended for members of the trial arm in a study being actively received by some participants in the other (control) arm. This excludes cross-over between arms that is known to the researchers and can be adjusted for, and **excludes** non-adherence, where participants have not utilised the intervention to its full potential.

Mechanisms of contamination in educational trials could include, for example:

- educational media being passed on to members of the control group by trial arm participants
- discussion between trained and untrained subjects
- transfer of trained and untrained people between intervention and control sites.

Educational interventions: one or more methods to alter their knowledge, skills and/or behaviours.

Single faceted intervention: one component only, e.g. a leaflet, a single nurse contact.

Multi-faceted intervention: whole intervention includes more than one component, e.g. a training meeting PLUS written handbook PLUS a follow-up visit to check progress.

Major risk: contamination is so likely to occur that cluster randomisation should be considered instead of individual randomisation.

Cluster randomisation: the randomised allocation of a group of participants to one or other arms of a study.

Intraclass correlation: the extent to which characteristics of participants in one cluster or group are acting independently or not.

Individually randomised parallel group trial: a trial that compares two groups of subjects, one of which receives the intervention of interest and one of which is a control group. The allocation of subjects to the control or intervention arm is made by randomly allocating each subject to a treatment group and following them up at the same time.

Controlled before and after study (CBA): CBAs incorporate a non-randomised control group that, it is hoped, will experience the same secular and sudden changes as the intervention group. Data are collected on the control and intervention groups before the intervention is introduced and then further data are collected after the intervention has been introduced. Also known as a quasi-experiment.

Repeated or interrupted time series (RTS/ITS): the repeated time series design involves repeated measurement of trial subjects over time, encompassing periods both prior to and after implementation of the intervention. The goal of such an analysis is to assess whether the treatment has 'interrupted' or changed a pattern established prior to the trial's implementation.

Health Technology Assessment reports published to date

Volume 1, 1997

No. 1
Home parenteral nutrition: a systematic review.
By Richards DM, Deeks JJ, Sheldon TA, Shaffer JL.

No. 2
Diagnosis, management and screening of early localised prostate cancer.
A review by Selley S, Donovan J, Faulkner A, Coast J, Gillatt D.

No. 3
The diagnosis, management, treatment and costs of prostate cancer in England and Wales.
A review by Chamberlain J, Melia J, Moss S, Brown J.

No. 4
Screening for fragile X syndrome.
A review by Murray J, Cuckle H, Taylor G, Hewison J.

No. 5
A review of near patient testing in primary care.
By Hobbs FDR, Delaney BC, Fitzmaurice DA, Wilson S, Hyde CJ, Thorpe GH, *et al*.

No. 6
Systematic review of outpatient services for chronic pain control.
By McQuay HJ, Moore RA, Eccleston C, Morley S, de C Williams AC.

No. 7
Neonatal screening for inborn errors of metabolism: cost, yield and outcome.
A review by Pollitt RJ, Green A, McCabe CJ, Booth A, Cooper NJ, Leonard JV, *et al*.

No. 8
Preschool vision screening.
A review by Snowdon SK, Stewart-Brown SL.

No. 9
Implications of socio-cultural contexts for the ethics of clinical trials.
A review by Ashcroft RE, Chadwick DW, Clark SRL, Edwards RHT, Frith L, Hutton JL.

No. 10
A critical review of the role of neonatal hearing screening in the detection of congenital hearing impairment.
By Davis A, Bamford J, Wilson I, Ramkalawan T, Forshaw M, Wright S.

No. 11
Newborn screening for inborn errors of metabolism: a systematic review.
By Seymour CA, Thomason MJ, Chalmers RA, Addison GM, Bain MD, Cockburn F, *et al*.

No. 12
Routine preoperative testing: a systematic review of the evidence.
By Munro J, Booth A, Nicholl J.

No. 13
Systematic review of the effectiveness of laxatives in the elderly.
By Petticrew M, Watt I, Sheldon T.

No. 14
When and how to assess fast-changing technologies: a comparative study of medical applications of four generic technologies.
A review by Mowatt G, Bower DJ, Brebner JA, Cairns JA, Grant AM, McKee L.

Volume 2, 1998

No. 1
Antenatal screening for Down's syndrome.
A review by Wald NJ, Kennard A, Hackshaw A, McGuire A.

No. 2
Screening for ovarian cancer: a systematic review.
By Bell R, Petticrew M, Luengo S, Sheldon TA.

No. 3
Consensus development methods, and their use in clinical guideline development.
A review by Murphy MK, Black NA, Lamping DL, McKee CM, Sanderson CFB, Askham J, *et al*.

No. 4
A cost–utility analysis of interferon beta for multiple sclerosis.
By Parkin D, McNamee P, Jacoby A, Miller P, Thomas S, Bates D.

No. 5
Effectiveness and efficiency of methods of dialysis therapy for end-stage renal disease: systematic reviews.
By MacLeod A, Grant A, Donaldson C, Khan I, Campbell M, Daly C, *et al*.

No. 6
Effectiveness of hip prostheses in primary total hip replacement: a critical review of evidence and an economic model.
By Faulkner A, Kennedy LG, Baxter K, Donovan J, Wilkinson M, Bevan G.

No. 7
Antimicrobial prophylaxis in colorectal surgery: a systematic review of randomised controlled trials.
By Song F, Glenny AM.

No. 8
Bone marrow and peripheral blood stem cell transplantation for malignancy.
A review by Johnson PWM, Simnett SJ, Sweetenham JW, Morgan GJ, Stewart LA.

No. 9
Screening for speech and language delay: a systematic review of the literature.
By Law J, Boyle J, Harris F, Harkness A, Nye C.

No. 10
Resource allocation for chronic stable angina: a systematic review of effectiveness, costs and cost-effectiveness of alternative interventions.
By Sculpher MJ, Petticrew M, Kelland JL, Elliott RA, Holdright DR, Buxton MJ.

No. 11
Detection, adherence and control of hypertension for the prevention of stroke: a systematic review.
By Ebrahim S.

No. 12
Postoperative analgesia and vomiting, with special reference to day-case surgery: a systematic review.
By McQuay HJ, Moore RA.

No. 13
Choosing between randomised and nonrandomised studies: a systematic review.
By Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C.

No. 14
Evaluating patient-based outcome measures for use in clinical trials.
A review by Fitzpatrick R, Davey C, Buxton MJ, Jones DR.

No. 15
Ethical issues in the design and conduct of randomised controlled trials.
A review by Edwards SJL, Lilford RJ, Braunholtz DA, Jackson JC, Hewison J, Thornton J.

No. 16
Qualitative research methods in health technology assessment: a review of the literature.
By Murphy E, Dingwall R, Greatbatch D, Parker S, Watson P.

No. 17
The costs and benefits of paramedic skills in pre-hospital trauma care.
By Nicholl J, Hughes S, Dixon S, Turner J, Yates D.

No. 18
Systematic review of endoscopic ultrasound in gastro-oesophageal cancer.
By Harris KM, Kelly S, Berry E, Hutton J, Roderick P, Cullingworth J, *et al*.

No. 19
Systematic reviews of trials and other studies.
By Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F.

No. 20
Primary total hip replacement surgery: a systematic review of outcomes and modelling of cost-effectiveness associated with different prostheses.
A review by Fitzpatrick R, Shortall E, Sculpher M, Murray D, Morris R, Lodge M, *et al*.

Volume 3, 1999

No. 1
Informed decision making: an annotated bibliography and systematic review.
By Bekker H, Thornton JG, Airey CM, Connelly JB, Hewison J, Robinson MB, *et al*.

No. 2
Handling uncertainty when performing economic evaluation of healthcare interventions.
A review by Briggs AH, Gray AM.

No. 3
The role of expectancies in the placebo effect and their use in the delivery of health care: a systematic review.
By Crow R, Gage H, Hampson S, Hart J, Kimber A, Thomas H.

No. 4
A randomised controlled trial of different approaches to universal antenatal HIV testing: uptake and acceptability. Annex: Antenatal HIV testing – assessment of a routine voluntary approach.
By Simpson WM, Johnstone FD, Boyd FM, Goldberg DJ, Hart GJ, Gormley SM, *et al*.

No. 5
Methods for evaluating area-wide and organisation-based interventions in health and health care: a systematic review.
By Ukoumunne OC, Gulliford MC, Chinn S, Sterne JAC, Burney PGJ.

No. 6
Assessing the costs of healthcare technologies in clinical trials.
A review by Johnston K, Buxton MJ, Jones DR, Fitzpatrick R.

No. 7
Cooperatives and their primary care emergency centres: organisation and impact.
By Hallam L, Henthorne K.

No. 8
Screening for cystic fibrosis.
A review by Murray J, Cuckle H, Taylor G, Littlewood J, Hewison J.

No. 9
A review of the use of health status measures in economic evaluation.
By Brazier J, Deverill M, Green C, Harper R, Booth A.

No. 10
Methods for the analysis of quality-of-life and survival data in health technology assessment.
A review by Billingham LJ, Abrams KR, Jones DR.

No. 11
Antenatal and neonatal haemoglobinopathy screening in the UK: review and economic analysis.
By Zeuner D, Ades AE, Karnon J, Brown J, Dezateux C, Anionwu EN.

No. 12
Assessing the quality of reports of randomised trials: implications for the conduct of meta-analyses.
A review by Moher D, Cook DJ, Jadad AR, Tugwell P, Moher M, Jones A, *et al*.

No. 13
'Early warning systems' for identifying new healthcare technologies.
By Robert G, Stevens A, Gabbay J.

No. 14
A systematic review of the role of human papillomavirus testing within a cervical screening programme.
By Cuzick J, Sasieni P, Davies P, Adams J, Normand C, Frater A, *et al*.

No. 15
Near patient testing in diabetes clinics: appraising the costs and outcomes.
By Grieve R, Beech R, Vincent J, Mazurkiewicz J.

No. 16
Positron emission tomography: establishing priorities for health technology assessment.
A review by Robert G, Milne R.

No. 17 (Pt 1)
The debridement of chronic wounds: a systematic review.
By Bradley M, Cullum N, Sheldon T.

No. 17 (Pt 2)
Systematic reviews of wound care management: (2) Dressings and topical agents used in the healing of chronic wounds.
By Bradley M, Cullum N, Nelson EA, Petticrew M, Sheldon T, Torgerson D.

No. 18
A systematic literature review of spiral and electron beam computed tomography: with particular reference to clinical applications in hepatic lesions, pulmonary embolus and coronary artery disease.
By Berry E, Kelly S, Hutton J, Harris KM, Roderick P, Boyce JC, *et al*.

No. 19
What role for statins? A review and economic model.
By Ebrahim S, Davey Smith G, McCabe C, Payne N, Pickin M, Sheldon TA, *et al*.

No. 20
Factors that limit the quality, number and progress of randomised controlled trials.
A review by Prescott RJ, Counsell CE, Gillespie WJ, Grant AM, Russell IT, Kiauka S, *et al*.

No. 21
Antimicrobial prophylaxis in total hip replacement: a systematic review.
By Glenny AM, Song F.

No. 22
Health promoting schools and health promotion in schools: two systematic reviews.
By Lister-Sharp D, Chapman S, Stewart-Brown S, Sowden A.

No. 23
Economic evaluation of a primary care-based education programme for patients with osteoarthritis of the knee.
A review by Lord J, Victor C, Littlejohns P, Ross FM, Axford JS.

Volume 4, 2000

No. 1
The estimation of marginal time preference in a UK-wide sample (TEMPUS) project.
A review by Cairns JA, van der Pol MM.

No. 2
Geriatric rehabilitation following fractures in older people: a systematic review.
By Cameron I, Crotty M, Currie C, Finnegan T, Gillespie L, Gillespie W, *et al*.

No. 3
Screening for sickle cell disease and thalassaemia: a systematic review with supplementary research.
By Davies SC, Cronin E, Gill M, Greengross P, Hickman M, Normand C.

No. 4
Community provision of hearing aids and related audiology services.
A review by Reeves DJ, Alborz A, Hickson FS, Bamford JM.

No. 5
False-negative results in screening programmes: systematic review of impact and implications.
By Petticrew MP, Sowden AJ, Lister-Sharp D, Wright K.

No. 6
Costs and benefits of community postnatal support workers: a randomised controlled trial.
By Morrell CJ, Spiby H, Stewart P, Walters S, Morgan A.

