Cluster randomization: a trap for the unwary

MARTIN UNDERWOOD

ADRIAN BARNETT

STEVEN HAJIOFF

SUMMARY

Controlled trials that randomize by practice can provide robust evidence to inform patient care. However, compared with randomizing by each individual patient, this approach may have substantial implications for sample size calculations and the interpretation of results. An increased awareness of these effects will improve the quality of research based on randomization by practice.

Keywords: randomized controlled trials; statistical method; primary care; survey design.

Introduction

BOTH the current fashion for evidence-based medicine and the move towards a primary care led National Health Service mean that there is an increased requirement for scientifically robust data from general practice on which to base clinical and administrative decisions. Randomized controlled trials (RCTs) based in general practice are one of the best sources of such data.

A conventional RCT, randomizing by individual patient, does not always lend itself to hypothesis testing within general practice. There are situations in which it is appropriate to randomize the intervention by practice rather than by individual. Examples of these situations arise when:

- It would be difficult or inappropriate to randomize to deny access to some patients within a practice; e.g. for a health promotion initiative when waiting room gossip or promotional material in the waiting room could affect the control group.
- The intervention or resource is expensive and therefore would need to be used fully to be cost-effective; e.g. if specialist diagnostic or computer equipment was being used.
- The intervention is, by necessity, practice or clinician based; e.g. an education programme aimed at general practitioners or other members of the primary health care team.

Cluster randomization

Randomization by practice (cluster randomization) can have a large effect on sample size requirements for, and analysis of, RCTs. This has not always been taken into account in published trials. Donner found that only three out of 16 trials randomizing by cluster produced a sample size justification based on cluster randomization.[1] More recently, Butler found that only three out of 10 trials of smoking cessation, based in primary care, had corrected for the effect of randomizing by cluster.[2]

M Underwood, MBChB, clinical scientist; and A Barnett, BSc, statistician, Medical Research Council Epidemiology and Medical Care Unit, London. S Hajioff, MBBCh, clinical lecturer, Department of General Practice and Primary Care, Queen Mary and Westfield College, London.
Submitted: 17 April 1997; accepted: 18 August 1997.

The usual requirements for calculating a sample size for an RCT include the following:

- that the subjects are expected to behave independently
- that a principal outcome measure has been defined that will be sensitive to differences between the two groups
- that a clinically significant difference between intervention and control groups is defined, and
- that the required probabilities of a Type I error (rejecting the null hypothesis when it is in fact true) and a Type II error (accepting the null hypothesis when it is in fact false) have been defined.

The impact on sample size of cluster randomization is caused by the tendency for patients from the same practice to behave similarly owing to factors within the practice. Thus, individual patients cannot be said to react with total independence, thereby invalidating one of the basic assumptions of most statistical analyses.

Differences between practices that are measurable, such as the age or social class of patients, can, to some extent, be corrected for in the analysis. However, other factors that are not quantifiable, such as the physical environment of the practice, the personal characteristics of the care providers, or the type of person attracted to a particular practice, cannot be corrected for in the same way.

Intra-cluster correlation coefficient

The magnitude of the effect of cluster randomization is quantified by the intra-cluster correlation coefficient (ICC), which is a statistical measure derived from the 'between' cluster and the 'within' cluster variation of the subjects.[3,4,5] If each individual's behaviour is unaffected by membership of the cluster, it will have no effect on the sample size calculation and the ICC will be zero. If all the individuals in one cluster behave in an identical manner, no statistical advantage will be gained from entering more than one individual from each cluster and the ICC will be one. So the ICC is a measure of the similarity of individuals (patients) within a cluster.

Assuming that the clusters are of similar sizes, the amount by which the overall sample size requirement has to be multiplied can be calculated from the equation $1+(\tilde{n}-1)\rho$, where ρ is the value of the ICC and $\tilde{n}$ is the average number of individuals in each cluster.[1] Primary care studies that have taken cluster effects into account in the analysis do not always state the value of the ICC in the final text.[6,7] Hence, calculating sample size can be difficult and estimates may be based on guesswork rather than genuine figures. In the North of England Study of Standards and Performance in General Practice, the value of ρ was greater than 0.1; for some intermediate outcomes, such as recording of general practitioner activity, and for some final outcome measures, such as prescribing rates, ρ was less than 0.01 (L Russell, personal communication, 1996). Values from general practice studies are commonly between 0.01 and 0.05 (M Campbell, personal communication, 1996).

The effect of cluster randomization can be demonstrated by two examples:

1. If there is a small ICC of 0.01 (i.e. individual behaviour is only affected to a minor degree by cluster membership) in a study that plans to recruit 10 patients in each of 10 practices,

the inflation factor will be 1.09. This will have little overall effect on the study design, only increasing the numbers required from 100 to 109 compared with randomization by individual.

2. If there is an ICC larger than 0.05 in a trial that plans to recruit 50 patients in each of 10 practices, then the inflation factor will be 3.45. This will have a major effect on study design, increasing the numbers required from 500 to 1725.

