

## Clustering by health professional in individually randomised trials

Katherine J Lee, Simon G Thompson

Patient outcomes in many randomised trials depend crucially on the health professional delivering the intervention, but the resulting clustering is rarely considered in the analysis

MRC Biostatistics Unit, Institute of Public Health, Cambridge CB2 2SR  
Katherine J Lee  
*PhD student*  
Simon G Thompson  
*director*

Correspondence to: K J Lee  
kjl27@cam.ac.uk

BMJ 2005;330:142-4

Almost all trials that randomise individuals assume that the observed outcomes of participants are independent. The validity of this assumption is doubtful, however, in some situations. One example is when more than one health professional (such as surgeons, nurses, general practitioners, or therapists) delivers a non-pharmaceutical intervention to participants. Because health professionals may vary in their effectiveness, observations on participants treated by the same professional may be somewhat similar or clustered. Clustering of outcomes may also appear less obviously (such as in clustering by centre in a multicentre trial) or in a more dominant form (as in cluster randomised trials). In each of these situations the assumption of independence is violated, which means that standard statistical methods are invalid and may give misleading conclusions. The presence of clustering in a trial inflates standard errors and reduces the effective sample size, thus reducing the power of the trial. We examine the prevalence and importance of potential clustering in individually randomised trials and present an example of the effect it can have on the overall results and conclusions of a trial.

**Table 1** Assessment of clustering in 42 individually randomised trials published in *BMJ* in 2002

	No of trials
Clustering present:	
Any	38
Natural	33
Imposed	17
Mentioned in text	6
Effect on results:	
Unlikely	20
Possible	3
Likely	19
Adequately addressed	1

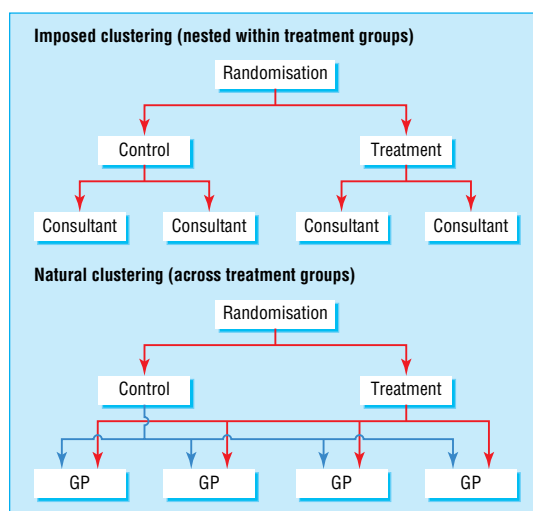
### Types of clustering

In a trial comparing a new one stop clinic with a dedicated breast clinic for breast cancer screening,<sup>1</sup> patients were randomised to a clinic, where they attended an appointment with one of several consultants. The main outcome was patient anxiety, which is likely to be influenced by the consultant treating the patient, yielding potential clustering by consultant. In this trial the clustering is imposed by the design of the trial because of the interventions being compared and is nested within treatment groups since each consultant participates in one treatment arm only (fig 1).

In another trial, comparing fusidic acid cream with placebo for the treatment of impetigo,<sup>2</sup> patients were recruited as they visited their general practitioner in their local practice. Outcomes may be more similar in the same practice than in different practices, either because of sociodemographic differences between patients or because of the general practitioner's or the practice's influence in delivering the intervention. In this trial, the clustering is natural rather than imposed, and clusters include patients in both treatment arms (fig 1).

### Is clustering common?

We reviewed all trials randomising individuals published in the *BMJ* during 2002, thus covering a wide



**Fig 1** Structure of imposed and natural clustering

References for trials and details of clustering and the statistical analysis are on [bmj.com](http://bmj.com)

range of medical areas. Trials randomising groups of individuals (cluster randomised trials) raise different issues and so were excluded.<sup>3</sup> We recorded information on:

- The type of clustering present—imposed or natural
- Whether the issue was recognised (irrespective of whether it was accounted for in the statistical analysis)
- How important we thought the clustering was, and
- Whether we felt it had been adequately accounted for in the statistical analysis.