No. 7
Implantable contraceptives (subdermal implants and hormonally impregnated intrauterine systems) versus other forms of reversible contraceptives: two systematic reviews to assess relative effectiveness, acceptability, tolerability and cost-effectiveness.
By French RS, Cowan FM, Mansour DJA, Morris S, Procter T, Hughes D, *et al*.

No. 8
An introduction to statistical methods for health technology assessment.
A review by White SJ, Ashby D, Brown PJ.

No. 9
Disease-modifying drugs for multiple sclerosis: a rapid and systematic review.
By Clegg A, Bryant J, Milne R.

No. 10
Publication and related biases.
A review by Song F, Eastwood AJ, Gilbody S, Duley L, Sutton AJ.

No. 11
Cost and outcome implications of the organisation of vascular services.
By Michaels J, Brazier J, Palfreyman S, Shackley P, Slack R.

No. 12
Monitoring blood glucose control in diabetes mellitus: a systematic review.
By Coster S, Gulliford MC, Seed PT, Powrie JK, Swaminathan R.

No. 13
The effectiveness of domiciliary health visiting: a systematic review of international studies and a selective review of the British literature.
By Elkan R, Kendrick D, Hewitt M, Robinson JJA, Tolley K, Blair M, *et al*.

No. 14
The determinants of screening uptake and interventions for increasing uptake: a systematic review.
By Jepson R, Clegg A, Forbes C, Lewis R, Sowden A, Kleijnen J.

No. 15
The effectiveness and cost-effectiveness of prophylactic removal of wisdom teeth.
A rapid review by Song F, O'Meara S, Wilson P, Golder S, Kleijnen J.

No. 16
Ultrasound screening in pregnancy: a systematic review of the clinical effectiveness, cost-effectiveness and women's views.
By Bricker L, Garcia J, Henderson J, Mugford M, Neilson J, Roberts T, *et al*.

No. 17
A rapid and systematic review of the effectiveness and cost-effectiveness of the taxanes used in the treatment of advanced breast and ovarian cancer.
By Lister-Sharp D, McDonagh MS, Khan KS, Kleijnen J.

No. 18
Liquid-based cytology in cervical screening: a rapid and systematic review.
By Payne N, Chilcott J, McGoogan E.

No. 19
Randomised controlled trial of non-directive counselling, cognitive–behaviour therapy and usual general practitioner care in the management of depression as well as mixed anxiety and depression in primary care.
By King M, Sibbald B, Ward E, Bower P, Lloyd M, Gabbay M, *et al*.

No. 20
Routine referral for radiography of patients presenting with low back pain: is patients' outcome influenced by GPs' referral for plain radiography?
By Kerry S, Hilton S, Patel S, Dundas D, Rink E, Lord J.

No. 21
Systematic reviews of wound care management: (3) antimicrobial agents for chronic wounds; (4) diabetic foot ulceration.
By O'Meara S, Cullum N, Majid M, Sheldon T.

No. 22
Using routine data to complement and enhance the results of randomised controlled trials.
By Lewsey JD, Leyland AH, Murray GD, Boddy FA.

No. 23
Coronary artery stents in the treatment of ischaemic heart disease: a rapid and systematic review.
By Meads C, Cummins C, Jolly K, Stevens A, Burls A, Hyde C.

No. 24
Outcome measures for adult critical care: a systematic review.
By Hayes JA, Black NA, Jenkinson C, Young JD, Rowan KM, Daly K, *et al*.

No. 25
A systematic review to evaluate the effectiveness of interventions to promote the initiation of breastfeeding.
By Fairbank L, O'Meara S, Renfrew MJ, Woolridge M, Sowden AJ, Lister-Sharp D.

No. 26
Implantable cardioverter defibrillators: arrhythmias. A rapid and systematic review.
By Parkes J, Bryant J, Milne R.

No. 27
Treatments for fatigue in multiple sclerosis: a rapid and systematic review.
By Brañas P, Jordan R, Fry-Smith A, Burls A, Hyde C.

No. 28
Early asthma prophylaxis, natural history, skeletal development and economy (EASE): a pilot randomised controlled trial.
By Baxter-Jones ADG, Helms PJ, Russell G, Grant A, Ross S, Cairns JA, *et al*.

No. 29
Screening for hypercholesterolaemia versus case finding for familial hypercholesterolaemia: a systematic review and cost-effectiveness analysis.
By Marks D, Wonderling D, Thorogood M, Lambert H, Humphries SE, Neil HAW.

No. 30
A rapid and systematic review of the clinical effectiveness and cost-effectiveness of glycoprotein IIb/IIIa antagonists in the medical management of unstable angina.
By McDonagh MS, Bachmann LM, Golder S, Kleijnen J, ter Riet G.

No. 31
A randomised controlled trial of prehospital intravenous fluid replacement therapy in serious trauma.
By Turner J, Nicholl J, Webber L, Cox H, Dixon S, Yates D.

No. 32
Intrathecal pumps for giving opioids in chronic pain: a systematic review.
By Williams JE, Louw G, Towlerton G.

No. 33
Combination therapy (interferon alfa and ribavirin) in the treatment of chronic hepatitis C: a rapid and systematic review.
By Shepherd J, Waugh N, Hewitson P.

No. 34
A systematic review of comparisons of effect sizes derived from randomised and non-randomised studies.
By MacLehose RR, Reeves BC, Harvey IM, Sheldon TA, Russell IT, Black AMS.

No. 35
Intravascular ultrasound-guided interventions in coronary artery disease: a systematic literature review, with decision-analytic modelling, of outcomes and cost-effectiveness.
By Berry E, Kelly S, Hutton J, Lindsay HSJ, Blaxill JM, Evans JA, *et al*.

No. 36
A randomised controlled trial to evaluate the effectiveness and cost-effectiveness of counselling patients with chronic depression.
By Simpson S, Corney R, Fitzgerald P, Beecham J.

No. 37
Systematic review of treatments for atopic eczema.
By Hoare C, Li Wan Po A, Williams H.

No. 38
Bayesian methods in health technology assessment: a review.
By Spiegelhalter DJ, Myles JP, Jones DR, Abrams KR.

No. 39
The management of dyspepsia: a systematic review.
By Delaney B, Moayyedi P, Deeks J, Innes M, Soo S, Barton P, *et al*.

No. 40
A systematic review of treatments for severe psoriasis.
By Griffiths CEM, Clark CM, Chalmers RJG, Li Wan Po A, Williams HC.

Volume 5, 2001

No. 1
Clinical and cost-effectiveness of donepezil, rivastigmine and galantamine for Alzheimer's disease: a rapid and systematic review.
By Clegg A, Bryant J, Nicholson T, McIntyre L, De Broe S, Gerard K, *et al*.

No. 2
The clinical effectiveness and cost-effectiveness of riluzole for motor neurone disease: a rapid and systematic review.
By Stewart A, Sandercock J, Bryan S, Hyde C, Barton PM, Fry-Smith A, *et al*.

No. 3
Equity and the economic evaluation of healthcare.
By Sassi F, Archard L, Le Grand J.

No. 4
Quality-of-life measures in chronic diseases of childhood.
By Eiser C, Morse R.

No. 5
Eliciting public preferences for healthcare: a systematic review of techniques.
By Ryan M, Scott DA, Reeves C, Bate A, van Teijlingen ER, Russell EM, *et al*.

No. 6
General health status measures for people with cognitive impairment: learning disability and acquired brain injury.
By Riemsma RP, Forbes CA, Glanville JM, Eastwood AJ, Kleijnen J.

No. 7
An assessment of screening strategies for fragile X syndrome in the UK.
By Pembrey ME, Barnicoat AJ, Carmichael B, Bobrow M, Turner G.

No. 8
Issues in methodological research: perspectives from researchers and commissioners.
By Lilford RJ, Richardson A, Stevens A, Fitzpatrick R, Edwards S, Rock F, *et al*.

No. 9
Systematic reviews of wound care management: (5) beds; (6) compression; (7) laser therapy, therapeutic ultrasound, electrotherapy and electromagnetic therapy.
By Cullum N, Nelson EA, Flemming K, Sheldon T.

No. 10
Effects of educational and psychosocial interventions for adolescents with diabetes mellitus: a systematic review.
By Hampson SE, Skinner TC, Hart J, Storey L, Gage H, Foxcroft D, *et al*.

No. 11
Effectiveness of autologous chondrocyte transplantation for hyaline cartilage defects in knees: a rapid and systematic review.
By Jobanputra P, Parry D, Fry-Smith A, Burls A.

No. 12
Statistical assessment of the learning curves of health technologies.
By Ramsay CR, Grant AM, Wallace SA, Garthwaite PH, Monk AF, Russell IT.

No. 13
The effectiveness and cost-effectiveness of temozolomide for the treatment of recurrent malignant glioma: a rapid and systematic review.
By Dinnes J, Cave C, Huang S, Major K, Milne R.

No. 14
A rapid and systematic review of the clinical effectiveness and cost-effectiveness of debriding agents in treating surgical wounds healing by secondary intention.
By Lewis R, Whiting P, ter Riet G, O'Meara S, Glanville J.

No. 15
Home treatment for mental health problems: a systematic review.
By Burns T, Knapp M, Catty J, Healey A, Henderson J, Watt H, *et al*.

No. 16
How to develop cost-conscious guidelines.
By Eccles M, Mason J.

No. 17
The role of specialist nurses in multiple sclerosis: a rapid and systematic review.
By De Broe S, Christopher F, Waugh N.

No. 18
A rapid and systematic review of the clinical effectiveness and cost-effectiveness of orlistat in the management of obesity.
By O'Meara S, Riemsma R, Shirran L, Mather L, ter Riet G.

No. 19
The clinical effectiveness and cost-effectiveness of pioglitazone for type 2 diabetes mellitus: a rapid and systematic review.
By Chilcott J, Wight J, Lloyd Jones M, Tappenden P.

No. 20
Extended scope of nursing practice: a multicentre randomised controlled trial of appropriately trained nurses and preregistration house officers in pre-operative assessment in elective general surgery.
By Kinley H, Czoski-Murray C, George S, McCabe C, Primrose J, Reilly C, *et al*.

No. 21
Systematic reviews of the effectiveness of day care for people with severe mental disorders: (1) Acute day hospital versus admission; (2) Vocational rehabilitation; (3) Day hospital versus outpatient care.
By Marshall M, Crowther R, Almaraz-Serrano A, Creed F, Sledge W, Kluiter H, *et al*.

No. 22
The measurement and monitoring of surgical adverse events.
By Bruce J, Russell EM, Mollison J, Krukowski ZH.

No. 23
Action research: a systematic review and guidance for assessment.
By Waterman H, Tillen D, Dickson R, de Koning K.

No. 24
A rapid and systematic review of the clinical effectiveness and cost-effectiveness of gemcitabine for the treatment of pancreatic cancer.
By Ward S, Morris E, Bansback N, Calvert N, Crellin A, Forman D, *et al*.

No. 25
A rapid and systematic review of the evidence for the clinical effectiveness and cost-effectiveness of irinotecan, oxaliplatin and raltitrexed for the treatment of advanced colorectal cancer.
By Lloyd Jones M, Hummel S, Bansback N, Orr B, Seymour M.

No. 26
Comparison of the effectiveness of inhaler devices in asthma and chronic obstructive airways disease: a systematic review of the literature.
By Brocklebank D, Ram F, Wright J, Barry P, Cates C, Davies L, *et al*.

No. 27
The cost-effectiveness of magnetic resonance imaging for investigation of the knee joint.
By Bryan S, Weatherburn G, Bungay H, Hatrick C, Salas C, Parry D, *et al*.

No. 28
A rapid and systematic review of the clinical effectiveness and cost-effectiveness of topotecan for ovarian cancer.
By Forbes C, Shirran L, Bagnall A-M, Duffy S, ter Riet G.

No. 29
Superseded by a report published in a later volume.

No. 30
The role of radiography in primary care patients with low back pain of at least 6 weeks duration: a randomised (unblinded) controlled trial.
By Kendrick D, Fielding K, Bentley E, Miller P, Kerslake R, Pringle M.

No. 31
Design and use of questionnaires: a review of best practice applicable to surveys of health service staff and patients.
By McColl E, Jacoby A, Thomas L, Soutter J, Bamford C, Steen N, *et al*.

