Randomised trials – cluster versus individual randomisation

Primary Care Alliance for Clinical Trials (PACT) network

C Raina Elley, BA, MBCHB, FRNZCGP, PhD, is Senior Lecturer, Department of General Practice, Wellington School of Medicine, University of Otago, New Zealand.

Patty Chondros, BSc, GDipEpid, MSc, is a biostatistician, Department of General Practice, University of Melbourne, Victoria.

Ngaire M Kerse, BHB, MBChB, FRNZCGP, PhD, is Associate Professor, Department of General Practice and Primary Health Care, University of Auckland, New Zealand.

Have you made a difference to your patients with a new intervention? Usually this question is best answered by a randomised controlled trial (RCT). In RCTs, the effect of interventions is compared with 'usual care' or control group patients. This ensures (as much as possible) that any observed changes in the intervention group patients are due to the intervention itself, rather then systematic differences between the groups, or some background change in management practices that may be going on at the time of the trial.

However, many interventions appropriate for primary health care involve lifestyle interventions such as smoking cessation, dietary, or physical activity programs,[1–4] or educational programs for general practitioners.[5] What type of RCT design is appropriate for testing these interventions?

Individual randomisation

In a traditional RCT, it is the individual or patient that is allocated to an intervention (may be more than one intervention group) or control group, and simple statistical analyses on participant outcomes are used to evaluate if the intervention was effective. These analyses assume that all participants are completely independent (ie. unlike each other, do not influence each other, and any outcomes measured on them are influenced by the intervention or usual care in the same way). This is easy to imagine when the intervention is a drug that can be given to an intervention or control group participant. When asking whether an intervention makes a difference in general practice, schools, workplaces and other group settings, it can be more difficult to randomise at the individual level because:

- the intervention being tested involves the way the practice/school or workplace operates and cannot be turned on or off depending on which experimental group the participant is in, eg. if the intervention were the introduction of computerised medical records, all attending patients would have contact with the intervention
- the intervention can be affected by close proximity of control participants, eg. if a trial were being carried out for a treatment for head lice or impetigo, randomisation by school classroom would be sensible to avoid re-infection of intervention patients by control patients, and
- control participants may receive the intervention by 'mistake' and there may be risk of 'contamination'.

Contamination

Contamination[6] is defined as the proportion of individuals in the intervention group that are not exposed to the intervention, and the proportion of individuals in the control group that are exposed to the intervention.[7,8]

In general practice trials, participants are often aware of which group they have been allocated to. Because patients in one practice often know each other, there is a risk of control patients receiving some of the intervention (eg. information or advice about smoking cessation) if individuals from the same practice are randomised to the active or control group. Concern among control patients who believe they are 'missing out' could also influence behaviour change within the control group and alter the study outcome.

Contamination can also occur during the delivery of the intervention. The interventions tested are often related to the GP's behaviour or a practice's way of doing things. Health providers may have difficulty turning on and off the intervention (eg. delivery of advice to patients) depending on whether their individual patients have been allocated to an intervention or control group. Health providers may sometimes forget to deliver the advice to intervention patients. Alternatively, asking the provider not to deliver the advice to control patients may alter 'usual care', as some lifestyle advice may traditionally be part of 'usual care'.

Contamination will dilute the demonstrated effect, which may lead to a type 2 error (finding no effect when one actually exists). If the sample size is inflated sufficiently, it is possible to allow for this dilution, and subsequently find a true positive effect

$$IF = \frac{1}{(1 - \text{contamination})^2}$$

where contamination = p1 + p2

*p*1 = proportion of intervention group not exposed to intervention

*p*2 = proportion of control group exposed to intervention

Figure 1. Inflation factor to adjust sample size for contamination

$$DEFF = 1 + (m-1) \times ICC$$

m is the average cluster size

ICC is the intra-class correlation coefficient

Figure 2. Design effect for cluster randomisation

of the intervention.

Slymen and Hovell[7] provide an 'inflation factor' (*Figure 1*) to adjust the sample size to allow for contamination. However, it is difficult to predict how much contamination is likely to occur and therefore how much inflation of sample size is necessary.

Cluster randomisation

Cluster randomised trial designs can be used to overcome some of the problems associated with 'lifestyle intervention' trials in general practice. A cluster randomisation trial is one in which intact social units, or clusters of individuals rather than individuals themselves, are randomised to different intervention groups.[9]

When randomisation occurs at the group level, all participants recruited from the practice, school or workplace are allocated to either the intervention or the control group. The outcomes to measure the effect of the intervention are still assessed at the individual level, but the level at which the comparison is made is the practice, school or workplace.

