Contamination in trials: is cluster randomisation the answer?

David J Torgerson

Centre for Health Economics, University of York, York YO10 5DD
David J Torgerson
senior research fellow
djt6@york.ac.uk

BMJ 2001;322:355–7

Most randomised trials allocate individual participants to different treatments. However, cluster randomised trials in which groups of subjects are allocated to different treatments are becoming increasingly popular.[1] Cluster randomisation is often advocated to minimise treatment "contamination" between intervention and control participants. For example, in a trial of dietary change, people in the control group might learn about the experimental diet and adopt it themselves.

Contamination of control participants has two related effects. It reduces the point estimate of an intervention's effectiveness and this apparent reduction may lead to a type II error—that is, rejection of an effective intervention as ineffective because the observed effect size was neither statistically nor clinically significant.

Although the threat of contamination is an issue in some controlled trials, it may be not be of much practical importance in many. Trialists should use individual randomisation if possible because of the drawbacks of cluster allocation. Cluster trials are associated with problems of recruitment bias and the need for larger samples than would be required in similar, individually randomised trials. In recruitment bias, different sorts of participants are selected into the various arms of the trial, thereby defeating the objective of randomisation, while a larger sample size may increase the cost of a trial, its length, or its complexity. This paper describes the difficulties of cluster trials and argues that the problem of contamination can often be dealt with by individual randomisation.

Summary points

Cluster trials are often used to prevent "contamination" between intervention and control groups

Cluster trials are usually very much larger than individually randomised trials and can be susceptible to recruitment bias

The problem of contamination can often be overcome by increasing the sample size

In terms of total sample size, cluster trials are only more efficient where contamination exceeds 30%

Sample size

Members of clusters cannot be treated as independent, and the effect of this on outcomes leads to a need to increase the sample size.[1 2] This problem can also be described as follows: for any given sample size, the correlation between cluster members will reduce the overall power of the study. The difficulty is well known and much has been published recently on sample size and analytical issues.[1–4]

Selection bias in cluster trials

The randomised trial is the ideal study design in evaluative research because of its ability to deal with known and unknown confounding factors. If randomisation is successful, these factors should be balanced across the treatment groups. Balance in cluster trials can be achieved through randomisation across the trial arms at the cluster level if there are sufficient clusters. However, simple randomisation of clusters, even with relatively large numbers, can still result in an imbalance. For instance, in a randomised trial of breast screening there was an imbalance in socioeconomic groups between the two study arms even though 87 clusters (general practices) and 50 000 women were included.[5] This imbalance would have been extremely unlikely had the trial used individual randomisation.

Cluster trials generally use the postrandomised consent method of trial design, which is similar in concept to Zelen's method in individually randomised designs.[6] Participants are not asked for their consent to randomisation; they are usually asked for consent to treatment (although this is not always done in cluster trials) and are asked whether they consent to inclusion in the study analysis. When cluster trials are analysed at the individual level using this method, balance between the randomised groups can only be guaranteed if one of two conditions are fulfilled: all members of the cluster must be included in the trial, or a random sample of participants must be included in the trial's analysis. The latter approach may be used to increase the statistical power of the study for a given sample size by increasing the number of clusters rather than the total number of patients.[4] For instance, Kerse et al took a random sample of patients from each cluster rather than all randomised individuals.[7] Neither of these conditions has been met in recently published cluster trials. Often participants are asked after randomisation whether they will participate in the trial in terms of providing data for follow up. If a significant proportion refuses, there is a possibility of selection bias.

Observable characteristics

An example of selection bias was seen in a recent trial of counselling to reduce the risk of cardiovascular disease.[8] Participants were allowed the choice of accepting the intervention and providing data on outcomes. Participation in the control arm was more desirable than participation in the intervention, as a result of which nearly twice as many subjects were recruited as controls. Moreover, the participants in the intervention group had some observable characteristics, such as a lower prevalence of smoking, which put them at a lower risk of coronary heart disease than the control group. Thus, the modest effect associated with the

bmj.com

A statistical appendix appears on the BMJ's website

Table 1 Numbers of participants required in an individually randomised trial to detect a difference between two groups where the true effect size is 25%