No. 32
A rapid and systematic review of the clinical effectiveness and cost-effectiveness of paclitaxel, docetaxel, gemcitabine and vinorelbine in non-small-cell lung cancer.
By Clegg A, Scott DA, Sidhu M, Hewitson P, Waugh N.

No. 33
Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives.
By Brookes ST, Whitley E, Peters TJ, Mulheran PA, Egger M, Davey Smith G.

No. 34
Depot antipsychotic medication in the treatment of patients with schizophrenia: (1) Meta-review; (2) Patient and nurse attitudes.
By David AS, Adams C.

No. 35
A systematic review of controlled trials of the effectiveness and cost-effectiveness of brief psychological treatments for depression.
By Churchill R, Hunot V, Corney R, Knapp M, McGuire H, Tylee A, *et al*.

No. 36
Cost analysis of child health surveillance.
By Sanderson D, Wright D, Acton C, Duree D.

Volume 6, 2002

No. 1
A study of the methods used to select review criteria for clinical audit.
By Hearnshaw H, Harker R, Cheater F, Baker R, Grimshaw G.

No. 2
Fludarabine as second-line therapy for B cell chronic lymphocytic leukaemia: a technology assessment.
By Hyde C, Wake B, Bryan S, Barton P, Fry-Smith A, Davenport C, *et al*.

No. 3
Rituximab as third-line treatment for refractory or recurrent Stage III or IV follicular non-Hodgkin's lymphoma: a systematic review and economic evaluation.
By Wake B, Hyde C, Bryan S, Barton P, Song F, Fry-Smith A, *et al*.

No. 4
A systematic review of discharge arrangements for older people.
By Parker SG, Peet SM, McPherson A, Cannaby AM, Baker R, Wilson A, *et al*.

No. 5
The clinical effectiveness and cost-effectiveness of inhaler devices used in the routine management of chronic asthma in older children: a systematic review and economic evaluation.
By Peters J, Stevenson M, Beverley C, Lim J, Smith S.

No. 6
The clinical effectiveness and cost-effectiveness of sibutramine in the management of obesity: a technology assessment.
By O'Meara S, Riemsma R, Shirran L, Mather L, ter Riet G.

No. 7
The cost-effectiveness of magnetic resonance angiography for carotid artery stenosis and peripheral vascular disease: a systematic review.
By Berry E, Kelly S, Westwood ME, Davies LM, Gough MJ, Bamford JM, *et al*.

No. 8
Promoting physical activity in South Asian Muslim women through 'exercise on prescription'.
By Carroll B, Ali N, Azam N.

No. 9
Zanamivir for the treatment of influenza in adults: a systematic review and economic evaluation.
By Burls A, Clark W, Stewart T, Preston C, Bryan S, Jefferson T, *et al*.

No. 10
A review of the natural history and epidemiology of multiple sclerosis: implications for resource allocation and health economic models.
By Richards RG, Sampson FC, Beard SM, Tappenden P.

No. 11
Screening for gestational diabetes: a systematic review and economic evaluation.
By Scott DA, Loveman E, McIntyre L, Waugh N.

No. 12
The clinical effectiveness and cost-effectiveness of surgery for people with morbid obesity: a systematic review and economic evaluation.
By Clegg AJ, Colquitt J, Sidhu MK, Royle P, Loveman E, Walker A.

No. 13
The clinical effectiveness of trastuzumab for breast cancer: a systematic review.
By Lewis R, Bagnall A-M, Forbes C, Shirran E, Duffy S, Kleijnen J, *et al*.

No. 14
The clinical effectiveness and cost-effectiveness of vinorelbine for breast cancer: a systematic review and economic evaluation.
By Lewis R, Bagnall A-M, King S, Woolacott N, Forbes C, Shirran L, *et al*.

No. 15
A systematic review of the effectiveness and cost-effectiveness of metal-on-metal hip resurfacing arthroplasty for treatment of hip disease.
By Vale L, Wyness L, McCormack K, McKenzie L, Brazzelli M, Stearns SC.

No. 16
The clinical effectiveness and cost-effectiveness of bupropion and nicotine replacement therapy for smoking cessation: a systematic review and economic evaluation.
By Woolacott NF, Jones L, Forbes CA, Mather LC, Sowden AJ, Song FJ, *et al*.

No. 17
A systematic review of effectiveness and economic evaluation of new drug treatments for juvenile idiopathic arthritis: etanercept.
By Cummins C, Connock M, Fry-Smith A, Burls A.

No. 18
Clinical effectiveness and cost-effectiveness of growth hormone in children: a systematic review and economic evaluation.
By Bryant J, Cave C, Mihaylova B, Chase D, McIntyre L, Gerard K, *et al*.

No. 19
Clinical effectiveness and cost-effectiveness of growth hormone in adults in relation to impact on quality of life: a systematic review and economic evaluation.
By Bryant J, Loveman E, Chase D, Mihaylova B, Cave C, Gerard K, *et al*.

No. 20
Clinical medication review by a pharmacist of patients on repeat prescriptions in general practice: a randomised controlled trial.
By Zermansky AG, Petty DR, Raynor DK, Lowe CJ, Freementle N, Vail A.

No. 21
The effectiveness of infliximab and etanercept for the treatment of rheumatoid arthritis: a systematic review and economic evaluation.
By Jobanputra P, Barton P, Bryan S, Burls A.

No. 22
A systematic review and economic evaluation of computerised cognitive behaviour therapy for depression and anxiety.
By Kaltenthaler E, Shackley P, Stevens K, Beverley C, Parry G, Chilcott J.

No. 23
A systematic review and economic evaluation of pegylated liposomal doxorubicin hydrochloride for ovarian cancer.
By Forbes C, Wilby J, Richardson G, Sculpher M, Mather L, Reimsma R.

No. 24
A systematic review of the effectiveness of interventions based on a stages-of-change approach to promote individual behaviour change.
By Riemsma RP, Pattenden J, Bridle C, Sowden AJ, Mather L, Watt IS, *et al*.

No. 25
A systematic review update of the clinical effectiveness and cost-effectiveness of glycoprotein IIb/IIIa antagonists.
By Robinson M, Ginnelly L, Sculpher M, Jones L, Riemsma R, Palmer S, *et al*.

No. 26
A systematic review of the effectiveness, cost-effectiveness and barriers to implementation of thrombolytic and neuroprotective therapy for acute ischaemic stroke in the NHS.
By Sandercock P, Berge E, Dennis M, Forbes J, Hand P, Kwan J, *et al*.

No. 27
A randomised controlled crossover trial of nurse practitioner versus doctor-led outpatient care in a bronchiectasis clinic.
By Caine N, Sharples LD, Hollingworth W, French J, Keogan M, Exley A, *et al*.

No. 28
Clinical effectiveness and cost – consequences of selective serotonin reuptake inhibitors in the treatment of sex offenders.
By Adi Y, Ashcroft D, Browne K, Beech A, Fry-Smith A, Hyde C.

No. 29
Treatment of established osteoporosis: a systematic review and cost–utility analysis.
By Kanis JA, Brazier JE, Stevenson M, Calvert NW, Lloyd Jones M.

No. 30
Which anaesthetic agents are cost-effective in day surgery? Literature review, national survey of practice and randomised controlled trial.
By Elliott RA Payne K, Moore JK, Davies LM, Harper NJN, St Leger AS, *et al*.

No. 31
Screening for hepatitis C among injecting drug users and in genitourinary medicine clinics: systematic reviews of effectiveness, modelling study and national survey of current practice.
By Stein K, Dalziel K, Walker A, McIntyre L, Jenkins B, Horne J, *et al*.

No. 32
The measurement of satisfaction with healthcare: implications for practice from a systematic review of the literature.
By Crow R, Gage H, Hampson S, Hart J, Kimber A, Storey L, *et al*.

No. 33
The effectiveness and cost-effectiveness of imatinib in chronic myeloid leukaemia: a systematic review.
By Garside R, Round A, Dalziel K, Stein K, Royle R.

No. 34
A comparative study of hypertonic saline, daily and alternate-day rhDNase in children with cystic fibrosis.
By Suri R, Wallis C, Bush A, Thompson S, Normand C, Flather M, *et al*.

No. 35
A systematic review of the costs and effectiveness of different models of paediatric home care.
By Parker G, Bhakta P, Lovett CA, Paisley S, Olsen R, Turner D, *et al*.

Volume 7, 2003

No. 1
How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study.
By Egger M, Jüni P, Bartlett C, Holenstein F, Sterne J.

No. 2
Systematic review of the effectiveness and cost-effectiveness, and economic evaluation, of home versus hospital or satellite unit haemodialysis for people with end-stage renal failure.
By Mowatt G, Vale L, Perez J, Wyness L, Fraser C, MacLeod A, *et al*.

No. 3
Systematic review and economic evaluation of the effectiveness of infliximab for the treatment of Crohn's disease.
By Clark W, Raftery J, Barton P, Song F, Fry-Smith A, Burls A.

No. 4
A review of the clinical effectiveness and cost-effectiveness of routine anti-D prophylaxis for pregnant women who are rhesus negative.
By Chilcott J, Lloyd Jones M, Wight J, Forman K, Wray J, Beverley C, *et al*.

No. 5
Systematic review and evaluation of the use of tumour markers in paediatric oncology: Ewing's sarcoma and neuroblastoma.
By Riley RD, Burchill SA, Abrams KR, Heney D, Lambert PC, Jones DR, *et al*.

No. 6
The cost-effectiveness of screening for *Helicobacter pylori* to reduce mortality and morbidity from gastric cancer and peptic ulcer disease: a discrete-event simulation model.
By Roderick P, Davies R, Raftery J, Crabbe D, Pearce R, Bhandari P, *et al*.

No. 7
The clinical effectiveness and cost-effectiveness of routine dental checks: a systematic review and economic evaluation.
By Davenport C, Elley K, Salas C, Taylor-Weetman CL, Fry-Smith A, Bryan S, *et al*.

No. 8
A multicentre randomised controlled trial assessing the costs and benefits of using structured information and analysis of women's preferences in the management of menorrhagia.
By Kennedy ADM, Sculpher MJ, Coulter A, Dwyer N, Rees M, Horsley S, *et al*.

No. 9
Clinical effectiveness and cost–utility of photodynamic therapy for wet age-related macular degeneration: a systematic review and economic evaluation.
By Meads C, Salas C, Roberts T, Moore D, Fry-Smith A, Hyde C.

No. 10
Evaluation of molecular tests for prenatal diagnosis of chromosome abnormalities.
By Grimshaw GM, Szczepura A, Hultén M, MacDonald F, Nevin NC, Sutton F, *et al*.

No. 11
First and second trimester antenatal screening for Down's syndrome: the results of the Serum, Urine and Ultrasound Screening Study (SURUSS).
By Wald NJ, Rodeck C, Hackshaw AK, Walters J, Chitty L, Mackinson AM.

No. 12
The effectiveness and cost-effectiveness of ultrasound locating devices for central venous access: a systematic review and economic evaluation.
By Calvert N, Hind D, McWilliams RG, Thomas SM, Beverley C, Davidson A.

No. 13
A systematic review of atypical antipsychotics in schizophrenia.
By Bagnall A-M, Jones L, Lewis R, Ginnelly L, Glanville J, Torgerson D, *et al*.

No. 14
Prostate Testing for Cancer and Treatment (ProtecT) feasibility study.
By Donovan J, Hamdy F, Neal D, Peters T, Oliver S, Brindle L, *et al*.

No. 15
Early thrombolysis for the treatment of acute myocardial infarction: a systematic review and economic evaluation.
By Boland A, Dundar Y, Bagust A, Haycox A, Hill R, Mujica Mota R, *et al*.

No. 16
Screening for fragile X syndrome: a literature review and modelling.
By Song FJ, Barton P, Sleightholme V, Yao GL, Fry-Smith A.

No. 17
Systematic review of endoscopic sinus surgery for nasal polyps.
By Dalziel K, Stein K, Round A, Garside R, Royle P.

No. 18
Towards efficient guidelines: how to monitor guideline use in primary care.
By Hutchinson A, McIntosh A, Cox S, Gilbert C.

No. 19
Effectiveness and cost-effectiveness of acute hospital-based spinal cord injuries services: systematic review.
By Bagnall A-M, Jones L, Richardson G, Duffy S, Riemsma R.