Cluster RCT design is recommended when delivery of an intervention is likely to affect others within the group or cluster.[6,9–11] Cluster RCTs are being used increasingly where delivery of an intervention is at a group (or practice) level,[12,13] and outcomes are measured at the patient level.[14]

Table 1. Effect on sample size inflation using a cluster design of different cluster sizes in a lifestyle intervention trial

Sample size required for each group if randomised by individual	Cluster size (m)	ICC used	DEFF	Sample size required for each group adjusted for clustering
1400	93	0.02	2.8	3976
1400	80	0.02	2.58	3612
1400	50	0.02	1.98	2772
1400	20	0.02	1.38	1936
1400	10	0.02	1.18	1652
1400	1	0.02	NA	1400

What happens if you choose cluster randomisation?

Methodological implications

There are clear advantages with using a cluster RCT. Randomising the entire practice (or the GP) to the intervention or control group will reduce the risk of contamination. Members from intervention and control groups are less likely to have direct contact with each other and are less likely to pass on components of the intervention to the control group. There may also be increased compliance due to group participation. In addition, GPs or practices can be consistent in their management.

However, there are disadvantages in using a cluster design rather than an individual randomisation design. If a cluster design is chosen, the two traditional assumptions of RCTs are violated: ie. that all individuals are independent, and that analysis is at the level of randomisation.[9] Specific statistical methods need to be employed to adjust for these factors.

In general practice, patients (individuals) are not really independent. For example, people from families are more like each other than those not in their family, people from the same area are alike in socioeconomic status, and people consulting the same GP share some characteristics by the fact that

Table 2. The impact of different intra-class correlation coefficients for different variables on sample size inflation using a cluster design assuming a cluster size of 20[19,20]

Outcome variable	Baseline mean (SD)	Estimated change	ICC	DEFF	N (in each group)	Adjusted N
Systolic blood pressure (mmHg)	135 (18.8)	–3	0.018	1.342	617	828
Diastolic blood pressure (mmHg)	82 (12.2)	–2	0.046	1.874	585	1096
Weight (kg)	82 (19)	–3.5	0.043	1.817	463	841
Cholesterol (mmol/L)	5.7 (1.02)	–0.2	0.004	1.076	409	440
HDL (mmol/L)	1.34 (0.38)	0.06	0.039	1.741	630	1096
CHD risk (%)	5.6 (2.72)	–0.4	0.022	1.418	726	1029
General health	63 (21)	5	0	1	277	277

they chose that GP. Patients within one practice – as members of a group – both influence and are influenced by group membership.[15] People who share characteristics are more likely to respond to an intervention in a similar way than those who do not share those characteristics. This results in a loss of power, as the variation in outcomes between different practices is greater than would be expected in individually randomised trials. Therefore, using a cluster RCT design requires a larger sample than an individually randomised trial to overcome the loss of statistical power associated with randomising in groups, and the lack of independence between participants.[16,17]

Sample size adjustment with a cluster design

The design effect (DEFF) of randomising clusters instead of individuals is given as the ratio of the total number of subjects required using cluster randomisation to the number required using individual randomisation.[18] The DEFF factor we use to inflate our sample size to allow for the interdependency of individuals from the same clusters for different outcome variables is given in *Figure 2*. The intra-class correlation coefficient (ICC) is a statistical measure of the interdependence within each cluster and is calculated by taking the ratio of the variance between groups compared with variance within groups. The larger the ICC, or the larger the cluster size, the greater the effect on sample size when using clustering. Larger cluster sizes can drastically increase the overall sample size required. *Table 1* shows the effect on sample size for a lifestyle intervention trial designed to lower disease risk.

Table 2 shows that the ICC of different outcome variables has differential impact on the design effect of sample size calculations.[19,20] The design effect needs to be estimated for all outcome variables to ensure that sample size is adequate for these outcomes. Appropriate sample size calculation requires knowledge of ICCs. While this is sometimes estimated, better estimates of ICCs can be obtained from previous research.

Table 3. Number of participants required to detect lower weight in the lifestyle intervention group compared with the control group in an individually RCT with varying degrees of contamination

Contamination (%)	0%	5%	10%	15%	20%	25%	30%
Contamination effect (IF)	1	1.11	1.23	1.38	1.56	1.78	2.04
Sample size per group	1400	1554	1722	1932	2184	2492	2856

Table 4. Number of participants required to detect lower weight in the lifestyle intervention group compared with the control group in a cluster RCT

ICC	0	0.01	0.02	0.05	0.1	0.15	0.2
Average cluster size=72 subjects							
DEFF	1	1.71	2.42	4.56	8.11	11.67	15.22
Sample size	1400	2394	3388	6384	11354	16338	21308
Average cluster size=18 subjects							
DEFF	1	1.18	1.36	1.89	2.78	3.67	4.56
Sample size	1400	1652	1904	2646	3892	5138	6384

Therefore, it is important to publish the ICCs obtained from cluster RCTs in primary health care so they can be used in the design of future trials.[18,21]

Analysis of cluster RCTs

Because clustered participants are not independent and randomisation is at the group level with outcomes measured at the individual level, statistical adjustments are necessary during the analysis comparing intervention and control group outcomes.[16,22–24] Failure to adjust for the clustering of subjects in the analysis will produce results with artificially narrow confidence intervals, and increase the chance of finding a significant result when one does not exist (type 1 error).