	Contamination (%)						
	0	5	10	15	20	25	30
Sample size	116	126	144	164	182	200	236
Contamination effect	1	1.09	1.24	1.41	1.58	1.72	2.03

intervention could have been due to the recruitment of different types of participant rather than the intervention. Similarly, in a cluster trial of diabetes care the intervention clusters identified about 25% more eligible patients than the control clusters, although the total size of the list of patients was slightly greater in the control clusters.[9]

Unobserved characteristics

Even when recruitment between the intervention and control clusters is similar and the observed baseline characteristics between allocation groups seem balanced, if not all the members of the cluster or a random sample of them are included in the analysis, the trial may still be unbalanced because of unobserved characteristics. For example, in a trial of a behavioural intervention to prevent violence in schools, the number of pupils in the two groups was similar, but only 66% of children—those whose parents had returned consent forms—were included in the study.[10] However, the intervention was still delivered to children whose parents had not returned consent forms. Although the trial recorded a creditable 82% follow up, this represented only 54% of the entire randomised sample. In contrast, a trial of accident prevention in children included all relevant members of the cluster in the analysis (children aged 3-12 months) and managed to collect and analyse outcome data on more than 92%, thus reassuring us that little, if any, selection bias had taken place.[11]

Reduced power: Zelen's method

When Zelen's method is used in an individually randomised trial there is usually a loss of statistical power because there is incomplete penetration of the intervention. Thus, uptake of a new treatment that is less than 100% will lead to a dilution of the effect of treatment and will reduce the measurable effect size.[12] Because cluster trials also use Zelen's method, this dilution of effect applies to them too. For example, in a cluster trial of injury prevention in young children the intervention was not delivered to about 25% of parents in the experimental arm.[11] This problem is not generally recognised in estimating the sample size for cluster trials, and if it is taken into account it will lead to a general increase in sample sizes. Thus, some cluster trials not only need to account for the clustering effects on statistical power, they also need to boost sample sizes because of the dilution effect.

Dealing with contamination

Before considering cluster randomisation as a way of addressing contamination we need to be certain that contamination is a real rather than a theoretical possibility. Are patients who are receiving counselling really going to pass on the intervention in such a way as to alter behaviour in the control arm of a trial? This thesis could be tested empirically by undertaking a pilot trial using individual randomisation with contamination of the control group as the outcome. However, even if contamination were a problem and a sizeable proportion of the control group was affected, individual randomisation might still be best.

As an example, consider a trial of counselling to promote healthy behaviour in adults at high risk of cardiovascular disease.[8] To detect a 9% reduction in smoking prevalence from 50% to 41% with 90% power at a significance value of 5%, a total of 1282 participants would have been needed. However, because the trialists elected to use a cluster design a sample of 2000 was required. The trialists could have chosen individual randomisation with a sample size of 2000 and utilised the extra power of this increased sample to cope with any contamination. Thus, a sample of 2000 would have been able to detect a 7% difference in smoking prevalence with 90% power and 5% significance, allowing for up to a 20% contamination of the control group. Furthermore, this estimate of the allowable contamination assumes that the effectiveness of the intervention through contamination is as powerful as if it had been delivered by healthcare professionals. However, if we assume that it was only half as effective, contamination could be much greater. Indeed, if contamination in this particular trial had been theoretical, the non-significant effect size noted in the trial would have been significant if individual randomisation had been used.

Effects of contamination on sample size

The probable scale of contamination will clearly differ with the intervention, and I am not aware of any published estimates in trials that might have used a cluster design to avoid contamination. However, a review of Zelen's method in cancer trials that did not use cluster randomisation showed that the mean proportion of patients who crossed from one treatment to another, which is analogous to contamination, was 18% (range 10%-36%).[12] The tables show the impact that different intracluster correlation coefficients and contamination rates have on estimates of sample size or contamination. Around 30% contamination can be sustained before the sample size has to be doubled to take into account the reduced effect size from such contamination. However, use of cluster randomisation rapidly leads to a doubling of the sample size. Even if the intracluster coefficient of correlation is very low, the clusters are small, and there are relatively high levels of contamination, individual randomisation can still result in a smaller sample size.