No. 20
Prioritisation of health technology assessment. The PATHS model: methods and case studies.
By Townsend J, Buxton M, Harper G.

No. 21
Systematic review of the clinical effectiveness and cost-effectiveness of tension-free vaginal tape for treatment of urinary stress incontinence.
By Cody J, Wyness L, Wallace S, Glazener C, Kilonzo M, Stearns S, *et al*.

No. 22
The clinical and cost-effectiveness of patient education models for diabetes: a systematic review and economic evaluation.
By Loveman E, Cave C, Green C, Royle P, Dunn N, Waugh N.

No. 23
The role of modelling in prioritising and planning clinical trials.
By Chilcott J, Brennan A, Booth A, Karnon J, Tappenden P.

No. 24
Cost–benefit evaluation of routine influenza immunisation in people 65–74 years of age.
By Allsup S, Gosney M, Haycox A, Regan M.

No. 25
The clinical and cost-effectiveness of pulsatile machine perfusion versus cold storage of kidneys for transplantation retrieved from heart-beating and non-heart-beating donors.
By Wight J, Chilcott J, Holmes M, Brewer N.

No. 26
Can randomised trials rely on existing electronic data? A feasibility study to explore the value of routine data in health technology assessment.
By Williams JG, Cheung WY, Cohen DR, Hutchings HA, Longo MF, Russell IT.

No. 27
Evaluating non-randomised intervention studies.
By Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, *et al*.

No. 28
A randomised controlled trial to assess the impact of a package comprising a patient-orientated, evidence-based self-help guidebook and patient-centred consultations on disease management and satisfaction in inflammatory bowel disease.
By Kennedy A, Nelson E, Reeves D, Richardson G, Roberts C, Robinson A, *et al*.

No. 29
The effectiveness of diagnostic tests for the assessment of shoulder pain due to soft tissue disorders: a systematic review.
By Dinnes J, Loveman E, McIntyre L, Waugh N.

No. 30
The value of digital imaging in diabetic retinopathy.
By Sharp PF, Olson J, Strachan F, Hipwell J, Ludbrook A, O'Donnell M, *et al*.

No. 31
Lowering blood pressure to prevent myocardial infarction and stroke: a new preventive strategy.
By Law M, Wald N, Morris J.

No. 32
Clinical and cost-effectiveness of capecitabine and tegafur with uracil for the treatment of metastatic colorectal cancer: systematic review and economic evaluation.
By Ward S, Kaltenthaler E, Cowan J, Brewer N.

No. 33
Clinical and cost-effectiveness of new and emerging technologies for early localised prostate cancer: a systematic review.
By Hummel S, Paisley S, Morgan A, Currie E, Brewer N.

No. 34
Literature searching for clinical and cost-effectiveness studies used in health technology assessment reports carried out for the National Institute for Clinical Excellence appraisal system.
By Royle P, Waugh N.

No. 35
Systematic review and economic decision modelling for the prevention and treatment of influenza A and B.
By Turner D, Wailoo A, Nicholson K, Cooper N, Sutton A, Abrams K.

No. 36
A randomised controlled trial to evaluate the clinical and cost-effectiveness of Hickman line insertions in adult cancer patients by nurses.
By Boland A, Haycox A, Bagust A, Fitzsimmons L.

No. 37
Redesigning postnatal care: a randomised controlled trial of protocol-based midwifery-led care focused on individual women's physical and psychological health needs.
By MacArthur C, Winter HR, Bick DE, Lilford RJ, Lancashire RJ, Knowles H, *et al*.

No. 38
Estimating implied rates of discount in healthcare decision-making.
By West RR, McNabb R, Thompson AGH, Sheldon TA, Grimley Evans J.

No. 39
Systematic review of isolation policies in the hospital management of methicillin-resistant *Staphylococcus aureus*: a review of the literature with epidemiological and economic modelling.
By Cooper BS, Stone SP, Kibbler CC, Cookson BD, Roberts JA, Medley GF, *et al*.

No. 40
Treatments for spasticity and pain in multiple sclerosis: a systematic review.
By Beard S, Hunn A, Wight J.

No. 41
The inclusion of reports of randomised trials published in languages other than English in systematic reviews.
By Moher D, Pham B, Lawson ML, Klassen TP.

No. 42
The impact of screening on future health-promoting behaviours and health beliefs: a systematic review.
By Bankhead CR, Brett J, Bukach C, Webster P, Stewart-Brown S, Munafo M, *et al*.

Volume 8, 2004

No. 1
What is the best imaging strategy for acute stroke?
By Wardlaw JM, Keir SL, Seymour J, Lewis S, Sandercock PAG, Dennis MS, *et al*.

No. 2
Systematic review and modelling of the investigation of acute and chronic chest pain presenting in primary care.
By Mant J, McManus RJ, Oakes RAL, Delaney BC, Barton PM, Deeks JJ, *et al*.

No. 3
The effectiveness and cost-effectiveness of microwave and thermal balloon endometrial ablation for heavy menstrual bleeding: a systematic review and economic modelling.
By Garside R, Stein K, Wyatt K, Round A, Price A.

No. 4
A systematic review of the role of bisphosphonates in metastatic disease.
By Ross JR, Saunders Y, Edmonds PM, Patel S, Wonderling D, Normand C, *et al*.

No. 5
Systematic review of the clinical effectiveness and cost-effectiveness of capecitabine (Xeloda®) for locally advanced and/or metastatic breast cancer.
By Jones L, Hawkins N, Westwood M, Wright K, Richardson G, Riemsma R.

No. 6
Effectiveness and efficiency of guideline dissemination and implementation strategies.
By Grimshaw JM, Thomas RE, MacLennan G, Fraser C, Ramsay CR, Vale L, *et al*.

No. 7
Clinical effectiveness and costs of the Sugarbaker procedure for the treatment of pseudomyxoma peritonei.
By Bryant J, Clegg AJ, Sidhu MK, Brodin H, Royle P, Davidson P.

No. 8
Psychological treatment for insomnia in the regulation of long-term hypnotic drug use.
By Morgan K, Dixon S, Mathers N, Thompson J, Tomeny M.

No. 9
Improving the evaluation of therapeutic interventions in multiple sclerosis: development of a patient-based measure of outcome.
By Hobart JC, Riazi A, Lamping DL, Fitzpatrick R, Thompson AJ.

No. 10
A systematic review and economic evaluation of magnetic resonance cholangiopancreatography compared with diagnostic endoscopic retrograde cholangiopancreatography.
By Kaltenthaler E, Bravo Vergel Y, Chilcott J, Thomas S, Blakeborough T, Walters SJ, *et al*.

No. 11
The use of modelling to evaluate new drugs for patients with a chronic condition: the case of antibodies against tumour necrosis factor in rheumatoid arthritis.
By Barton P, Jobanputra P, Wilson J, Bryan S, Burls A.

No. 12
Clinical effectiveness and cost-effectiveness of neonatal screening for inborn errors of metabolism using tandem mass spectrometry: a systematic review.
By Pandor A, Eastham J, Beverley C, Chilcott J, Paisley S.

No. 13
Clinical effectiveness and cost-effectiveness of pioglitazone and rosiglitazone in the treatment of type 2 diabetes: a systematic review and economic evaluation.
By Czoski-Murray C, Warren E, Chilcott J, Beverley C, Psyllaki MA, Cowan J.

No. 14
Routine examination of the newborn: the EMREN study. Evaluation of an extension of the midwife role including a randomised controlled trial of appropriately trained midwives and paediatric senior house officers.
By Townsend J, Wolke D, Hayes J, Davé S, Rogers C, Bloomfield L, *et al*.

No. 15
Involving consumers in research and development agenda setting for the NHS: developing an evidence-based approach.
By Oliver S, Clarke-Jones L, Rees R, Milne R, Buchanan P, Gabbay J, *et al*.

No. 16
A multi-centre randomised controlled trial of minimally invasive direct coronary bypass grafting versus percutaneous transluminal coronary angioplasty with stenting for proximal stenosis of the left anterior descending coronary artery.
By Reeves BC, Angelini GD, Bryan AJ, Taylor FC, Cripps T, Spyt TJ, *et al*.

No. 17
Does early magnetic resonance imaging influence management or improve outcome in patients referred to secondary care with low back pain? A pragmatic randomised controlled trial.
By Gilbert FJ, Grant AM, Gillan MGC, Vale L, Scott NW, Campbell MK, *et al*.

No. 18
The clinical and cost-effectiveness of anakinra for the treatment of rheumatoid arthritis in adults: a systematic review and economic analysis.
By Clark W, Jobanputra P, Barton P, Burls A.

No. 19
A rapid and systematic review and economic evaluation of the clinical and cost-effectiveness of newer drugs for treatment of mania associated with bipolar affective disorder.
By Bridle C, Palmer S, Bagnall A-M, Darba J, Duffy S, Sculpher M, *et al*.

No. 20
Liquid-based cytology in cervical screening: an updated rapid and systematic review and economic analysis.
By Karnon J, Peters J, Platt J, Chilcott J, McGoogan E, Brewer N.

No. 21
Systematic review of the long-term effects and economic consequences of treatments for obesity and implications for health improvement.
By Avenell A, Broom J, Brown TJ, Poobalan A, Aucott L, Stearns SC, *et al*.

No. 22
Autoantibody testing in children with newly diagnosed type 1 diabetes mellitus.
By Dretzke J, Cummins C, Sandercock J, Fry-Smith A, Barrett T, Burls A.

No. 23
Clinical effectiveness and cost-effectiveness of prehospital intravenous fluids in trauma patients.
By Dretzke J, Sandercock J, Bayliss S, Burls A.

No. 24
Newer hypnotic drugs for the short-term management of insomnia: a systematic review and economic evaluation.
By Dündar Y, Boland A, Strobl J, Dodd S, Haycox A, Bagust A, *et al*.

No. 25
Development and validation of methods for assessing the quality of diagnostic accuracy studies.
By Whiting P, Rutjes AWS, Dinnes J, Reitsma JB, Bossuyt PMM, Kleijnen J.

No. 26
EVALUATE hysterectomy trial: a multicentre randomised trial comparing abdominal, vaginal and laparoscopic methods of hysterectomy.
By Garry R, Fountain J, Brown J, Manca A, Mason S, Sculpher M, *et al*.

No. 27
Methods for expected value of information analysis in complex health economic models: developments on the health economics of interferon-β and glatiramer acetate for multiple sclerosis.
By Tappenden P, Chilcott JB, Eggington S, Oakley J, McCabe C.

No. 28
Effectiveness and cost-effectiveness of imatinib for first-line treatment of chronic myeloid leukaemia in chronic phase: a systematic review and economic analysis.
By Dalziel K, Round A, Stein K, Garside R, Price A.

No. 29
VenUS I: a randomised controlled trial of two types of bandage for treating venous leg ulcers.
By Iglesias C, Nelson EA, Cullum NA, Torgerson DJ on behalf of the VenUS Team.

No. 30
Systematic review of the effectiveness and cost-effectiveness, and economic evaluation, of myocardial perfusion scintigraphy for the diagnosis and management of angina and myocardial infarction.
By Mowatt G, Vale L, Brazzelli M, Hernandez R, Murray A, Scott N, *et al*.

No. 31
A pilot study on the use of decision theory and value of information analysis as part of the NHS Health Technology Assessment programme.
By Claxton K, Ginnelly L, Sculpher M, Philips Z, Palmer S.

No. 32
The Social Support and Family Health Study: a randomised controlled trial and economic evaluation of two alternative forms of postnatal support for mothers living in disadvantaged inner-city areas.
By Wiggins M, Oakley A, Roberts I, Turner H, Rajan L, Austerberry H, *et al*.

No. 33
Psychosocial aspects of genetic screening of pregnant women and newborns: a systematic review.
By Green JM, Hewison J, Bekker HL, Bryant, Cuckle HS.

No. 34
Evaluation of abnormal uterine bleeding: comparison of three outpatient procedures within cohorts defined by age and menopausal status.
By Critchley HOD, Warner P, Lee AJ, Brechin S, Guise J, Graham B.

No. 35
Coronary artery stents: a rapid systematic review and economic evaluation.
By Hill R, Bagust A, Bakhai A, Dickson R, Dündar Y, Haycox A, *et al*.