The simplest analysis is to calculate summary statistics for each cluster and analyse the summary statistics using the usual inferential statistics. However, this analysis is limited, as it does not allow for adjustment of confounders and what may have seemed like a large data set is reduced to a summary measure for each cluster. With the advancement of statistical programs, complex statistical models can be applied to the individual values of subjects that allow for the effect of clustering.[13,16] Statistical packages such as 'STATA' or 'SAS', have capabilities for the complex statistical techniques such as generalised estimating and hierarchical modelling techniques and can be used to analyse the results of a cluster RCT.[16]

Although some studies have used cluster randomisation in health services research, many have not accounted for the cluster design during calculation of sample size or during analysis, therefore producing underpowered studies or overestimated results.[6,16] Reviews of RCTs using cluster designs have also highlighted the problem of inappropriate analysis techniques for cluster randomisation and lack of information about levels of intra-

class correlation.[21,25] There is usually an appropriate unit of analysis for intervention, depending on who delivers the intervention and what environmental influences are likely to have an effect on the outcome of the intervention. 'Errors' in the selected unit of analysis were found in 70% of studies in one review of intervention trials aimed at influencing physician practice.[26]

Other issues with cluster randomisation

Recruitment bias

There is a risk of recruitment bias if a cluster design is used. This is because researchers often know whether practices have been allocated to intervention or control before starting recruitment of subjects. This is known as 'pre-randomisation'. Knowing whether the group is going to be intervention or control may advertently or inadvertently influence who is recruited into the trial in each group. This introduces recruitment bias that can affect the validity and generalisability of the results. For example, patients with a poorer prognosis may be selected into the control group.

To avoid selection bias, all members of the cluster should be included in the trial. Alternatively, a random selection[14] or a systematic and representative sample of people from the cluster should be approached. Participant characteristics can be measured and compared between groups to check for differences and to assess whether there may have been recruitment bias. However, even when individuals are drawn from clusters randomly, there may be some imbalance of potential confounding factors between the groups by chance alone, especially if there are few clusters or if clusters differ markedly from each other. Comparisons of baseline characteristics of participants will help assess this and adjustments can be made using multivariate analysis.

'Blinding'

The traditional double blind trial, where neither the participant nor the outcome assessor know the experimental group allocation, is not possible in a cluster design. Blinding of outcome assessors is also difficult. If it becomes obvious which group one patient is in, then the assessor will be aware of the allocation of randomisation of all the patients in that group. In the instance of randomisation by practice, 'unblinding' can easily occur. If outcome assessment (eg. smoking status, level of activity) is determined by someone who knows the participant has received an intervention (nonblinded), the effect of the intervention can be overestimated by up to 18%.[27]

In addition, patients may be difficult to 'blind' in the traditional sense, especially if the intervention is for example, physiotherapy or counselling. Participants may tell the assessor they are in the intervention group. They may also report a more favourable outcome because they themselves know they are in the intervention group and 'should' be better. Objective outcomes should be used wherever possible to reduce the risk of assessment bias.

Sample size required and choice of trial design

If the inflation of the sample size required due to clustering is less than the inflation required to counter contamination in an individually RCT, then a cluster randomised trial would be more cost efficient.[7,8] *Table 3* and *Table 4* illustrate the different sample sizes required, depending on contamination in an individually randomised design and the ICC in a cluster randomised design. The example used is a theoretical lifestyle intervention trial designed to decrease obesity where 1400 patients are required in each study group to detect an effect on weight. The equation in *Figure 1* is used to calculate values in *Table 3*, and the equation in *Figure 2* to calculate values in *Table 4*. It is also evident that increasing the number of clusters is a more 'efficient' strategy than increasing the number in each cluster.

Conclusion

The correct answer to the question: 'Should we use cluster or individual randomisation?' will depend on the:

- type of intervention being tested
- level of intervention delivery, and
- risk of contamination.

If a cluster design is chosen (eg. at the level of GP, practice or community), increasing the number of clusters is more efficient in terms of sample size than increasing the number of participants per cluster. The decision whether to use individual or cluster randomisation should be considered early in the design of a study. Sample size calculations and analytical techniques need to be carried out using methods appropriate to the chosen design in order to obtain a valid result when assessing the effect of an intervention in primary health care.