Reduced effect size

Although some contamination can be dealt with by increasing the sample size, a trial might still show a statistically significant effect but one that is too small to be of clinical relevance. Trials are rarely powered on the basis of a minimum clinical significance. Rather, they are powered on the likely effect that other trials have detected or on the basis of observational work or on logistical factors. Thus, the reduced effect size caused

by contamination could be used to calculate sample sizes.

Discussion

Often a cluster design is the correct design for a trial. For instance, a randomised trial of training teachers to promote smoking cessation demanded a cluster design as it would not have been practical to randomise children to different teachers.[13] In these instances one must be aware of the potential problems of selection bias creeping into the study. However, there is uncertainty with many trials. Thus, recent cluster trials which probably could have used an individually randomised design include a vitamin A supplementation study,[14] an accident prevention trial,[11] and a lifestyle behaviour changes study.[8] Because of the problems outlined in this paper, trialists ought to consider very carefully the cluster approach to trial design. Substantial contamination can be tolerated within the usual individual randomised trial before a cluster design is better in terms of total sample size.

I thank Hazel Inskip, the *BMJ*'s referee, for helpful comments on the original manuscript.

Competing interests: None declared.

1 Campbell MK, Grimshaw JM. Cluster randomised trials: time for improvement. *BMJ* 1998;317:1171-2.
2 Kerry SM, Bland M. The intracluster correlation coefficient in cluster randomisation. *BMJ* 1998;316:1455-60.
3 Ukoumunne OC, Gulliford MC, Chinn S, Sterne J, Burney P, Donner A. Evaluations of health care interventions at area and organisation level. In: Black N, Brazier J, Fitzpatrick R, Reeve B, eds. *Health services research methods: a guide to best practice.* London: BMJ Publishing, 1998.
4 Campbell MK, Grimshaw J, Steen N. Sample size calculations for cluster randomised trials. *J Health Services Res Policy* 2000;5:12-6.
5 Alexander FE, Anderson TJ, Brown HK, Forrest APM, Hepburn W, Kirkpatrick AE, et al. 14 years of follow-up from the Edinburgh randomised trial of breast-cancer screening. *Lancet* 1999;353:1903-8.
6 Zelen M. A new design for randomized clinical trials. *N Engl J Med* 1979;300:1242-5.

Table 2 Numbers of participants required in a cluster trial to detect a difference between two groups where the true effect size is 25%

	Cluster size						
Intracluster correlation coefficient	**1**	**10**	**20**	**30**	**40**	**50**	**100**
0.01:							
Sample size	116	126	139	150	161	173	231
Design effect	1	1.09	1.20	1.29	1.39	1.49	1.99
0.02:							
Sample size	116	137	160	184	207	230	346
Design effect	1	1.18	1.38	1.59	1.78	1.98	2.98
0.03							
Sample size	116	148	182	218	252	287	461
Design effect	1	1.27	1.57	1.88	2.17	2.47	3.97

7 Kerse NM, Flicker L, Jolley D, Arroll B, Young D. Improving the health behaviours of elderly people: randomised controlled trial of a general practice education programme. *BMJ* 1999;319:683-7.
8 Steptoe A, Doherty S, Rink E, Kerry S, Kendrick T, Hilton S. Behavioural counselling in general practice for the promotion of healthy behaviour among adults at increased risk of coronary heart disease: randomised trial. *BMJ* 1999;319:943-8.
9 Kinmouth AL, Woodcock A, Griffin S, Spiegal N, Campbell MJ. Randomised controlled trial of patient centred care of diabetes in general practice: impact on current wellbeing and future disease risk. *BMJ* 1998;317:1202-8.
10 Grossman DC, Neckerman HJ, Koepsell TD, Liu PY, Asher KN, Beland K, et al. Effectiveness of a violence prevention curriculum among children in elementary school: a randomised trial. *JAMA* 1997;277:1605-11.
11 Kendrick D, Marsh P, Fielding K, Miller P. Preventing injuries in children: cluster randomised controlled trial in primary care. *BMJ* 1999;318:980-3.
12 Altman DG, Whitehead J, Parmar MKB, Stenning SP, Fayers PM, Machin D. Randomised consent designs in cancer clinical trials. *Eur J Cancer* 1995;31A:1934-44.
13 Aveyard P, Cheng KK, Almond J, Sherratt E, Lancashire R, Lawrence T, et al. Cluster randomised controlled trial of expert system based on the transtheoretical ("stages of change") model for smoking prevention and cessation in schools. *BMJ* 1999;319:948-53.
14 West KP, Katz J, LeClerq SC, Pradham EK, Sheresta SR, Connor PB, et al. Double blind, cluster randomised trial of low dose supplementation with vitamin A or β carotene on mortality related to pregnancy in Nepal. *BMJ* 1999;318:570-5.