No. 36
Review of guidelines for good practice in decision-analytic modelling in health technology assessment.
By Philips Z, Ginnelly L, Sculpher M, Claxton K, Golder S, Riemsma R, *et al*.

No. 37
Rituximab (MabThera®) for aggressive non-Hodgkin's lymphoma: systematic review and economic evaluation.
By Knight C, Hind D, Brewer N, Abbott V.

No. 38
Clinical effectiveness and cost-effectiveness of clopidogrel and modified-release dipyridamole in the secondary prevention of occlusive vascular events: a systematic review and economic evaluation.
By Jones L, Griffin S, Palmer S, Main C, Orton V, Sculpher M, *et al.*

No. 39
Pegylated interferon α-2a and -2b in combination with ribavirin in the treatment of chronic hepatitis C: a systematic review and economic evaluation.
By Shepherd J, Brodin H, Cave C, Waugh N, Price A, Gabbay J.

No. 40
Clopidogrel used in combination with aspirin compared with aspirin alone in the treatment of non-ST-segment-elevation acute coronary syndromes: a systematic review and economic evaluation.
By Main C, Palmer S, Griffin S, Jones L, Orton V, Sculpher M, *et al.*

No. 41
Provision, uptake and cost of cardiac rehabilitation programmes: improving services to under-represented groups.
By Beswick AD, Rees K, Griebsch I, Taylor FC, Burke M, West RR, *et al.*

No. 42
Involving South Asian patients in clinical trials.
By Hussain-Gambles M, Leese B, Atkin K, Brown J, Mason S, Tovey P.

No. 43
Clinical and cost-effectiveness of continuous subcutaneous insulin infusion for diabetes.
By Colquitt JL, Green C, Sidhu MK, Hartwell D, Waugh N.

No. 44
Identification and assessment of ongoing trials in health technology assessment reviews.
By Song FJ, Fry-Smith A, Davenport C, Bayliss S, Adi Y, Wilson JS, *et al*.

No. 45
Systematic review and economic evaluation of a long-acting insulin analogue, insulin glargine
By Warren E, Weatherley-Jones E, Chilcott J, Beverley C.

No. 46
Supplementation of a home-based exercise programme with a class-based programme for people with osteoarthritis of the knees: a randomised controlled trial and health economic analysis.
By McCarthy CJ, Mills PM, Pullen R, Richardson G, Hawkins N, Roberts CR, *et al*.

No. 47
Clinical and cost-effectiveness of once-daily versus more frequent use of same potency topical corticosteroids for atopic eczema: a systematic review and economic evaluation.
By Green C, Colquitt JL, Kirby J, Davidson P, Payne E.

No. 48
Acupuncture of chronic headache disorders in primary care: randomised controlled trial and economic analysis.
By Vickers AJ, Rees RW, Zollman CE, McCarney R, Smith CM, Ellis N, *et al*.

No. 49
Generalisability in economic evaluation studies in healthcare: a review and case studies.
By Sculpher MJ, Pang FS, Manca A, Drummond MF, Golder S, Urdahl H, *et al*.

No. 50
Virtual outreach: a randomised controlled trial and economic evaluation of joint teleconferenced medical consultations.
By Wallace P, Barber J, Clayton W, Currell R, Fleming K, Garner P, *et al*.

Volume 9, 2005

No. 1
Randomised controlled multiple treatment comparison to provide a cost-effectiveness rationale for the selection of antimicrobial therapy in acne.
By Ozolins M, Eady EA, Avery A, Cunliffe WJ, O'Neill C, Simpson NB, *et al*.

No. 2
Do the findings of case series studies vary significantly according to methodological characteristics?
By Dalziel K, Round A, Stein K, Garside R, Castelnuovo E, Payne L.

No. 3
Improving the referral process for familial breast cancer genetic counselling: findings of three randomised controlled trials of two interventions.
By Wilson BJ, Torrance N, Mollison J, Wordsworth S, Gray JR, Haites NE, *et al*.

No. 4
Randomised evaluation of alternative electrosurgical modalities to treat bladder outflow obstruction in men with benign prostatic hyperplasia.
By Fowler C, McAllister W, Plail R, Karim O, Yang Q.

No. 5
A pragmatic randomised controlled trial of the cost-effectiveness of palliative therapies for patients with inoperable oesophageal cancer.
By Shenfine J, McNamee P, Steen N, Bond J, Griffin SM.

No. 6
Impact of computer-aided detection prompts on the sensitivity and specificity of screening mammography.
By Taylor P, Champness J, Given-Wilson R, Johnston K, Potts H.

No. 7
Issues in data monitoring and interim analysis of trials.
By Grant AM, Altman DG, Babiker AB, Campbell MK, Clemens FJ, Darbyshire JH, *et al*.

No. 8
Lay public's understanding of equipoise and randomisation in randomised controlled trials.
By Robinson EJ, Kerr CEP, Stevens AJ, Lilford RJ, Braunholtz DA, Edwards SJ, *et al*.

No. 9
Clinical and cost-effectiveness of electroconvulsive therapy for depressive illness, schizophrenia, catatonia and mania: systematic reviews and economic modelling studies.
By Greenhalgh J, Knight C, Hind D, Beverley C, Walters S.

No. 10
Measurement of health-related quality of life for people with dementia: development of a new instrument (DEMQOL) and an evaluation of current methodology.
By Smith SC, Lamping DL, Banerjee S, Harwood R, Foley B, Smith P, *et al*.

No. 11
Clinical effectiveness and cost-effectiveness of drotrecogin alfa (activated) (Xigris®) for the treatment of severe sepsis in adults: a systematic review and economic evaluation.
By Green C, Dinnes J, Takeda A, Shepherd J, Hartwell D, Cave C, *et al*.

No. 12
A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy.
By Dinnes J, Deeks J, Kirby J, Roderick P.

No. 13
Cervical screening programmes: can automation help? Evidence from systematic reviews, an economic analysis and a simulation modelling exercise applied to the UK.
By Willis BH, Barton P, Pearmain P, Bryan S, Hyde C.

No. 14
Laparoscopic surgery for inguinal hernia repair: systematic review of effectiveness and economic evaluation.
By McCormack K, Wake B, Perez J, Fraser C, Cook J, McIntosh E, *et al*.

No. 15
Clinical effectiveness, tolerability and cost-effectiveness of newer drugs for epilepsy in adults: a systematic review and economic evaluation.
By Wilby J, Kainth A, Hawkins N, Epstein D, McIntosh H, McDaid C, *et al*.

No. 16
A randomised controlled trial to compare the cost-effectiveness of tricyclic antidepressants, selective serotonin reuptake inhibitors and lofepramine.
By Peveler R, Kendrick T, Buxton M, Longworth L, Baldwin D, Moore M, *et al*.

No. 17
Clinical effectiveness and cost-effectiveness of immediate angioplasty for acute myocardial infarction: systematic review and economic evaluation.
By Hartwell D, Colquitt J, Loveman E, Clegg AJ, Brodin H, Waugh N, *et al*.

No. 18
A randomised controlled comparison of alternative strategies in stroke care.
By Kalra L, Evans A, Perez I, Knapp M, Swift C, Donaldson N.

No. 19
The investigation and analysis of critical incidents and adverse events in healthcare.
By Woloshynowych M, Rogers S, Taylor-Adams S, Vincent C.

No. 20
Potential use of routine databases in health technology assessment.
By Raftery J, Roderick P, Stevens A.

No. 21
Clinical and cost-effectiveness of newer immunosuppressive regimens in renal transplantation: a systematic review and modelling study.
By Woodroffe R, Yao GL, Meads C, Bayliss S, Ready A, Raftery J, *et al*.

No. 22
A systematic review and economic evaluation of alendronate, etidronate, risedronate, raloxifene and teriparatide for the prevention and treatment of postmenopausal osteoporosis.
By Stevenson M, Lloyd Jones M, De Nigris E, Brewer N, Davis S, Oakley J.

No. 23
A systematic review to examine the impact of psycho-educational interventions on health outcomes and costs in adults and children with difficult asthma.
By Smith JR, Mugford M, Holland R, Candy B, Noble MJ, Harrison BDW, *et al*.

No. 24
An evaluation of the costs, effectiveness and quality of renal replacement therapy provision in renal satellite units in England and Wales.
By Roderick P, Nicholson T, Armitage A, Mehta R, Mullee M, Gerard K, *et al*.

No. 25
Imatinib for the treatment of patients with unresectable and/or metastatic gastrointestinal stromal tumours: systematic review and economic evaluation.
By Wilson J, Connock M, Song F, Yao G, Fry-Smith A, Raftery J, *et al*.

No. 26
Indirect comparisons of competing interventions.
By Glenny AM, Altman DG, Song F, Sakarovitch C, Deeks JJ, D'Amico R, *et al*.

No. 27
Cost-effectiveness of alternative strategies for the initial medical management of non-ST elevation acute coronary syndrome: systematic review and decision-analytical modelling.
By Robinson M, Palmer S, Sculpher M, Philips Z, Ginnelly L, Bowens A, *et al*.

No. 28
Outcomes of electrically stimulated gracilis neosphincter surgery.
By Tillin T, Chambers M, Feldman R.

No. 29
The effectiveness and cost-effectiveness of pimecrolimus and tacrolimus for atopic eczema: a systematic review and economic evaluation.
By Garside R, Stein K, Castelnuovo E, Pitt M, Ashcroft D, Dimmock P, *et al*.

No. 30
Systematic review on urine albumin testing for early detection of diabetic complications.
By Newman DJ, Mattock MB, Dawnay ABS, Kerry S, McGuire A, Yaqoob M, *et al*.

No. 31
Randomised controlled trial of the cost-effectiveness of water-based therapy for lower limb osteoarthritis.
By Cochrane T, Davey RC, Matthes Edwards SM.

No. 32
Longer term clinical and economic benefits of offering acupuncture care to patients with chronic low back pain.
By Thomas KJ, MacPherson H, Ratcliffe J, Thorpe L, Brazier J, Campbell M, *et al*.

No. 33
Cost-effectiveness and safety of epidural steroids in the management of sciatica.
By Price C, Arden N, Coglan L, Rogers P.

No. 34
The British Rheumatoid Outcome Study Group (BROSG) randomised controlled trial to compare the effectiveness and cost-effectiveness of aggressive versus symptomatic therapy in established rheumatoid arthritis.
By Symmons D, Tricker K, Roberts C, Davies L, Dawes P, Scott DL.

No. 35
Conceptual framework and systematic review of the effects of participants' and professionals' preferences in randomised controlled trials.
By King M, Nazareth I, Lampe F, Bower P, Chandler M, Morou M, *et al*.

No. 36
The clinical and cost-effectiveness of implantable cardioverter defibrillators: a systematic review.
By Bryant J, Brodin H, Loveman E, Payne E, Clegg A.

No. 37
A trial of problem-solving by community mental health nurses for anxiety, depression and life difficulties among general practice patients. The CPN-GP study.
By Kendrick T, Simons L, Mynors-Wallis L, Gray A, Lathlean J, Pickering R, *et al*.

No. 38
The causes and effects of socio-demographic exclusions from clinical trials.
By Bartlett C, Doyal L, Ebrahim S, Davey P, Bachmann M, Egger M, *et al*.

No. 39
Is hydrotherapy cost-effective? A randomised controlled trial of combined hydrotherapy programmes compared with physiotherapy land techniques in children with juvenile idiopathic arthritis.
By Epps H, Ginnelly L, Utley M, Southwood T, Gallivan S, Sculpher M, *et al*.

No. 40
A randomised controlled trial and cost-effectiveness study of systematic screening (targeted and total population screening) versus routine practice for the detection of atrial fibrillation in people aged 65 and over. The SAFE study.
By Hobbs FDR, Fitzmaurice DA, Mant J, Murray E, Jowett S, Bryan S, *et al*.

No. 41
Displaced intracapsular hip fractures in fit, older people: a randomised comparison of reduction and fixation, bipolar hemiarthroplasty and total hip arthroplasty.
By Keating JF, Grant A, Masson M, Scott NW, Forbes JF.

No. 42
Long-term outcome of cognitive behaviour therapy clinical trials in central Scotland.
By Durham RC, Chambers JA, Power KG, Sharp DM, Macdonald RR, Major KA, *et al*.