The final two items are subjective. Clustering judged as likely to be important was directly related to the outcomes in the trial (such as a therapist delivering an intervention). All multicentre trials were recorded as having natural clustering irrespective of the number of centres. In a validation exercise, a second reader independently reviewed a random sample of seven trials; there was 95% agreement on the items recorded.

We identified 42 trials (see [bmj.com](http://bmj.com) for list). Thirty eight had some form of clustering, with 17 (40%) having clustering by health professional imposed by the design of the trial (table 1). Only six out of the 38 trials mentioned clustering as a potential issue. Four of these allowed in some way for clustering in the analysis of the trial's results, although three of the four failed to recognise multiple sources of clustering.

We classified clustering as unlikely to affect the results in 20 trials; four had no clustering (either single centre trials or trials where the intervention was delivered by a single health professional); and the remaining 16 were trials with natural clustering not directly related to the outcomes being assessed. Nineteen trials (45%) showed clustering that was likely to affect their results. Of these, only one attempted to take the issue into account. This trial, of community nurses specialising in Parkinson's disease,<sup>4</sup> explicitly investigated the potential variability in results between nurses, although the researchers found it insignificant and disregarded it in the final analysis. We conclude that some potential for clustering exists in almost all trials, with over a third of trials having clustering by health professional imposed by the design. This clustering is infrequently acknowledged, and even more rarely adequately addressed.

### Is clustering important?

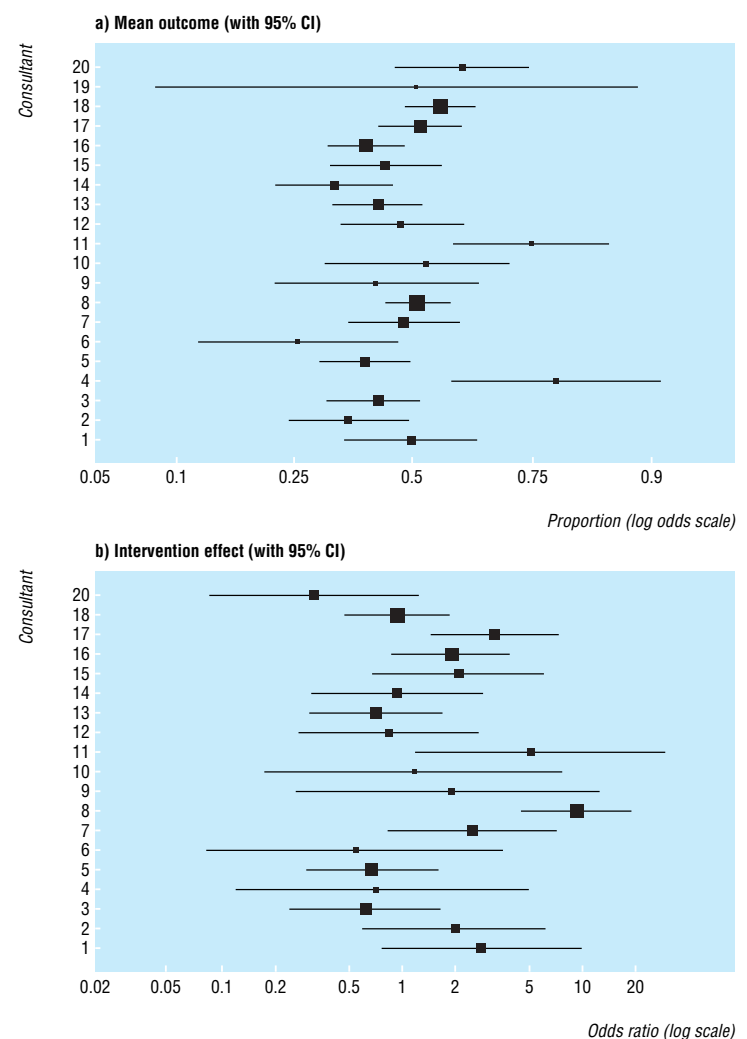
To show how clustering can affect a trial's results and conclusions, we use the example of a large trial of teleconsultation.<sup>5</sup> The trial of 2094 patients compared teleconsultations (video linked consultations between general practitioners and hospital consultants) with standard outpatient appointments for the referral of patients from primary to secondary care. The intervention was delivered by a range of consultants, who have a direct impact on the outcome, resulting in potentially important clustering. The crude results for the primary outcome (the proportion of patients subsequently offered a follow up appointment) were 52% in teleconsultation group and 41% in the standard outpatient group, corresponding to an odds ratio of 1.52.