(Accepted 24 July 2000)

A memorable patient

A life saving procedure

A 34 year old temple priest from the Maharashtra district of central India was on his way home from a pilgrimage to a temple in northern India. He had been suffering from diarrhoea and vomiting for the past 12 hours. He thought he would cope with it until he arrived home and saw his doctor. But he was wrong. He was fast losing body water and getting colder. I know because I was travelling in the same train, in the same compartment. He was weak and could barely open his eyes. His pulse was weak, thready, and rapid, his extremities were cold and he was perspiring. I told his wife that he needed at least two lines in his arms and lots of fluids.

I requested the ticket checker to bring an emergency kit, but when I opened it I found it did not have intravenous fluids. All I could do was to give him an injection to keep up his blood pressure. The next stop was three and a half hours away, and that was too long for him. I discussed this again with his wife, and by now everybody in the compartment was involved. A unanimous decision was made to pull the chain and stop the train at some station on the way.

But then came the next problem. All the chains turned out to be jammed. It was a real surprise and nobody knew what to do next. There was no way of communication between the ticket checker and the driver.

I then suggested that when the train slowed down at some station on the way we should drop a note carrying a message for the station master to alert the master at the next station to halt our train there. We wrapped the note round a five rupee coin and threw it near enough for the signalman to notice. It worked. The train was stopped at the next station and the priest was lifted away to hospital, while his wife kept on thanking me.

I felt frustrated at not being of much help to him, and angry at the rail authorities for their negligence. Also I thought I would never be able to find out how the priest was.

Surprisingly, his wife had checked on the list of passengers and found my address, because when I reached home I found a letter of thanks from the priest. He had made an uneventful recovery after two days in hospital. After that he kept writing to me because he thought I had helped save his life—not because of the injection but because I had come up with the idea of stopping the train. Recently he has invited me to his elder daughter's wedding.

Kalpesh Shah *clinical observer, Ipswich*

We welcome articles of up to 600 words on topics such as *A memorable patient, A paper that changed my practice, My most unfortunate mistake*, or any other piece conveying instruction, pathos, or humour. If possible the article should be supplied on a disk. Permission is needed from the patient or a relative if an identifiable patient is referred to. We also welcome contributions for "Endpieces," consisting of quotations of up to 80 words (but most are considerably shorter) from any source, ancient or modern, which have appealed to the reader.

Full Text of this article
Email this article to a friend
Respond to this article
Alert me when:
New articles cite this article

Dealing with contamination in estimating sample size

The effect of potential contamination can be built into the sample size estimations as follows. Assume δ represents the difference to be detected in the absence of contamination of the control group then δ_c, the reduced difference because of contamination, is:

$$\delta_c=(\rho_1-\rho_2)(1-c)$$

where ρ_1 is the proportion with a positive outcome in the control group without contamination and ρ_2 is the proportion in the intervention group with c the estimated level of contamination.

The new sample size S_c is given by the formula:

$$S_c=(\delta^2/\delta_c^2)S$$

where S is the original sample size assuming no contamination.

For example, assume that there is a treatment to reduce the event rate from 50% in the control group to 25% in the intervention arm, and that this would require a sample of 116 patients (for 80% power and 5% significance). Assuming an estimated 20% contamination of the control group, the effect size would be reduced from 25% to 20% (that is, (0.50-0.25)(1-0.2)). The new sample size would therefore be 182 (that is, (0.25x0.25)/(0.20x0.20)x116), an increase of 56%. Although this increase is substantial, it is still less than the 62% increase that would have been needed had a cluster design been adopted (assuming an intracluster correlation coefficient of 0.02 and a cluster size of 30).

Full Text of this article
Email this article to a friend
Respond to this article
Alert me when:
New articles cite this article