No. 43
The effectiveness and cost-effectiveness of dual-chamber pacemakers compared with single-chamber pacemakers for bradycardia due to atrioventricular block or sick sinus syndrome: systematic review and economic evaluation.
By Castelnuovo E, Stein K, Pitt M, Garside R, Payne E.

No. 44
Newborn screening for congenital heart defects: a systematic review and cost-effectiveness analysis.
By Knowles R, Griebsch I, Dezateux C, Brown J, Bull C, Wren C.

No. 45
The clinical and cost-effectiveness of left ventricular assist devices for end-stage heart failure: a systematic review and economic evaluation.
By Clegg AJ, Scott DA, Loveman E, Colquitt J, Hutchinson J, Royle P, *et al*.

No. 46
The effectiveness of the Heidelberg Retina Tomograph and laser diagnostic glaucoma scanning system (GDx) in detecting and monitoring glaucoma.
By Kwartz AJ, Henson DB, Harper RA, Spencer AF, McLeod D.

No. 47
Clinical and cost-effectiveness of autologous chondrocyte implantation for cartilage defects in knee joints: systematic review and economic evaluation.
By Clar C, Cummins E, McIntyre L, Thomas S, Lamb J, Bain L, *et al*.

No. 48
Systematic review of effectiveness of different treatments for childhood retinoblastoma.
By McDaid C, Hartley S, Bagnall A-M, Ritchie G, Light K, Riemsma R.

No. 49
Towards evidence-based guidelines for the prevention of venous thromboembolism: systematic reviews of mechanical methods, oral anticoagulation, dextran and regional anaesthesia as thromboprophylaxis.
By Roderick P, Ferris G, Wilson K, Halls H, Jackson D, Collins R, *et al*.

No. 50
The effectiveness and cost-effectiveness of parent training/education programmes for the treatment of conduct disorder, including oppositional defiant disorder, in children.
By Dretzke J, Frew E, Davenport C, Barlow J, Stewart-Brown S, Sandercock J, *et al*.

Volume 10, 2006

No. 1
The clinical and cost-effectiveness of donepezil, rivastigmine, galantamine and memantine for Alzheimer's disease.
By Loveman E, Green C, Kirby J, Takeda A, Picot J, Payne E, *et al*.

No. 2
FOOD: a multicentre randomised trial evaluating feeding policies in patients admitted to hospital with a recent stroke.
By Dennis M, Lewis S, Cranswick G, Forbes J.

No. 3
The clinical effectiveness and cost-effectiveness of computed tomography screening for lung cancer: systematic reviews.
By Black C, Bagust A, Boland A, Walker S, McLeod C, De Verteuil R, *et al*.

No. 4
A systematic review of the effectiveness and cost-effectiveness of neuroimaging assessments used to visualise the seizure focus in people with refractory epilepsy being considered for surgery.
By Whiting P, Gupta R, Burch J, Mujica Mota RE, Wright K, Marson A, *et al*.

No. 5
Comparison of conference abstracts and presentations with full-text articles in the health technology assessments of rapidly evolving technologies.
By Dundar Y, Dodd S, Dickson R, Walley T, Haycox A, Williamson PR.

No. 6
Systematic review and evaluation of methods of assessing urinary incontinence.
By Martin JL, Williams KS, Abrams KR, Turner DA, Sutton AJ, Chapple C, *et al*.

No. 7
The clinical effectiveness and cost-effectiveness of newer drugs for children with epilepsy. A systematic review.
By Connock M, Frew E, Evans B-W, Bryan S, Cummins C, Fry-Smith A, *et al*.

No. 8
Surveillance of Barrett's oesophagus: exploring the uncertainty through systematic review, expert workshop and economic modelling.
By Garside R, Pitt M, Somerville M, Stein K, Price A, Gilbert N.

No. 9
Topotecan, pegylated liposomal doxorubicin hydrochloride and paclitaxel for second-line or subsequent treatment of advanced ovarian cancer: a systematic review and economic evaluation.
By Main C, Bojke L, Griffin S, Norman G, Barbieri M, Mather L, *et al*.

No. 10
Evaluation of molecular techniques in prediction and diagnosis of cytomegalovirus disease in immunocompromised patients.
By Szczepura A, Westmoreland D, Vinogradova Y, Fox J, Clark M.

No. 11
Screening for thrombophilia in high-risk situations: systematic review and cost-effectiveness analysis. The Thrombosis: Risk and Economic Assessment of Thrombophilia Screening (TREATS) study.
By Wu O, Robertson L, Twaddle S, Lowe GDO, Clark P, Greaves M, *et al*.

No. 12
A series of systematic reviews to inform a decision analysis for sampling and treating infected diabetic foot ulcers.
By Nelson EA, O'Meara S, Craig D, Iglesias C, Golder S, Dalton J, *et al*.

No. 13
Randomised clinical trial, observational study and assessment of cost-effectiveness of the treatment of varicose veins (REACTIV trial).
By Michaels JA, Campbell WB, Brazier JE, MacIntyre JB, Palfreyman SJ, Ratcliffe J, *et al*.

No. 14
The cost-effectiveness of screening for oral cancer in primary care.
By Speight PM, Palmer S, Moles DR, Downer MC, Smith DH, Henriksson M *et al*.

No. 15
Measurement of the clinical and cost-effectiveness of non-invasive diagnostic testing strategies for deep vein thrombosis.
By Goodacre S, Sampson F, Stevenson M, Wailoo A, Sutton A, Thomas S, *et al*.

No. 16
Systematic review of the effectiveness and cost-effectiveness of HealOzone® for the treatment of occlusal pit/fissure caries and root caries.
By Brazzelli M, McKenzie L, Fielding S, Fraser C, Clarkson J, Kilonzo M, *et al*.

No. 17
Randomised controlled trials of conventional antipsychotic versus new atypical drugs, and new atypical drugs versus clozapine, in people with schizophrenia responding poorly to, or intolerant of, current drug treatment.
By Lewis SW, Davies L, Jones PB, Barnes TRE, Murray RM, Kerwin R, *et al*.

No. 18
Diagnostic tests and algorithms used in the investigation of haematuria: systematic reviews and economic evaluation.
By Rodgers M, Nixon J, Hempel S, Aho T, Kelly J, Neal D, *et al*.

No. 19
Cognitive behavioural therapy in addition to antispasmodic therapy for irritable bowel syndrome in primary care: randomised controlled trial.
By Kennedy TM, Chalder T, McCrone P, Darnley S, Knapp M, Jones RH, *et al*.

No. 20
A systematic review of the clinical effectiveness and cost-effectiveness of enzyme replacement therapies for Fabry's disease and mucopolysaccharidosis type 1.
By Connock M, Juarez-Garcia A, Frew E, Mans A, Dretzke J, Fry-Smith A, *et al*.

No. 21
Health benefits of antiviral therapy for mild chronic hepatitis C: randomised controlled trial and economic evaluation.
By Wright M, Grieve R, Roberts J, Main J, Thomas HC on behalf of the UK Mild Hepatitis C Trial Investigators.

No. 22
Pressure relieving support surfaces: a randomised evaluation.
By Nixon J, Nelson EA, Cranny G, Iglesias CP, Hawkins K, Cullum NA, *et al*.

No. 23
A systematic review and economic model of the effectiveness and cost-effectiveness of methylphenidate, dexamfetamine and atomoxetine for the treatment of attention deficit hyperactivity disorder in children and adolescents.
By King S, Griffin S, Hodges Z, Weatherly H, Asseburg C, Richardson G, *et al*.

No. 24
The clinical effectiveness and cost-effectiveness of enzyme replacement therapy for Gaucher's disease: a systematic review.
By Connock M, Burls A, Frew E, Fry-Smith A, Juarez-Garcia A, McCabe C, *et al*.

No. 25
Effectiveness and cost-effectiveness of salicylic acid and cryotherapy for cutaneous warts. An economic decision model.
By Thomas KS, Keogh-Brown MR, Chalmers JR, Fordham RJ, Holland RC, Armstrong SJ, *et al*.

No. 26
A systematic literature review of the effectiveness of non-pharmacological interventions to prevent wandering in dementia and evaluation of the ethical implications and acceptability of their use.
By Robinson L, Hutchings D, Corner L, Beyer F, Dickinson H, Vanoli A, *et al*.

No. 27
A review of the evidence on the effects and costs of implantable cardioverter defibrillator therapy in different patient groups, and modelling of cost-effectiveness and cost–utility for these groups in a UK context.
By Buxton M, Caine N, Chase D, Connelly D, Grace A, Jackson C, *et al*.

No. 28
Adefovir dipivoxil and pegylated interferon alfa-2a for the treatment of chronic hepatitis B: a systematic review and economic evaluation.
By Shepherd J, Jones J, Takeda A, Davidson P, Price A.

No. 29
An evaluation of the clinical and cost-effectiveness of pulmonary artery catheters in patient management in intensive care: a systematic review and a randomised controlled trial.
By Harvey S, Stevens K, Harrison D, Young D, Brampton W, McCabe C, *et al*.

No. 30
Accurate, practical and cost-effective assessment of carotid stenosis in the UK.
By Wardlaw JM, Chappell FM, Stevenson M, De Nigris E, Thomas S, Gillard J, *et al*.

No. 31
Etanercept and infliximab for the treatment of psoriatic arthritis: a systematic review and economic evaluation.
By Woolacott N, Bravo Vergel Y, Hawkins N, Kainth A, Khadjesari Z, Misso K, *et al*.

No. 32
The cost-effectiveness of testing for hepatitis C in former injecting drug users.
By Castelnuovo E, Thompson-Coon J, Pitt M, Cramp M, Siebert U, Price A, *et al*.

No. 33
Computerised cognitive behaviour therapy for depression and anxiety update: a systematic review and economic evaluation.
By Kaltenthaler E, Brazier J, De Nigris E, Tumur I, Ferriter M, Beverley C, *et al*.

No. 34
Cost-effectiveness of using prognostic information to select women with breast cancer for adjuvant systemic therapy.
By Williams C, Brunskill S, Altman D, Briggs A, Campbell H, Clarke M, *et al*.

No. 35
Psychological therapies including dialectical behaviour therapy for borderline personality disorder: a systematic review and preliminary economic evaluation.
By Brazier J, Tumur I, Holmes M, Ferriter M, Parry G, Dent-Brown K, *et al*.

No. 36
Clinical effectiveness and cost-effectiveness of tests for the diagnosis and investigation of urinary tract infection in children: a systematic review and economic model.
By Whiting P, Westwood M, Bojke L, Palmer S, Richardson G, Cooper J, *et al*.

No. 37
Cognitive behavioural therapy in chronic fatigue syndrome: a randomised controlled trial of an outpatient group programme.
By O'Dowd H, Gladwell P, Rogers CA, Hollinghurst S, Gregory A.

No. 38
A comparison of the cost-effectiveness of five strategies for the prevention of non-steroidal anti-inflammatory drug-induced gastrointestinal toxicity: a systematic review with economic modelling.
By Brown TJ, Hooper L, Elliott RA, Payne K, Webb R, Roberts C, *et al*.

No. 39
The effectiveness and cost-effectiveness of computed tomography screening for coronary artery disease: systematic review.
By Waugh N, Black C, Walker S, McIntyre L, Cummins E, Hillis G.

No. 40
What are the clinical outcome and cost-effectiveness of endoscopy undertaken by nurses when compared with doctors? A Multi-Institution Nurse Endoscopy Trial (MINuET).
By Williams J, Russell I, Durai D, Cheung W-Y, Farrin A, Bloor K, *et al*.

No. 41
The clinical and cost-effectiveness of oxaliplatin and capecitabine for the adjuvant treatment of colon cancer: systematic review and economic evaluation.
By Pandor A, Eggington S, Paisley S, Tappenden P, Sutcliffe P.

No. 42
A systematic review of the effectiveness of adalimumab, etanercept and infliximab for the treatment of rheumatoid arthritis in adults and an economic evaluation of their cost-effectiveness.
By Chen Y-F, Jobanputra P, Barton P, Jowett S, Bryan S, Clark W, *et al*.

No. 43
Telemedicine in dermatology: a randomised controlled trial.
By Bowns IR, Collins K, Walters SJ, McDonagh AJG.

No. 44
Cost-effectiveness of cell salvage and alternative methods of minimising perioperative allogeneic blood transfusion: a systematic review and economic model.
By Davies L, Brown TJ, Haynes S, Payne K, Elliott RA, McCollum C.

