BMJ

ENDGAMES

STATISTICAL QUESTION

Cluster sampling

Philip Sedgwick *reader in medical statistics and medical education*

Centre for Medical and Healthcare Education, St George's, University of London, London, UK

Researchers investigated the effectiveness of providing smoking cessation support to adult smokers admitted to hospital. A cluster randomised controlled trial study design was used. The intervention comprised systematic smoking ascertainment and default provision of behavioural support and cessation pharmacotherapy for the duration of the hospital stay, with follow-up and referral to community services after discharge. Control treatment comprised usual care, with cessation support delivered at the initiative and discretion of clinical staff.[1]

Participants were recruited using cluster sampling. Smokers and recent ex-smokers admitted to 18 acute medical wards in one large UK teaching hospital between 11 October 2010 and 9 August 2011 were invited to take part. In total, 493 patients were recruited. Wards were allocated to treatment using random allocation, stratified by the number of discharges per week. Nine wards were allocated to each treatment, resulting in 264 patients receiving the intervention and 229 the control treatment.

The primary outcome measure was smoking cessation at four weeks, validated by measuring exhaled carbon monoxide. A greater proportion of the intervention group achieved smoking cessation at four weeks compared with controls (adjusted odds ratio 2.10, 95% confidence interval 0.96 to 4.61; P=0.06). It was concluded that improvements in smoking cessation among smokers and recent ex-smokers admitted to hospital can be achieved by systematic ascertainment and delivery of cessation support in secondary care.

Which of the following statements, if any, are true?

a) Cluster sampling involved recruiting a random sample of adult patients from each hospital ward

b) Cluster sampling was used to minimise contamination between groups in the delivery of treatment

c) By definition, cluster sampling constitutes probability sampling

Answers

Statements *b* and *c* are true, whereas *a* is false.

The purpose of the trial was to investigate the effectiveness of the systematic provision of smoking cessation support to adult smokers and recent ex-smokers admitted to hospital. Control treatment was usual care, with cessation support delivered at the initiative and discretion of clinical staff. Participants were recruited using cluster sampling, with all adult smokers and recent ex-smokers admitted to 18 wards in a UK teaching hospital between 11 October 2010 and 9 August 2011 invited to take part.

Clusters are natural groupings of people—for example, electoral wards, general practices, and schools. Cluster sampling involves obtaining a random sample of clusters from the population, with all members of each selected cluster invited to participate (*a* is false). It is necessary to construct a sampling frame listing all clusters in the population. A sample of a fixed number of clusters is selected at random from this list. Each cluster has the same probability of being selected, independently of all others. However, if the size of clusters varies then the probability of selection may be proportional to the size of the cluster, with larger clusters having a larger probability of selection. Obtaining a random sample of clusters can be time consuming, expensive, and impractical, not least because clusters will be diverse geographically. Therefore, cluster sampling sometimes uses a random sample of clusters from a conveniently selected geographical region.

In the example above, simple random sampling could have been used to recruit patients from the population. It would be necessary to construct a sampling frame—that is, a list of all people belonging to the population. A sample of a fixed size would be selected at random from this list, with all members of the population having the same probability of being selected, independently of all others. However, such sampling would have resulted in a sample whose members were widely dispersed geographically. Therefore, cluster sampling of hospital wards was used because it concentrated resources in fewer places. A sample of 18 wards from a single UK teaching hospital was selected. The wards were all in the same hospital, and therefore in effect constituted a convenience sample. Convenience sampling has been described in a previous question.[2] All adult smokers and recent ex-smokers admitted to these wards were invited to participate (*a* is false). However, not all those patients who were invited consented to taking part in the trial.

p.sedgwick@sgul.ac.uk

The use of cluster sampling in the trial above facilitated cluster allocation—that is, the allocation of wards rather than of the patients themselves to the intervention or control. Wards were randomly allocated to treatment, stratified by the number of discharges per week. Stratified random allocation has been described in a previous question.[3] All the patients in each ward then received the same treatment—intervention or control—that their ward had been allocated. Cluster allocation minimised practical and contamination problems (*b* is true) that would probably have arisen if simple random allocation had been used. Simple random allocation would have resulted in patients in the same ward being allocated to different treatment groups. If simple random allocation had been used, it may have been difficult for staff to implement the intervention within a ward for some but not other patients. Furthermore, within a ward patients allocated to the intervention may have influenced the activity of those allocated to the control treatment, or vice versa. The allocation of clusters to treatment is essential to the design of cluster randomised controlled trials.[4]

Two types of sampling method can be used to recruit participants to a study—random sampling (sometimes called probability sampling) and non-random sampling (sometimes called non-probability sampling). Random sampling involves some form of random selection of the population members. Simple random sampling (sometimes referred to simply as random sampling), described above, is the most straightforward type of probability sampling.

By definition, cluster sampling constitutes probability sampling (*c* is true). As described above, cluster sampling involves obtaining a random sample of clusters from the population, with all members of each selected cluster invited to participate. Therefore, each member of the population has a fixed probability of being selected independently of all others. In the example above, the wards were in effect chosen using convenience sampling—it was convenient to sample the 18 wards because they were all in the same hospital and easily accessible to the researchers. Convenience sampling constitutes non-probability sampling because the resulting sample is not chosen using any random selection of the members of the population.[2] Therefore, the particular cluster sampling approach used above is probably best described as non-probability sampling. In particular, not all of the hospital wards in the population had the opportunity of being selected and included in the sample.

Cluster sampling is commonly implemented as part of multistage cluster sampling, often referred to simply as multistage sampling. An example of multistage sampling has been given in a previous question.[5] Multistage sampling entails two or more stages of random sampling based on the hierarchical structure of natural clusters within the population. A different type of cluster is sampled at each stage, with the clusters nested within each other at successive stages. The final stage of sampling involves choosing a random sample of people in the clusters selected at the penultimate stage. Simple cluster sampling as used in the study described here is not regarded as multistage sampling, not least because only one type of cluster (the ward) was sampled. Furthermore, all patients in the wards were invited to participate in the study. This is in contrast to multistage sampling, where a random sample of participants is recruited in the final stage of sampling.

Competing interests: None declared.

1 Murray RL, Leonardi-Bee J, Marsh J, Jayes L, Li J, Parrott S, et al. Systematic identification and treatment of smokers by hospital based cessation practitioners in a secondary care setting: cluster randomised controlled trial. *BMJ* 2013;347:f4004.
2 Sedgwick P. Convenience sampling. *BMJ* 2013;347:f6304.
3 Sedgwick P. Stratified random allocation. *BMJ* 2013;346:f822.
4 Sedgwick P. Cluster randomised controlled trials. *BMJ* 2012;345:e4654.
5 Sedgwick P. Stratified cluster sampling. *BMJ* 2013;347:f7016.

Cite this as: *BMJ* 2014;348:g1215