In the trial, patients initially attended an appointment with a general practitioner in their local practice. Clustering of outcomes may exist for patients seen by the same general practitioner or in the same practice

(see [bmj.com](http://bmj.com)). Patients were then referred to consultants depending on their ailment, and it is the potential clustering by consultant that we focus on here.

Consultants may vary in their tendency to offer a follow up appointment, leading to heterogeneity in the outcomes across consultants. Disjoint confidence intervals for data from individual consultants in the trial shows this was the case (fig 2a). When we adjusted for this heterogeneity using random effects, while assuming that the intervention effect (the effect of teleconsultation) is the same across consultants, it widened the confidence interval for the estimated proportion of patients in the control group receiving an offer of a follow up appointment (table 2).

More importantly, the intervention effect may also vary between consultants. Some consultants may be happier with teleconsultation and offer fewer follow up appointments after a teleconsultation than others. When we allowed for the heterogeneity of the intervention effect across consultants (fig 2b) it made a substantial difference to the conclusions (table 2). The



**Fig 2** Variability among consultants in outcome (proportion of patients offered follow up appointment) and intervention effect (odds ratio of follow up appointment in intervention group v control group) in teleconsultation trial.<sup>5</sup> (One consultant had only one patient in the control arm and is not included in the intervention effect because a 95% confidence interval could not be calculated)

**Table 2** Effect of clustering on estimates of the proportion of patients offered a follow up appointment and the effect of teleconsultation in large trial<sup>a</sup>

	No clustering	Clustering of outcomes	Clustering of intervention effect
Overall proportion in control group (95% CI)	0.41 (0.38 to 0.45)	0.42 (0.35 to 0.48)	0.43 (0.35 to 0.52)
Odds ratio for intervention effect (95% CI)	1.52 (1.27 to 1.82)	1.55 (1.30 to 1.87)	1.36 (0.85 to 2.13)

<sup>a</sup>Details of model calculations are on [bmj.com](http://bmj.com).

intervention effect was reduced and its standard error inflated more than twofold. As a result, the confidence interval of the odds ratio includes 1, so that the increased overall proportion of patients offered follow up appointments is no longer convincing. This change in the overall conclusions is caused by the large variability between consultants (fig 2). Some of this variability is likely to be explained by the medical specialties of the consultants, as found in the original analysis.<sup>5</sup> A fuller set of results is on [bmj.com](http://bmj.com). This example shows that ignoring important clustering in the analysis of a trial can overstate the precision of results by yielding confidence intervals that are too narrow and P values that are too extreme. This can in turn produce misleading conclusions.

## Discussion

Our analysis shows that potential clustering of outcomes is common in trials that randomise individuals but is usually ignored in the analysis of results. A review of trials in psychotherapy research also found that two thirds of trials ignored clustering.<sup>6</sup> Clustering is most likely to have an important effect on results when health professionals actively deliver the intervention. Patients' outcomes may then depend crucially on the skill and enthusiasm of the professional involved. This may also be an issue in cluster randomised trials in which, although the clustering used in the randomisation process is recognised and generally adjusted for, further important forms of clustering may not be identified.

In the teleconsultation trial, the intervention effect became non-significant once clustering had been accounted for. Similar results have been reported in a trial comparing the delivery of minor acute care by the patient's general practitioner with that of commercial

deputising services,<sup>7</sup> in allowing for physician level clustering on the quality of diabetes care between specialty groups,<sup>8</sup> and in studies of small group teaching.<sup>9</sup>