Conflict of interest: this manuscript was accepted and edited by *AFP* staff members who have no university affiliations.

References

1. Steptoe A, Kerry S, Rink E, Hilton S. The impact of behavioural counselling on stage of change in fat intake, physical activity, and cigarette smoking in adults at increased risk of coronary heart disease. Am J Pub Health 2001;91:265–269.
2. Anonymous. Effectiveness of health checks conducted by nurses in primary care: final results of the OXCHECK study. Imperial Cancer Research Fund OXCHECK Study Group. BMJ 1995;310:1099–1104.
3. Ashenden R, Silagy C, Weller D. A systematic review of the effectiveness of promoting lifestyle change in general practice. Fam Pract 1997;14:160–176.
4. Burton LC, Paglia MJ, German PS, Shapiro S, Damiano AM. The effect among older persons of a general preventive visit on three health behaviours: smoking, excessive alcohol drinking, and sedentary lifestyle. The Medicare Preventive Services Research Team. Prev Med 1995;24:492–497.
5. Kerse N. Health promotion and older people: a general practice intervention study. Melbourne: The University of Melbourne, 1998.
6. Edwards SJ, Braunholtz DA, Lilford RJ, Stevens AJ. Ethical issues in the design and conduct of cluster randomised controlled trials. BMJ 1999;318:1407–1409.

7. Slymen DJ, Hovell MF. Cluster versus individual randomization in adolescent tobacco and alcohol studies: illustrations for design decisions. Int J Epidemiol 1997;26:765–771.
8. Torgerson DJ. Contamination in trials: is cluster randomisation the answer? BMJ 2001;322:355–357.
9. Donner A, Klar N. Design and analysis of cluster randomisation trials in health research. New York: Oxford University Press Inc, 2000.
10. Campbell MK, Grimshaw JM. Cluster randomised trials: time for improvement. The implications of adopting a cluster design are still largely being ignored. BMJ 1998;317: 1171–1172.
11. Eccles M, Grimshaw J, Steen N, et al. The design and analysis of a randomised controlled trial to evaluate computerised decision support in primary care: The COGENT study. Fam Pract 2000;17:180–186.
12. Yudkin PL, Moher M. Putting theory into practice: a cluster randomised trial with a small number of clusters. Stat Med 2001;20:341–349.
13. Raab GM, Butcher I. Balance in cluster randomised trials. Stat Med 2001;20:351–365.
14. 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 program. BMJ 1999; 319:683–687.
15. Goldstein MG, Pinto BM, Marcus BH, et al. Physician-based physical activity counselling for middle aged and older adults: a randomised trial. Ann Behav Med 1999;21:40–47.
16. Campbell MK, Mollison J, Steen N, Grimshaw JM, Eccles M. Analysis of cluster randomised trials in primary care: a practical approach. Fam Pract 2000;17:192–196.
17. Moore H, Summerbell C, Vail A, Greenwood DC, Adamson AJ. The design features and practicalities of conducting a pragmatic cluster randomised trial of obesity management in primary care. Stat Med 2001;20:331–340.
18. Kerry SM, Bland JM. The intracluster correlation coefficient in cluster randomisation. BMJ 1998;316:1455.
19. Elley C, Kerse N, Arroll B. Why target sedentary adults in primary health care? Baseline results from the Waikato Heart, Health and Activity Study. Prev Med 2003;37:342–348.
20. Elley C, Kerse N, Arroll B, Robinson E. Effectiveness of counselling patients on physical activity in general practice: cluster randomised controlled trial. BMJ 2003;362:793–796.
21. Donner A, Brown K, Brasher P. A methodological review of nontherapeutic intervention trials employing cluster randomisation 1979–1989. Int J Epidemiol 1990;19:795–800.
22. Bland J, Kerry S. Trials randomised in clusters. BMJ 1997;315:600.
23. Sashegyi AI, Brown KS, Farrell PJ. Application of a generalized random effects regression model for cluster correlated longitudinal data to a school based smoking prevention trial. Am J Epidemiol 2000;152:1192–1200.
24. Wood J, Freemantle N. Choosing an appropriate unit of analysis in trials of interventions that attempt to influence practice. J Health Serv Res Policy 1999;4:44–48.
25. Simpson JM, Klar N, Donnor A. Accounting for cluster randomisation: a review of primary prevention trials, 1990 through 1993. Am J Public Health 1995;85:1378–1383.
26. Divine GW, Brown JT, Frazier LM. The unit of analysis error in studies about physicians' patient care behaviour. J Gen Intern Med 1992;7:623–629.
27. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 1995;273:408–412.

AFP

Correspondence
Email: c.elley@wnmeds.ac.nz