No. 45
Clinical effectiveness and cost-effectiveness of laparoscopic surgery for colorectal cancer: systematic reviews and economic evaluation.
By Murray A, Lourenco T, de Verteuil R, Hernandez R, Fraser C, McKinley A, *et al*.

No. 46
Etanercept and efalizumab for the treatment of psoriasis: a systematic review.
By Woolacott N, Hawkins N, Mason A, Kainth A, Khadjesari Z, Bravo Vergel Y, *et al*.

No. 47
Systematic reviews of clinical decision tools for acute abdominal pain.
By Liu JLY, Wyatt JC, Deeks JJ, Clamp S, Keen J, Verde P, *et al*.

No. 48
Evaluation of the ventricular assist device programme in the UK.
By Sharples L, Buxton M, Caine N, Cafferty F, Demiris N, Dyer M, *et al*.

No. 49
A systematic review and economic model of the clinical and cost-effectiveness of immunosuppressive therapy for renal transplantation in children.
By Yao G, Albon E, Adi Y, Milford D, Bayliss S, Ready A, *et al*.

No. 50
Amniocentesis results: investigation of anxiety. The ARIA trial.
By Hewison J, Nixon J, Fountain J, Cocks K, Jones C, Mason G, *et al*.

Volume 11, 2007

No. 1
Pemetrexed disodium for the treatment of malignant pleural mesothelioma: a systematic review and economic evaluation.
By Dundar Y, Bagust A, Dickson R, Dodd S, Green J, Haycox A, *et al*.

No. 2
A systematic review and economic model of the clinical effectiveness and cost-effectiveness of docetaxel in combination with prednisone or prednisolone for the treatment of hormone-refractory metastatic prostate cancer.
By Collins R, Fenwick E, Trowman R, Perard R, Norman G, Light K, *et al*.

No. 3
A systematic review of rapid diagnostic tests for the detection of tuberculosis infection.
By Dinnes J, Deeks J, Kunst H, Gibson A, Cummins E, Waugh N, *et al*.

No. 4
The clinical effectiveness and cost-effectiveness of strontium ranelate for the prevention of osteoporotic fragility fractures in postmenopausal women.
By Stevenson M, Davis S, Lloyd-Jones M, Beverley C.

No. 5
A systematic review of quantitative and qualitative research on the role and effectiveness of written information available to patients about individual medicines.
By Raynor DK, Blenkinsopp A, Knapp P, Grime J, Nicolson DJ, Pollock K, *et al*.

No. 6
Oral naltrexone as a treatment for relapse prevention in formerly opioid-dependent drug users: a systematic review and economic evaluation.
By Adi Y, Juarez-Garcia A, Wang D, Jowett S, Frew E, Day E, *et al*.

No. 7
Glucocorticoid-induced osteoporosis: a systematic review and cost–utility analysis.
By Kanis JA, Stevenson M, McCloskey EV, Davis S, Lloyd-Jones M.

No. 8
Epidemiological, social, diagnostic and economic evaluation of population screening for genital chlamydial infection.
By Low N, McCarthy A, Macleod J, Salisbury C, Campbell R, Roberts TE, *et al*.

No. 9
Methadone and buprenorphine for the management of opioid dependence: a systematic review and economic evaluation.
By Connock M, Juarez-Garcia A, Jowett S, Frew E, Liu Z, Taylor RJ, *et al*.

No. 10
Exercise Evaluation Randomised Trial (EXERT): a randomised trial comparing GP referral for leisure centre-based exercise, community-based walking and advice only.
By Isaacs AJ, Critchley JA, See Tai S, Buckingham K, Westley D, Harridge SDR, *et al*.

No. 11
Interferon alfa (pegylated and non-pegylated) and ribavirin for the treatment of mild chronic hepatitis C: a systematic review and economic evaluation.
By Shepherd J, Jones J, Hartwell D, Davidson P, Price A, Waugh N.

No. 12
Systematic review and economic evaluation of bevacizumab and cetuximab for the treatment of metastatic colorectal cancer.
By Tappenden P, Jones R, Paisley S, Carroll C.

No. 13
A systematic review and economic evaluation of epoetin alfa, epoetin beta and darbepoetin alfa in anaemia associated with cancer, especially that attributable to cancer treatment.
By Wilson J, Yao GL, Raftery J, Bohlius J, Brunskill S, Sandercock J, *et al*.

No. 14
A systematic review and economic evaluation of statins for the prevention of coronary events.
By Ward S, Lloyd Jones M, Pandor A, Holmes M, Ara R, Ryan A, *et al*.

No. 15
A systematic review of the effectiveness and cost-effectiveness of different models of community-based respite care for frail older people and their carers.
By Mason A, Weatherly H, Spilsbury K, Arksey H, Golder S, Adamson J, *et al*.

No. 16
Additional therapy for young children with spastic cerebral palsy: a randomised controlled trial.
By Weindling AM, Cunningham CC, Glenn SM, Edwards RT, Reeves DJ.

No. 17
Screening for type 2 diabetes: literature review and economic modelling.
By Waugh N, Scotland G, McNamee P, Gillett M, Brennan A, Goyder E, *et al*.

No. 18
The effectiveness and cost-effectiveness of cinacalcet for secondary hyperparathyroidism in end-stage renal disease patients on dialysis: a systematic review and economic evaluation.
By Garside R, Pitt M, Anderson R, Mealing S, Roome C, Snaith A, *et al*.

No. 19
The clinical effectiveness and cost-effectiveness of gemcitabine for metastatic breast cancer: a systematic review and economic evaluation.
By Takeda AL, Jones J, Loveman E, Tan SC, Clegg AJ.

No. 20
A systematic review of duplex ultrasound, magnetic resonance angiography and computed tomography angiography for the diagnosis and assessment of symptomatic, lower limb peripheral arterial disease.
By Collins R, Cranny G, Burch J, Aguiar-Ibáñez R, Craig D, Wright K, *et al*.

No. 21
The clinical effectiveness and cost-effectiveness of treatments for children with idiopathic steroid-resistant nephrotic syndrome: a systematic review.
By Colquitt JL, Kirby J, Green C, Cooper K, Trompeter RS.

No. 22
A systematic review of the routine monitoring of growth in children of primary school age to identify growth-related conditions.
By Fayter D, Nixon J, Hartley S, Rithalia A, Butler G, Rudolf M, *et al*.

No. 23
Systematic review of the effectiveness of preventing and treating *Staphylococcus aureus* carriage in reducing peritoneal catheter-related infections.
By McCormack K, Rabindranath K, Kilonzo M, Vale L, Fraser C, McIntyre L, *et al*.

No. 24
The clinical effectiveness and cost of repetitive transcranial magnetic stimulation versus electroconvulsive therapy in severe depression: a multicentre pragmatic randomised controlled trial and economic analysis.
By McLoughlin DM, Mogg A, Eranti S, Pluck G, Purvis R, Edwards D, *et al*.

No. 25
A randomised controlled trial and economic evaluation of direct versus indirect and individual versus group modes of speech and language therapy for children with primary language impairment.
By Boyle J, McCartney E, Forbes J, O'Hare A.

No. 26
Hormonal therapies for early breast cancer: systematic review and economic evaluation.
By Hind D, Ward S, De Nigris E, Simpson E, Carroll C, Wyld L.

No. 27
Cardioprotection against the toxic effects of anthracyclines given to children with cancer: a systematic review.
By Bryant J, Picot J, Levitt G, Sullivan I, Baxter L, Clegg A.

No. 28
Adalimumab, etanercept and infliximab for the treatment of ankylosing spondylitis: a systematic review and economic evaluation.
By McLeod C, Bagust A, Boland A, Dagenais P, Dickson R, Dundar Y, *et al*.

No. 29
Prenatal screening and treatment strategies to prevent group B streptococcal and other bacterial infections in early infancy: cost-effectiveness and expected value of information analyses.
By Colbourn T, Asseburg C, Bojke L, Philips Z, Claxton K, Ades AE, *et al*.

No. 30
Clinical effectiveness and cost-effectiveness of bone morphogenetic proteins in the non-healing of fractures and spinal fusion: a systematic review.
By Garrison KR, Donell S, Ryder J, Shemilt I, Mugford M, Harvey I, *et al*.

No. 31
A randomised controlled trial of postoperative radiotherapy following breast-conserving surgery in a minimum-risk older population. The PRIME trial.
By Prescott RJ, Kunkler IH, Williams LJ, King CC, Jack W, van der Pol M, *et al*.

No. 32
Current practice, accuracy, effectiveness and cost-effectiveness of the school entry hearing screen.
By Bamford J, Fortnum H, Bristow K, Smith J, Vamvakas G, Davies L, *et al*.

No. 33
The clinical effectiveness and cost-effectiveness of inhaled insulin in diabetes mellitus: a systematic review and economic evaluation.
By Black C, Cummins E, Royle P, Philip S, Waugh N.

No. 34
Surveillance of cirrhosis for hepatocellular carcinoma: systematic review and economic analysis.
By Thompson Coon J, Rogers G, Hewson P, Wright D, Anderson R, Cramp M, *et al*.

No. 35
The Birmingham Rehabilitation Uptake Maximisation Study (BRUM). Home-based compared with hospital-based cardiac rehabilitation in a multi-ethnic population: cost-effectiveness and patient adherence.
By Jolly K, Taylor R, Lip GYH, Greenfield S, Raftery J, Mant J, *et al*.

No. 36
A systematic review of the clinical, public health and cost-effectiveness of rapid diagnostic tests for the detection and identification of bacterial intestinal pathogens in faeces and food.
By Abubakar I, Irvine L, Aldus CF, Wyatt GM, Fordham R, Schelenz S, *et al*.

No. 37
A randomised controlled trial examining the longer-term outcomes of standard versus new antiepileptic drugs. The SANAD trial.
By Marson AG, Appleton R, Baker GA, Chadwick DW, Doughty J, Eaton B, *et al*.

No. 38
Clinical effectiveness and cost-effectiveness of different models of managing long-term oral anticoagulation therapy: a systematic review and economic modelling.
By Connock M, Stevens C, Fry-Smith A, Jowett S, Fitzmaurice D, Moore D, *et al*.

No. 39
A systematic review and economic model of the clinical effectiveness and cost-effectiveness of interventions for preventing relapse in people with bipolar disorder.
By Soares-Weiser K, Bravo Vergel Y, Beynon S, Dunn G, Barbieri M, Duffy S, *et al*.

No. 40
Taxanes for the adjuvant treatment of early breast cancer: systematic review and economic evaluation.
By Ward S, Simpson E, Davis S, Hind D, Rees A, Wilkinson A.

No. 41
The clinical effectiveness and cost-effectiveness of screening for open angle glaucoma: a systematic review and economic evaluation.
By Burr JM, Mowatt G, Hernández R, Siddiqui MAR, Cook J, Lourenco T, *et al*.

No. 42
Acceptability, benefit and costs of early screening for hearing disability: a study of potential screening tests and models.
By Davis A, Smith P, Ferguson M, Stephens D, Gianopoulos I.

No. 43
Contamination in trials of educational interventions.
By Keogh-Brown MR, Bachmann MO, Shepstone L, Hewitt C, Howe A, Ramsay CR, *et al*.