The most efficient results will be obtained from cluster randomization where the size of each cluster is small. Taken to its extreme, if each cluster has only one individual, then the statistical power is the same as for individual randomization. There is little advantage in increasing the size of each cluster above 50. If, in the second example, a sample size of 1000 was required using randomization by individual, the extra number of patients required could be achieved either by doubling the size of each cluster to 100 (which would mean that 5950 individuals would be required), or by keeping the size of each cluster at 50 and doubling the number of clusters (which would mean that 3450 individuals would be required).

Trial findings

Even if the number of patients in each cluster is large, scientifically robust trials are possible. In a trial of the effectiveness of dietary advice by practice nurses in lowering coronary heart disease risk, which recruited 956 patients from eight practices, a modest reduction in serum cholesterol was shown.[8] The statistical section explicitly states that cluster effects were taken into consideration in both trial design and analysis. It is therefore possible to be confident that the findings are scientifically robust. However, the values of ρ used in the sample size calculation and in the final analysis were not stated.

A study of a computer-based, active clinical decision support system in the care of patients with diabetes, which was performed by one of the authors (SH), required significant redesign. Because computerized Hb_{A1C} results were easily available, all patients known to have diabetes were studied. Twenty-four practices, each with an average of 200 diabetic patients, were randomly allocated to intervention or control. To show a difference of 0.5% in Hb_{A1C} between the two groups, with a significance of 5% and a power of 90%, assuming a standard deviation of 3%, randomizing by individual, and using a standard *t*-test, a total of 1514 diabetic patients was required. The 4800 patients available to the study comfortably exceeded this. However, in another study[9] of diabetes in primary care, the value of ρ for Hb_{A1C} was 0.018 (A-L Kinmonth, personal communication, 1996). This suggests that the required inflation factor is 24.5. Thus, to be confident of an adequate sample, 37 093 patients with diabetes from 186 practices would be required. Alternative outcome measures are now being used for this study.

When, as in this example, a whole practice intervention is used and the cluster size cannot be controlled, then outcomes that are less susceptible to changes at a practice level are to be preferred. It might be expected that different practices would be affected differently by the intervention. There might be large cluster effects with some outcome measures that could be influenced by individual doctor behaviour, such as the proportion of patients who had had their lipids measured or a retinal examination performed. The effect on Hb_{A1C} levels, which could be affected by many aspects of the improved care, would be expected to be smaller.

Conclusion

The case will arise where an important question can only be answered using cluster randomization where there is a high ICC. Researchers and funding bodies need to be prepared for the possibly large increase in study size required to obtain meaningful results.

Randomization by practice can give valuable unbiased data that may not be accessible using a conventional randomization; if the possible effects of cluster randomization are not taken into account, there is a potentially serious trap for the unwary researcher. We would reiterate Donner's suggestions[4] that all studies using cluster randomization state clearly that corrections have been made to account for this effect, and that when the value of ρ has been calculated it is included in the results to help other workers in designing their studies.

References

1. Donner A, Brown KS, Brasher P. A methodological review of non-therapeutic intervention trials employing cluster randomization 1979-89. *Int J Epidemiol* 1990; **19:** 795-800.
2. Butler C, Bachmann M. Design and analysis of studies evaluating smoking cessation interventions where effects vary between practices or practitioners. *Fam Pract* 1996; **13:** 402-407.
3. Cornfield J. Randomization by group: a formal analysis. *Am J Epidemiol* 1978; **108:** 100-102.
4. Donner A, Birkett N, Buck C. Randomization by cluster: sample size requirements and analysis. *Am J Epidemiol* 1981; **114:** 906-114.
5. Hsieh FY. Sample size formulae for intervention studies with the cluster as unit of randomization. *Stat Med* 1988; **8:** 1195-1201.
6. Wood DA, Kinmonth A-L, Davies GA, *et al.* Randomized controlled trial evaluating cardiovascular screening and intervention in general practice: principal results of the British Family Heart Study. *BMJ* 1994; **308:** 313-320.
7. North of England Study of standards and performance in general practice. Medical audit in general practice I: effects on doctors' clinical behaviour for common childhood conditions. *BMJ* 1992; **304:** 1480-1484.
8. Roderick P, Ruddock V, Hunt P, Miller G. A randomized trial to evaluate the effectiveness of dietary advice by practice nurses in lowering diet-related coronary heart disease risk. *Br J Gen Pract* 1997; **47:** 7-11.
9. Kinmonth A-L, Woodcock A, Griffen S, *et al. Diabetes care from diagnosis: the achievement of successful recruitment in a multipractice randomized control trial (RCT).* Proceedings of the AUDGP 25th Annual Scientific Meeting, Newcastle, 17-19 July, 1996.

Address for correspondence

Dr M Underwood, Medical Research Council Epidemiology and Medical Care Unit, Wolfson Institute of Preventive Medicine, St Bartholemew's and the Royal London School of Medicine and Dentistry, Queen Mary and Westfield College, Charterhouse Square, London EC1M 6BQ.