Clustering reduces the effective sample size, reducing the power of a trial to detect an intervention effect.<sup>7-10</sup> Thus the best way to deal with the problem is to anticipate it at the time of design and to increase the sample size.<sup>9</sup> We do not recommend trying to identify clustering on the basis of a statistical test of significance. Such tests lack power, and a non-significant result does not rule out the presence of important clustering.<sup>11</sup> Rather, clustering should be anticipated on the basis of a trial's design, and so making allowance for it should follow as a consequence. Similar arguments apply both to cluster randomised trials<sup>3</sup> and to individually randomised multicentre or international trials.<sup>12-14</sup>

Clustering also affects the generalisability of conclusions. For example, in therapy trials, the sample of therapists in the trial should be representative of those who are going to deliver the intervention in practice. Even if this is the case, the analysis must acknowledge the clusters for the conclusions to be justified.<sup>6</sup> The issue of clustering of outcomes in randomised trials warrants much more attention than it has received so far, not only in design and analysis, but also in drawing justified and generalisable conclusions.

We thank the virtual outreach project team, in particular Paul Wallace and Julie Barber, for allowing us access to the data from the teleconsultation trial.

Contributors and sources: KJL is undertaking research on the topic of clustering in individually randomised trials. SGT has extensive experience in applied and methodological clinical trial research. SGT proposed the investigation. KJL reviewed the *BMJ* papers and reanalysed the teleconsultation trial. Both authors participated in drafting the article. KJL is guarantor.

Competing interests: None declared.

## Summary points

Clustering is common in individually randomised trials

The potential effects of clustering are generally ignored in the analysis of trial results

Clustering is particularly important when interventions are delivered by more than one health professional

Ignoring clustering can lead to incorrect conclusions

Bigger sample sizes are needed to accommodate the potential for clustering

- Dey P, Bundred N, Gibbs A, Hopwood P, Baildam A, Boggis C, et al. Costs and benefits of a one stop clinic compared with a dedicated breast clinic: randomised controlled trial. *BMJ* 2002;324:507-11.
- Koning S, Suijlekom-Smit LW, Nouwen JL, Verduin CM, Bernsen RM, Oranje AP, et al. Fusidic acid cream in the treatment of impetigo in general practice: double blind randomised placebo controlled trial. *BMJ* 2002;324:203-6.
- Donner A, Klar N. *Design and analysis of cluster randomization trials in health research*. London: Arnold, 2000.
- Jarman B, Hurwitz B, Cook A, Bajekal M, Lee A. Effects of community based nurses specialising in Parkinson's disease on health outcome and costs: randomised controlled trial. *BMJ* 2002;324:1072-5.
- Wallace P, Haines A, Harrison R, Barber J, Thompson SG, Jacklin P, et al. Joint teleconsultations (virtual outreach) versus standard outpatient appointments for patients referred by their general practitioner for a specialist opinion: a randomised trial. *Lancet* 2002;359:1961-8.
- Crits-Christoph P, Mintz J. Implications of therapist effects for the design and analysis of comparative studies of psychotherapies. *J Consult Clin Psychol* 1991;59:20-6.
- Roberts C. The implications of variation in outcome between health professionals for the design and analysis of randomised controlled trials. *Stat Med* 1999;18:2605-15.
- Greenfield S, Kaplan SH, Kahn R, Ninomiya J, Griffith JL. Profiling care provided by different groups of physicians: effects of patient case-mix (bias) and physician-level clustering on quality assessment results. *Ann Intern Med* 2002;136:111-21.
- Hoover DR. Clinical trials of behavioural interventions with heterogeneous teaching subgroup effects. *Stat Med* 2002;21:1351-64.
- Martindale C. The therapist as fixed effect fallacy in psychotherapy research. *J Consult Clin Psychol* 1978;46:1526-30.
- Paul SR, Donner A. Small sample performance of tests of homogeneity of odds ratios in k 2x2 tables. *Stat Med* 1992;11:159-65.
- Localio AR, Berlin JA, Ten Have TR, Kimmel SE. Adjustments for center in multicenter studies: an overview. *Ann Intern Med* 2001;135:112-23.
- Jones B, Teather D, Wang J, Lewis JA. A comparison of various estimators of a treatment difference for a multi-centre clinical trial. *Stat Med* 1998;17:1767-77.
- Senn S. Some controversies in planning and analysing multi-centre trials. *Stat Med* 1998;17:1753-65.

(Accepted 8 November 2004)