Health Technology Assessment Programme

Prioritisation Strategy Group

Members

HTA Commissioning Board

Members

Current and past membership details of all HTA 'committees' are available from the HTA website (www.hta.ac.uk)

Diagnostic Technologies & Screening Panel

Members

Chair,
Dr Ron Zimmern, Director of the Public Health Genetics Unit, Strangeways Research Laboratories, Cambridge

Ms Norma Armston, Freelance Consumer Advocate, Bolton

Professor Max Bachmann, Professor of Health Care Interfaces, Department of Health Policy and Practice, University of East Anglia

Professor Rudy Bilous Professor of Clinical Medicine & Consultant Physician, The Academic Centre, South Tees Hospitals NHS Trust

Ms Dea Birkett, Service User Representative, London

Dr Paul Cockcroft, Consultant Medical Microbiologist and Clinical Director of Pathology, Department of Clinical Microbiology, St Mary's Hospital, Portsmouth

Professor Adrian K Dixon, Professor of Radiology, University Department of Radiology, University of Cambridge Clinical School

Dr David Elliman, Consultant in Community Child Health, Islington PCT & Great Ormond Street Hospital, London

Professor Glyn Elwyn, Research Chair, Centre for Health Sciences Research, Cardiff University, Department of General Practice, Cardiff

Professor Paul Glasziou, Director, Centre for Evidence-Based Practice, University of Oxford

Dr Jennifer J Kurinczuk, Consultant Clinical Epidemiologist, National Perinatal Epidemiology Unit, Oxford

Dr Susanne M Ludgate, Clinical Director, Medicines & Healthcare Products Regulatory Agency, London

Mr Stephen Pilling, Director, Centre for Outcomes, Research & Effectiveness, Joint Director, National Collaborating Centre for Mental Health, University College London

Mrs Una Rennard, Service User Representative, Oxford

Dr Phil Shackley, Senior Lecturer in Health Economics, Academic Vascular Unit, University of Sheffield

Dr Margaret Somerville, Director of Public Health Learning, Peninsula Medical School, University of Plymouth

Dr Graham Taylor, Scientific Director & Senior Lecturer, Regional DNA Laboratory, The Leeds Teaching Hospitals

Professor Lindsay Wilson Turnbull, Scientific Director, Centre for MR Investigations & YCR Professor of Radiology, University of Hull

Professor Martin J Whittle, Clinical Co-director, National Co-ordinating Centre for Women's and Childhealth

Dr Dennis Wright, Consultant Biochemist & Clinical Director, The North West London Hospitals NHS Trust, Middlesex

Pharmaceuticals Panel

Members

Chair,
Professor Robin Ferner, Consultant Physician and Director, West Midlands Centre for Adverse Drug Reactions, City Hospital NHS Trust, Birmingham

Ms Anne Baileff, Consultant Nurse in First Contact Care, Southampton City Primary Care Trust, University of Southampton

Professor Imti Choonara, Professor in Child Health, Academic Division of Child Health, University of Nottingham

Professor John Geddes, Professor of Epidemiological Psychiatry, University of Oxford

Mrs Barbara Greggains, Non-Executive Director, Greggains Management Ltd

Dr Bill Gutteridge, Medical Adviser, National Specialist Commissioning Advisory Group (NSCAG), London

Mrs Sharon Hart, Consultant Pharmaceutical Adviser, Reading

Dr Jonathan Karnon, Senior Research Fellow, Health Economics and Decision Science, University of Sheffield

Dr Yoon Loke, Senior Lecturer in Clinical Pharmacology, University of East Anglia

Ms Barbara Meredith, Lay Member, Epsom

Dr Andrew Prentice, Senior Lecturer and Consultant Obstetrician & Gynaecologist, Department of Obstetrics & Gynaecology, University of Cambridge

Dr Frances Rotblat, CPMP Delegate, Medicines & Healthcare Products Regulatory Agency, London

Dr Martin Shelly, General Practitioner, Leeds

Mrs Katrina Simister, Assistant Director New Medicines, National Prescribing Centre, Liverpool

Dr Richard Tiner, Medical Director, Medical Department, Association of the British Pharmaceutical Industry, London

Therapeutic Procedures Panel

Members

Chair,
Professor Bruce Campbell, Consultant Vascular and General Surgeon, Department of Surgery, Royal Devon & Exeter Hospital

Dr Mahmood Adil, Deputy Regional Director of Public Health, Department of Health, Manchester

Dr Aileen Clarke, Consultant in Public Health, Public Health Resource Unit, Oxford

Professor Matthew Cooke, Professor of Emergency Medicine, Warwick Emergency Care and Rehabilitation, University of Warwick

Mr Mark Emberton, Senior Lecturer in Oncological Urology, Institute of Urology, University College Hospital

Professor Paul Gregg, Professor of Orthopaedic Surgical Science, Department of General Practice and Primary Care, South Tees Hospital NHS Trust, Middlesbrough

Ms Maryann L Hardy, Lecturer, Division of Radiography, University of Bradford

Dr Simon de Lusignan, Senior Lecturer, Primary Care Informatics, Department of Community Health Sciences, St George's Hospital Medical School, London

Dr Peter Martin, Consultant Neurologist, Addenbrooke's Hospital, Cambridge

Professor Neil McIntosh, Edward Clark Professor of Child Life & Health, Department of Child Life & Health, University of Edinburgh

Professor Jim Neilson, Professor of Obstetrics and Gynaecology, Department of Obstetrics and Gynaecology, University of Liverpool

Dr John C Pounsford, Consultant Physician, Directorate of Medical Services, North Bristol NHS Trust

Dr Karen Roberts, Nurse Consultant, Queen Elizabeth Hospital, Gateshead

Dr Vimal Sharma, Consultant Psychiatrist/Hon. Senior Lecturer, Mental Health Resource Centre, Cheshire and Wirral Partnership NHS Trust, Wallasey

Professor Scott Weich, Professor of Psychiatry, Division of Health in the Community, University of Warwick

Disease Prevention Panel

Members

Chair,
Dr Edmund Jessop, Medical Adviser, National Specialist Commissioning Advisory Group (NSCAG), London

Mrs Sheila Clark, Chief Executive, St James's Hospital, Portsmouth

Mr Richard Copeland, Lead Pharmacist: Clinical Economy/Interface, Wansbeck General Hospital, Northumberland

Dr Elizabeth Fellow-Smith, Medical Director, West London Mental Health Trust, Middlesex

Mr Ian Flack, Director PPI Forum Support, Council of Ethnic Minority Voluntary Sector Organisations, Stratford

Dr John Jackson, General Practitioner, Newcastle upon Tyne

Mrs Veronica James, Chief Officer, Horsham District Age Concern, Horsham

Professor Mike Kelly, Director, Centre for Public Health Excellence, National Institute for Health and Clinical Excellence, London

Professor Yi Mien Koh, Director of Public Health and Medical Director, London NHS (North West London Strategic Health Authority), London

Ms Jeanett Martin, Director of Clinical Leadership & Quality, Lewisham PCT, London

Dr Chris McCall, General Practitioner, Dorset

Dr David Pencheon, Director, Eastern Region Public Health Observatory, Cambridge

Dr Ken Stein, Senior Clinical Lecturer in Public Health, Director, Peninsula Technology Assessment Group, University of Exeter, Exeter

Dr Carol Tannahill, Director, Glasgow Centre for Population Health, Glasgow

Professor Margaret Thorogood, Professor of Epidemiology, University of Warwick, Coventry

Dr Ewan Wilkinson, Consultant in Public Health, Royal Liverpool University Hospital, Liverpool

Expert Advisory Network

Members

Professor Douglas Altman, Professor of Statistics in Medicine, Centre for Statistics in Medicine, University of Oxford

Professor John Bond, Director, Centre for Health Services Research, University of Newcastle upon Tyne, School of Population & Health Sciences, Newcastle upon Tyne

Professor Andrew Bradbury, Professor of Vascular Surgery, Solihull Hospital, Birmingham

Mr Shaun Brogan, Chief Executive, Ridgeway Primary Care Group, Aylesbury

Mrs Stella Burnside OBE, Chief Executive, Regulation and Improvement Authority, Belfast

Ms Tracy Bury, Project Manager, World Confederation for Physical Therapy, London

Professor Iain T Cameron, Professor of Obstetrics and Gynaecology and Head of the School of Medicine, University of Southampton

Dr Christine Clark, Medical Writer & Consultant Pharmacist, Rossendale

Professor Collette Clifford, Professor of Nursing & Head of Research, School of Health Sciences, University of Birmingham, Edgbaston, Birmingham

Professor Barry Cookson, Director, Laboratory of Healthcare Associated Infection, Health Protection Agency, London

Dr Carl Counsell, Clinical Senior Lecturer in Neurology, Department of Medicine & Therapeutics, University of Aberdeen

Professor Howard Cuckle, Professor of Reproductive Epidemiology, Department of Paediatrics, Obstetrics & Gynaecology, University of Leeds

Dr Katherine Darton, Information Unit, MIND – The Mental Health Charity, London

Professor Carol Dezateux, Professor of Paediatric Epidemiology, London

Dr Keith Dodd, Consultant Paediatrician, Derby

Mr John Dunning, Consultant Cardiothoracic Surgeon, Cardiothoracic Surgical Unit, Papworth Hospital NHS Trust, Cambridge

Mr Jonothan Earnshaw, Consultant Vascular Surgeon, Gloucestershire Royal Hospital, Gloucester

Professor Martin Eccles, Professor of Clinical Effectiveness, Centre for Health Services Research, University of Newcastle upon Tyne

Professor Pam Enderby, Professor of Community Rehabilitation, Institute of General Practice and Primary Care, University of Sheffield

Professor Gene Feder, Professor of Primary Care Research & Development, Centre for Health Sciences, Barts & The London Queen Mary's School of Medicine & Dentistry, London

Mr Leonard R Fenwick, Chief Executive, Newcastle upon Tyne Hospitals NHS Trust

Mrs Gillian Fletcher, Antenatal Teacher & Tutor and President, National Childbirth Trust, Henfield

Professor Jayne Franklyn, Professor of Medicine, Department of Medicine, University of Birmingham, Queen Elizabeth Hospital, Edgbaston, Birmingham

Dr Neville Goodman, Consultant Anaesthetist, Southmead Hospital, Bristol

Professor Robert E Hawkins, CRC Professor and Director of Medical Oncology, Christie CRC Research Centre, Christie Hospital NHS Trust, Manchester

Professor Allen Hutchinson, Director of Public Health & Deputy Dean of ScHARR, Department of Public Health, University of Sheffield

Professor Peter Jones, Professor of Psychiatry, University of Cambridge, Cambridge

Professor Stan Kaye, Cancer Research UK Professor of Medical Oncology, Section of Medicine, Royal Marsden Hospital & Institute of Cancer Research, Surrey

Dr Duncan Keeley, General Practitioner (Dr Burch & Ptnrs), The Health Centre, Thame

Dr Donna Lamping, Research Degrees Programme Director & Reader in Psychology, Health Services Research Unit, London School of Hygiene and Tropical Medicine, London

Mr George Levvy, Chief Executive, Motor Neurone Disease Association, Northampton

Professor James Lindesay, Professor of Psychiatry for the Elderly, University of Leicester, Leicester General Hospital

Professor Julian Little, Professor of Human Genome Epidemiology, Department of Epidemiology & Community Medicine, University of Ottawa

Professor Rajan Madhok, Consultant in Public Health, South Manchester Primary Care Trust, Manchester

Professor Alexander Markham, Director, Molecular Medicine Unit, St James's University Hospital, Leeds

Professor Alistaire McGuire, Professor of Health Economics, London School of Economics

Dr Peter Moore, Freelance Science Writer, Ashtead

Dr Andrew Mortimore, Public Health Director, Southampton City Primary Care Trust, Southampton

Dr Sue Moss, Associate Director, Cancer Screening Evaluation Unit, Institute of Cancer Research, Sutton

Mrs Julietta Patnick, Director, NHS Cancer Screening Programmes, Sheffield

Professor Robert Peveler, Professor of Liaison Psychiatry, Royal South Hants Hospital, Southampton

Professor Chris Price, Visiting Professor in Clinical Biochemistry, University of Oxford

Professor William Rosenberg, Professor of Hepatology and Consultant Physician, University of Southampton, Southampton

Professor Peter Sandercock, Professor of Medical Neurology, Department of Clinical Neurosciences, University of Edinburgh

Dr Susan Schonfield, Consultant in Public Health, Hillingdon PCT, Middlesex

Dr Eamonn Sheridan, Consultant in Clinical Genetics, Genetics Department, St James's University Hospital, Leeds

Professor Sarah Stewart-Brown, Professor of Public Health, University of Warwick, Division of Health in the Community Warwick Medical School, LWMS, Coventry

Professor Ala Szczepura, Professor of Health Service Research, Centre for Health Services Studies, University of Warwick

Dr Ross Taylor, Senior Lecturer, Department of General Practice and Primary Care, University of Aberdeen

Mrs Joan Webster, Consumer member, HTA – Expert Advisory Network

Feedback

The HTA Programme and the authors would like to know your views about this report.

The Correspondence Page on the HTA website (http://www.hta.ac.uk) is a convenient way to publish your comments. If you prefer, you can send your comments to the address below, telling us whether you would like us to transfer them to the website.

We look forward to hearing from you.

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@hta.ac.uk
http://www.hta.ac.uk

ISSN 1366-5278