

Statistics notes

Sample size in cluster randomisation

Sally M Kerry, J Martin Bland

Techniques for estimating sample size for randomised trials are well established,^{1,2} but most texts do not discuss sample size for trials which randomise groups (clusters) of people rather than individuals. For example, in a study of different preparations to control head lice all children in the same class were allocated to receive the same preparation. This was done to avoid contaminating the treatment groups through contact with control children in the same class.³ The children in the class cannot be considered independent of one another and the analysis should take this into account.^{4,5} There will be some loss of power due to randomising by cluster rather than individual and this should be reflected in the sample size calculations. Here we describe sample size calculations for a cluster randomised trial.

For a conventional randomised trial assessing the difference between two sample means the number of subjects required in each group, n , to detect a difference of d using a significance level of 5% and a power of 90% is given by $n = 21s^2/d^2$ where s is the standard deviation of the outcome measure. Other values of power and significance can be used.¹

For a trial using cluster randomisation we need to take the design into account. For a continuous outcome measurement such as serum cholesterol values, a simple method of analysis is based on the mean of the observations for all subjects in the cluster and compares these means between the treatment groups. We will denote the variance of observations within one cluster by s_w^2 and assume that this variance is the same for all clusters. If there are m subjects in each cluster then the variance of a single sample mean is s_w^2/m . The true cluster mean (unknown) will vary from cluster to cluster, with variance s_c^2 . The observed variance of the cluster means will be the sum of the variance between clusters and the variance within clusters—that is, variance of outcome = $s_c^2 + s_w^2/m$. Hence we can replace s^2 by $s_c^2 + s_w^2/m$ in the formula for sample size above to obtain the number of clusters required in each intervention group. To do this we need estimates of s_c^2 and s_w^2 .

For example, in a proposed study of a behavioural intervention in general practice to lower cholesterol concentrations practices were to be randomised into two groups, one to offer intensive dietary intervention by practice nurses using a behavioural approach and the other to offer usual general practice care. The outcome measure would be mean cholesterol values in patients attending each practice one year later. Estimates of between practice variance and within practice variance were obtained from the Medical Research Council thrombosis prevention trial⁶ and were $s_c^2 = 0.0046$ and $s_w^2 = 1.28$ respectively. The minimum difference considered to be clinically relevant was 0.1 mmol/l. If we recruit 50 patients per practice, we would have $s^2 = s_c^2 + s_w^2/m = 0.0046 + 1.28/50 = 0.0302$. The number of practices is given by $n = 21 \times 0.0302/0.1^2 = 63$ in each group. We would require 63 practices in each group to detect a difference of 0.1 mmol/l with a power of 90%

Total number of practices required to detect a difference of 0.1 mmol/l cholesterol with 90% power at 5% significance level

No of patients per practice (m)	Standard deviation	No of practices	No of patients	Design effect
10	0.364	558	5 580	1.04
25	0.236	234	5 850	1.09
50	0.173	126	6 300	1.17
100	0.132	74	7 400	1.38
500	0.085	32	16 000	2.98
No needed with individual randomisation			5 364	1.00

using a 5% significance level—a total of 3150 patients in each group.

It can be seen from the formula for the variance of the outcome that when the number of patients within a practice, m , is very large, s_w^2/m will be very small and so the overall variance is roughly the same as the variance between practices. In this situation, increasing the number of patients per practice will not increase the power of the study. The table shows the number of practices required for different values of m , the number of subjects per practice. In all situations the total number of subjects required is greater than if simple random allocation had been used.

The ratio of the total number of subjects required using cluster randomisation to the number required using simple randomisation is called the design effect. Thus a cluster randomised trial which has a large design effect will require many more subjects than a trial of the same intervention which randomises individuals. As the number of patients per practice increases so does the design effect. In the table, the design effect is very small when m is less than 10. This would involve recruiting a total of 558 practices, and the nature of the intervention and difficulties in recruiting practices made this impractical. Thus it was decided to recruit fewer practices. The design effect of using 126 practices with 50 patients from each practice was 1.17. This design requires the total sample size to be inflated by 17%. If the study involves training practice based staff it may be cost effective to reduce the number of practices even further. If we chose to use 32 practices then we would need 500 patients from each practice and the design effect would be 2.98. Thus the cluster design with 32 practices would require the total sample size to be trebled to maintain the same level of power.

We shall discuss the use of the intraclass correlation coefficient in these calculations in a future statistics note.

- 1 Florey C du V. Sample size for beginners. *BMJ* 1993;306:1181-4.
- 2 Machin D, Campbell MJ. *Statistical tables for the design of clinical trials*. Oxford: Blackwell, 1987.
- 3 Chosidow O, Chastang C, Brue C, Bouvet E, Izri M, Monteny N, et al. Controlled study of malathion and d-phenothrin lotions for *Pediculus humanus* var *capitis*-infested schoolchildren. *Lancet* 1994;344:1724-7.
- 4 Bland JM, Kerry SM. Trials randomised in clusters. *BMJ* 1997;315:600.
- 5 Kerry SM, Bland JM. Analysis of a trial randomised in clusters. *BMJ* 1998;316:54.
- 6 Meade TW, Roderick PJ, Brennan PJ, Wilkes HC, Kelleher CC. Extracranial bleeding and other symptoms due to low dose aspirin and low intensity oral anticoagulation. *Thromb Haemostasis* 1992;68:1-6.

Division of General Practice and Primary Care, St George's Hospital Medical School, London SW17 0RE

Sally M Kerry, statistician
Department of Public Health Sciences

J Martin Bland, professor of medical statistics

Correspondence to: Mrs Kerry.

BMJ 1998;316:549